Abdominal ultrasound and alpha-foetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease

Colli, Agostino; Nadarević, Tin; Miletić, Damir; Giljača, Vanja; Fraquelli, Mirella; Štimac, Davor; Casazza, Giovanni

Source / Izvornik: **Cochrane Database of Systematic Reviews, 2021, 2021**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.1002/14651858.CD013346.pub2

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:184:279175

Cochrane Library

Cochrane Database of Systematic Reviews

Abdominal ultrasound and alpha-foetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease (Review)

Colli A, Nadarevic T, Miletic D, Giljaca V, Fraquelli M, Štimac D, Casazza G

Colli A, Nadarevic T, Miletic D, Giljaca V, Fraquelli M, Štimac D, Casazza G.
Abdominal ultrasound and alpha-foetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease.
Cochrane Database of Systematic Reviews 2021, Issue 4. Art. No.: CD013346.
DOI: 10.1002/14651858.CD013346.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Abdominal ultrasound and alpha-foetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease

Agostino Colli[1], Tin Nadarevic[2], Damir Miletic[3], Vanja Giljaca[4], Mirella Fraquelli[5], Davor Štimac[6], Giovanni Casazza[7]

[1]Department of Transfusion Medicine and Haematology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy. [2]Department of Radiology, Clinical Hospital Centre Rijeka, Rijeka, Croatia. [3]Department of Radiology , Clinical Hospital Centre Rijeka, Rijeka, Croatia. [4]Department of Gastroenterology, Heart of England NHS Foundation Trust, Birmingham, UK. [5]Gastroenterology and Endoscopy Unit, Fondazione IRCCS Ca´ Granda - Ospedale Maggiore Policlinico, Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy. [6]Department of Gastroenterology, Clinical Hospital Centre Rijeka, Rijeka, Croatia. [7]Dipartimento di Scienze Biomediche e Cliniche "L. Sacco", Università degli Studi di Milano, Milan, Italy

Contact: Agostino Colli, colliagostino@gmail.com.

Editorial group: Cochrane Hepato-Biliary Group.
Publication status and date: New, published in Issue 4, 2021.

Citation: Colli A, Nadarevic T, Miletic D, Giljaca V, Fraquelli M, Štimac D, Casazza G. Abdominal ultrasound and alpha-foetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. *Cochrane Database of Systematic Reviews* 2021, Issue 4. Art. No.: CD013346. DOI: 10.1002/14651858.CD013346.pub2.

ABSTRACT

Background

Hepatocellular carcinoma (HCC) occurs mostly in people with chronic liver disease and ranks sixth in terms of global instances of cancer, and fourth in terms of cancer deaths for men. Despite that abdominal ultrasound (US) is used as an initial test to exclude the presence of focal liver lesions and serum alpha-foetoprotein (AFP) measurement may raise suspicion of HCC occurrence, further testing to confirm diagnosis as well as staging of HCC is required. Current guidelines recommend surveillance programme using US, with or without AFP, to detect HCC in high-risk populations despite the lack of clear benefits on overall survival. Assessing the diagnostic accuracy of US and AFP may clarify whether the absence of benefit in surveillance programmes could be related to under-diagnosis. Therefore, assessment of the accuracy of these two tests for diagnosing HCC in people with chronic liver disease, not included in surveillance programmes, is needed.

Objectives

Primary: the diagnostic accuracy of US and AFP, alone or in combination, for the diagnosis of HCC of any size and at any stage in adults with chronic liver disease, either in a surveillance programme or in a clinical setting.

Secondary: to assess the diagnostic accuracy of abdominal US and AFP, alone or in combination, for the diagnosis of resectable HCC; to compare the diagnostic accuracy of the individual tests versus the combination of both tests; to investigate sources of heterogeneity in the results.

Search methods

We searched the Cochrane Hepato-Biliary Group Controlled Trials Register, the Cochrane Hepato-Biliary Group Diagnostic-Test-Accuracy Studies Register, Cochrane Library, MEDLINE, Embase, LILACS, Science Citation Index Expanded, until 5 June 2020. We applied no language or document-type restrictions.

Selection criteria

Studies assessing the diagnostic accuracy of US and AFP, independently or in combination, for the diagnosis of HCC in adults with chronic liver disease, with cross-sectional and case-control designs, using one of the acceptable reference standards, such as pathology of the

explanted liver, histology of resected or biopsied focal liver lesion, or typical characteristics on computed tomography, or magnetic resonance imaging, all with a six-months follow-up.

Data collection and analysis

We independently screened studies, extracted data, and assessed the risk of bias and applicability concerns, using the QUADAS-2 checklist. We presented the results of sensitivity and specificity, using paired forest-plots, and tabulated the results. We used a hierarchical meta-analysis model where appropriate. We presented uncertainty of the accuracy estimates using 95% confidence intervals (CIs). We double-checked all data extractions and analyses.

Main results

We included 373 studies. The index-test was AFP (326 studies, 144,570 participants); US (39 studies, 18,792 participants); and a combination of AFP and US (eight studies, 5454 participants).

We judged at high-risk of bias all but one study. Most studies used different reference standards, often inappropriate to exclude the presence of the target condition, and the time-interval between the index test and the reference standard was rarely defined. Most studies with AFP had a case-control design. We also had major concerns for the applicability due to the characteristics of the participants.

As the primary studies with AFP used different cut-offs, we performed a meta-analysis using the hierarchical-summary-receiver-operating-characteristic model, then we carried out two meta-analyses including only studies reporting the most used cut-offs: around 20 ng/mL or 200 ng/mL.

AFP cut-off 20 ng/mL: for HCC (147 studies) sensitivity 60% (95% CI 58% to 62%), specificity 84% (95% CI 82% to 86%); for resectable HCC (six studies) sensitivity 65% (95% CI 62% to 68%), specificity 80% (95% CI 59% to 91%).

AFP cut-off 200 ng/mL: for HCC (56 studies) sensitivity 36% (95% CI 31% to 41%), specificity 99% (95% CI 98% to 99%); for resectable HCC (two studies) one with sensitivity 4% (95% CI 0% to 19%), specificity 100% (95% CI 96% to 100%), and one with sensitivity 8% (95% CI 3% to 18%), specificity 100% (95% CI 97% to 100%).

US: for HCC (39 studies) sensitivity 72% (95% CI 63% to 79%), specificity 94% (95% CI 91% to 96%); for resectable HCC (seven studies) sensitivity 53% (95% CI 38% to 67%), specificity 96% (95% CI 94% to 97%).

Combination of AFP (cut-off of 20 ng/mL) and US: for HCC (six studies) sensitivity 96% (95% CI 88% to 98%), specificity 85% (95% CI 73% to 93%); for resectable HCC (two studies) one with sensitivity 89% (95% CI 73% to 97%), specificity of 83% (95% CI 76% to 88%), and one with sensitivity 79% (95% CI 54% to 94%), specificity 87% (95% CI 79% to 94%).

The observed heterogeneity in the results remains mostly unexplained, and only in part referable to different cut-offs or settings (surveillance programme compared to clinical series). The sensitivity analyses, excluding studies published as abstracts, or with case-control design, showed no variation in the results.

We compared the accuracy obtained from studies with AFP (cut-off around 20 ng/mL) and US: a direct comparison in 11 studies (6674 participants) showed a higher sensitivity of US (81%, 95% CI 66% to 90%) versus AFP (64%, 95% CI 56% to 71%) with similar specificity: US 92% (95% CI 83% to 97%) versus AFP 89% (95% CI 79% to 94%). A direct comparison of six studies (5044 participants) showed a higher sensitivity (96%, 95% CI 88% to 98%) of the combination of AFP and US versus US (76%, 95% CI 56% to 89%) with similar specificity: AFP and US 85% (95% CI 73% to 92%) versus US 93% (95% CI 80% to 98%).

Authors' conclusions

In the clinical pathway for the diagnosis of HCC in adults, AFP and US, singularly or in combination, have the role of triage-tests. We found that using AFP, with 20 ng/mL as a cut-off, about 40% of HCC occurrences would be missed, and with US alone, more than a quarter. The combination of the two tests showed the highest sensitivity and less than 5% of HCC occurrences would be missed with about 15% of false-positive results. The uncertainty resulting from the poor study quality and the heterogeneity of included studies limit our ability to confidently draw conclusions based on our results.

PLAIN LANGUAGE SUMMARY

Abdominal ultrasound and alpha-foetoprotein for the diagnosis of hepatocellular carcinoma

Why is improving the diagnosis of hepatocellular carcinoma important?

Hepatocellular carcinoma (HCC), i.e. cancer originating in the liver, is sixth in terms of global occurrences of cancer and fourth in terms of cancer deaths in men. This cancer occurs mostly in people with chronic liver disease regardless of the cause. Ultrasound (US), which uses ultrasound waves to show abnormalities in the liver, can detect the presence of liver lesions suspected of being HCC. Alpha-foetoprotein (AFP), a glycoprotein, produced by the liver and measurable in the blood, is considered a tumour-marker because high levels can be associated with the presence of HCC. These two tests (US and AFP) are used, alone or in combination, to exclude the presence of HCC

in people at high risk of developing HCC. People at high risk are those who have chronic liver disease. Current guidelines recommend surveillance programmes, repeating abdominal US with or without AFP testing every six months to detect early HCC, amenable to surgical resection or other treatment.

What is the aim of this review?

To find out how accurate AFP, US, and a combination of AFP and US are for diagnosing HCC in people with chronic liver disease.

What was studied in this review?

AFP (tumour marker), that can easily be measured in the blood, using a commercial kit. Studies with AFP used various threshold values for defining the test as positive or negative.

US is an equipment, available worldwide. It produces images of liver and other abdominal organs. It can detect the presence of liver lesions suspected of being HCC.

A combination of AFP and US can detect or negate the presence of liver lesions suspected of being HCC.

What are the main results in this review?

We found 373 total studies in adults: AFP was analysed in 326 studies, 144,570 participants; US in 39 studies, 18,792 participants; and the combination of AFP and US in eight studies, 5454 participants.

- AFP with threshold of 20 ng/mL (147 studies): the test was positive in 60 out of 100 participants with HCC and in 16 out of 100 participants without HCC. AFP with threshold of 200 ng/mL (56 studies): the test was positive in 36 out of 100 participants with HCC and only in 1 out of 100 without HCC.
- US (39 studies): the test was positive in 72 out of 100 participants with HCC and in 6 out of 100 participants without HCC.
- The combination of AFP with threshold of 20 ng/mL and US (6 studies): one or both tests were positive in 96 out of 100 participants with HCC and in 15 out of 100 participants without HCC.

Thus, the combination of the two tests is better in detecting participants with HCC. Considering that people with chronic liver disease have HCC in 5 out of 100, one can assume that among 1000 people with chronic liver disease, 50 will have HCC, and, using AFP and abdominal US in combination, one can detect 48 out of the people with HCC, and 2 people will go undetected and will not receive appropriate treatment; 950 out of 1000 will have no HCC, and 143 of them will receive a wrong diagnosis of HCC, and will undergo further unnecessary testing such as computed tomography, magnetic resonance imaging, or biopsy.

How reliable are the results of the studies in this review?

All but one study had issues with risk of bias, especially in participants selection and in the correct definition on presence of HCC. These problems could impair the correct estimates of the diagnostic ability of the three tests.

Who do the results of this review apply to?

People with chronic liver disease

What are the implications of this review?

Using AFP, with 20 ng/mL, as threshold, about 40% of HCC occurrences would be missed, and with US alone, more than a quarter. The sensitivity was highest when the two tests were used in combination, and less than 5% of HCC occurrences would be missed with about 15% of false-positive results.

How up-to-date is this review?

5 June 2020

SUMMARY OF FINDINGS

Summary of findings 1. 'Summary of findings' table: diagnostic accuracy of AFP, US, and combination of AFP and US for the diagnosis of HCC

Review question: what is the diagnostic accuracy of alpha-foetoprotein (AFP), abdominal ultrasound (US), or of the combination of AFP and abdominal US for the diagnosis of hepatocellular carcinoma (HCC) in adults with chronic liver disease?

Population: adults with chronic liver disease

Setting: clinical setting (secondary or tertiary care setting) or surveillance programs

Study design: prospective and retrospective cross-sectional and case-control studies

Index tests

Serum alpha-foetoprotein (AFP) measurement with a cut-off value of 20 ng/mL

Serum alpha-foetoprotein (AFP) measurement with a cut-off value of 200 ng/mL

Abdominal ultrasound (US)

Combination of serum alpha-foetoprotein (AFP) measurement with a cut-off value of 20 ng/mL and abdominal ultrasound (US)

Target condition: HCC of any size, any stage

Reference standards:

the pathology of the explanted liver in case of transplantation; the histology of resected focal liver lesion(s), or the histology of resected or biopsied focal liver lesion(s) with a follow-up period of at least six months to exclude the presence of focal lesions non detected by the index test and synchronous lesions from the parenchyma surrounding the resected or biopsied area;

typical characteristics on cross-sectional multiphasic contrast computer tomography (CT) or magnetic resonance imaging (MRI), with a follow-up period of at least six months in order to allow the confirmation of an initial negative result on CT or on MRI.

Limitations in the evidence - Risk of bias/Applicability

Index test: serum alpha-foetoprotein (AFP) measurement cut-off value 20 ng/mL

- Participant selection: high/unclear risk of bias 141 studies (96%), high concern 115 studies (78%)

- Index tests: high/unclear risk of bias in 73 studies (50%) high concern: no study

- Reference standard: high/unclear risk of bias in 105 studies (71%) high concern 33 studies (22%)

- Flow and timing: high risk of bias in 143 studies (97%)

Index test: serum alpha-foetoprotein (AFP) measurement cut-off value 200 ng/mL

- Participant selection: high/unclear risk of bias 48 studies (86%), high concern 47(84%)

- Index tests: high/unclear risk of bias in 54 studies (96%) high concern no study

- Reference standard: high/unclear risk of bias in 39 studies (70%) high concern 13 studies (23%)

- Flow and timing: high risk of bias in 55 studies (98%)

Index test: abdominal ultrasound

- Participant selection: high/unclear risk of bias in 23 studies (59%) high concern 22 studies (56%)

- Index tests: high/unclear risk of bias in 15 studies (38%) high concern no study

- Reference standard: high/unclear risk of bias in 27 studies (69%) high concern 13 studies (33%)

- Flow and timing: high risk of bias in 27 studies (TN) (69%)

Index test: combination of serum alpha-foetoprotein (AFP) measurement with a cut-off value of 20 ng/mL and abdominal ultrasound

- Participant selection: high/unclear risk of bias in 2 studies (33%) high concern 2 studies (33%)

- Index tests: high/unclear risk of bias in 2 studies (33%) high concern no study

- Reference standard: high/unclear risk of bias in 4 studies (67%) high concern one study (17%)

- Flow and timing: high risk of bias in 6 studies (100%)

Findings

				Implications in a hypothetical cohort of 1000 people					
Index test	**Number of studies (participants)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Assumed prevalence of hepatocellular carcinoma (HCC)[a] %**	**True positives** will receive appropriately further necessary testing with CT or MRI, or contrast enhanced ultrasound (CEUS) and possibly treatment.	**False negatives** will be misdiagnosed and not receive appropriate treatment.	**True negatives** will not appropriately undergo unnecessary further testing with CT, MRI, CEUS, biopsy.	**False positives** will inappropriately undergo further unnecessary testing with CT, MRI, CEUS biopsy.	**Certainty of the evidence**
AFP (cut-off 20 ng/mL)	147 (52144)	59.8% (57.9% to 61.7%)	84.4% (82.3% to 86.3%)	5%	30	20	802	148	very low [b] ⊕◯◯◯
				30%	179	121	591	109	

AFP (cut-off 200 ng/mL)	56 (20452)	36% (31% to 41%)	99% (98% to 100%)	5%	18	32	940	10	very low [c] ⊕◯◯◯
				30%	108	192	693	7	
US	39 (18792)	72% (63% to 79%),	94% (91% to 96%)	5%	36	14	893	57	very low [d] ⊕◯◯◯
				30%	216	84	658	42	
Combination of AFP (cut-off 20 ng/mL) and US	6 (5044)	96% (88% to 98%)	85% (73% to 93%)	5%	48	2	807	143	low [e] ⊕⊕◯◯
				30%	288	12	595	105	

[a] We chose for exemplification two values of HCC prevalence: 5% for a population at low risk (compensated advanced chronic liver disease and chronic viral hepatitis) Lok 2009 and 30% for a population with high risk, a median of the prevalence in the included cross-sectional studies conducted in clinical cohorts.
[b] Downgraded by three levels: risk of bias, indirectness, and inconsistency. Risk of bias downgraded one level because all studies were judged at high risk of bias; indirectness downgraded one level as we considered most studies to have concern regarding applicability mainly in relation to the population (including disease spectrum); inconsistency downgraded one level as for individual studies ranged from 24% to 90% and we could not explain the heterogeneity by study quality or other factors
[c] Downgraded by three levels: risk of bias, indirectness, and inconsistency. Risk of bias downgraded one level because all studies were judged at high risk of bias; indirectness downgraded one level as we considered most studies to have concern regarding applicability mainly in relation to the population (including disease spectrum); inconsistency downgraded one level as for individual studies ranged from 4% to 83% and we could not explain the heterogeneity by study quality or other factors
[d] Downgraded by three levels: risk of bias, indirectness, and inconsistency. Risk of bias downgraded one level because all studies were judged at high risk of bias; indirectness downgraded one level as we considered most studies to have concern regarding applicability mainly in relation to the population (including disease spectrum); inconsistency downgraded one level as for individual studies ranged from 28%to 100% and we could not explain the heterogeneity by study quality or other factors
[e]Downgraded by two levels: risk of bias, indirectness. Risk of bias downgraded one level because all studies were judged at high risk of bias; indirectness downgraded one level as we considered most studies to have concern regarding applicability mainly in relation to the population (including disease spectrum).
GRADE certainty of the evidence
High: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.
The results presented in this table should not be interpreted in isolation from results of the individual included studies contributing to each summary test accuracy measure.

Summary of findings 2. 'Summary of findings' table: direct comparison of US, and combination of AFP and US

Review question: what is the diagnostic accuracy of the combination of alpha-foetoprotein (AFP) and abdominal ultrasound (US) compared to US for the diagnosis of hepatocellular carcinoma (HCC) in adults with chronic liver disease?
Population: adults with chronic liver disease

Setting: clinical setting (secondary or tertiary care setting) or surveillance programs

Study design: prospective and retrospective cross-sectional studies

Index tests:abdominal ultrasound; combination of serum alpha-foetoprotein (AFP) measurement with a cut-off value of 20 ng/mL and abdominal ultrasound

Target condition: HCC of any size, any stage

Reference standards:the pathology of the explanted liver in case of transplantation;the histology of resected focal liver lesion(s), or the histology of resected or biopsied focal liver lesion(s) with a follow-up period of at least six months to exclude the presence of focal lesions non detected by the index test and synchronous lesions from the parenchyma surrounding the resected or biopsied area;typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months in order to allow the confirmation of an initial negative result on computer tomography (CT) or on magnetic resonance imaging (MRI).

Limitations in the evidence

Risk of bias/ Applicability

- Participant selection: high/unclear risk of bias in 2 studies (33%)/ high concern 2 studies (33%)

- Index tests: high/unclear risk of bias in 2 studies (33%)/ high concern no study

- Reference standard: high/unclear risk of bias in 4 studies (67%)/ high concern 1 study (17%)

- Flow and timing: high risk of bias in 6 studies (100%)

Findings

						Implications in a hypothetical cohort of 1000 people					
Index test	**Number of studies (participants)**	**Sensitivity** (95% CI)	**Relative sensitivity (95% CI)** **P value**	**Specificity** **(95% CI)**	**Relative specificity** **(95% CI)** **P value**	**Assumed prevalence of hepatocellular carcinoma (HCC)[a]** **%**	**True positives** will receive appropriately further necessary testing with CT or MRI, or contrast enhanced ultrasound (CEUS) and possibly treatment .	**False negatives** will be misdiagnosed and not receive appropriate treatment.	**True negatives** will not appropriately undergo unnecessary further testing with CT, MRI, CEUS, biopsy	**False positives** will inappropriately undergo further unnecessary testing with CT, MRI, CEUS biopsy.	**Certainty of the evidence**
US	6 (5044)	76% (56% to 89%)	1.28 (1.03 to 1.539	93% (80% to 96%)	0.94, (0.87 to 1.01)	5%	38	12	883	67	low[b] ⊕⊕◯◯

		P = 0.014		P = 0.102	30%	228	72	651	49
Combination of AFP (cut-off 20 ng/mL) and US	96% (88% to 98%)		85% (73% to 82%)		5%	48	2	807	143
					30%	288	12	595	105

[a] We chose for exemplification two values of HCC prevalence: 5% for a population at low risk (compensated advanced chronic liver disease and chronic viral hepatitis) Lok 2009 and 30% for a population with high risk, a median of the prevalence in the included cross-sectional studies conducted in clinical cohorts.
[b]Downgraded by two levels: risk of bias, indirectness. Risk of bias downgraded one level because all studies were judged at high risk of bias; indirectness downgraded one level as we considered most studies to have concern regarding applicability mainly in relation to the population (including disease spectrum)
GRADE certainty of the evidence
High: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.
The results presented in this table should not be interpreted in isolation from results of the individual included studies contributing to each summary test accuracy measure.

BACKGROUND

Hepatocellular carcinoma (HCC) is the most common primary liver neoplasm, usually developing in the setting of chronic liver disease. It is the sixth most commonly diagnosed cancer and the fourth leading cause of death from cancer worldwide; there were 782,000 deaths due to HCC in 2018 (Bray 2018). In men, HCC ranks fifth in terms of global cases of cancer and second in terms of cancer deaths (Bray 2018). In Western countries, the incidence and mortality rates of HCC increased substantially between 1990 and 2015 (Ryerson 2016; GBD 2017). Most common risk factors include liver cirrhosis, severe liver fibrosis, hepatitis B, hepatitis C, alcohol intake, and non-alcoholic fatty liver disease (Yang 2011), although some people may develop HCC without the presence of known risk factors (Bralet 2000; Young 2012).

Clinically, HCC is frequently diagnosed in the late stages because of the absence of specific symptoms of the malignancy, other than those related to chronic liver disease. Only 20% of patients with HCC are eligible for curative treatments — such as liver resection, transplantation, or ablation — due to advanced tumour stage, liver dysfunction, or shortage of liver donors (Davila 2012). According to the current guidelines, HCC can only be considered as resectable and amenable to surgical radical resection if the cancer presents as either a single lesion with a maximum diameter of less than 5 cm, or up to three lesions, each with a maximum diameter of 3 cm (Mazzaferro 1996; EASL-EORTC 2012; Omata 2017; EASL 2018; Heimbach 2018). Furthermore, curative treatment options are not feasible for most patients due to severe clinical deterioration at the moment of diagnosis, or due to the inaccuracy of the preoperative clinical evaluation and staging procedure.

Despite the poor initial prognosis (the mortality-to-incidence overall ratio has been reported as 0.93; (Bray 2018)), a five-year survival rate of more than 50% can be achieved if HCC is detected at an early stage (Forner 2012). According to the Barcelona Clinic Liver Cancer staging system, only patients with early-stage HCC are eligible for curative treatment (Llovet 1999). Therefore, it is very important to make an accurate diagnosis of HCC as early as possible.

Abdominal ultrasound (US) has become an acceptable imaging modality in detecting HCC because it is non-invasive, acceptable to patients, has moderate costs, and no associated risks. A recent meta-analysis showed a pooled sensitivity of 84% of US surveillance in detecting HCC in people without any symptoms (Tzartzeva 2018). However, the same publication showed a poor result for US in the detection of early-stage HCC in people who are eligible for curative therapies, with a pooled sensitivity of only 47% (Tzartzeva 2018). Accordingly, detection of HCC poses a challenge. The sonographic liver tissue characteristics in people with fibrosis make it particularly difficult to detect and differentiate small neoplastic nodules from the surrounding parenchyma and from regenerative nodules. Furthermore, the performance of US can be influenced by the expertise of the operator and the quality of the equipment.

Alpha-foetoprotein (AFP) is a tumour marker which has been used as a diagnostic test for HCC since the 1970s, when most patients were diagnosed in the late stage and with clinical symptoms (Kew 1975). Although the test for AFP is widely available, inexpensive, and easy to perform, it has poor accuracy as a serological test for the early detection of HCC (Tateishi 2008). Levels of AFP increase not only in people with HCC, but also in people with active hepatitis, cirrhosis without HCC, or exacerbation of the underlying liver disease, due to pathophysiological changes of inflammation and regeneration; this means the test can have low specificity in the population at risk (Di Bisceglie 2005; Gopal 2014).

Surveillance programmes for early detection of HCC in high-risk patients have been implemented in the current medical practice in most Western and Asian-Pacific countries, despite the very low-certainty evidence regarding the effects on mortality (Kansagara 2014; Singal 2014). The American Association for the Study of Liver Disease (AASLD), European Association for the Study of the Liver with European Organization for Research and Treatment of Cancer (EASL-EORTC), and Asian Pacific Association for the Study of the Liver (APASL) recommend abdominal US as an imaging modality for surveillance of HCC every six months in people at risk. However, disagreement exists between using serum biomarker AFP as an additional test (EASL-EORTC 2012; Omata 2017; EASL 2018; Heimbach 2018).

There are several published systematic reviews which examine the accuracy of ultrasonography and AFP in detecting HCC (Colli 2006; Tateishi 2008; Singal 2009; Kansagara 2014; Singal 2014; Chou 2015; Tzartzeva 2018), but to our knowledge, there is no recent systematic review which compares AFP alone, US alone, and the combination of AFP and US in detecting HCC. Therefore, the aim of our review is to use Cochrane methodology to assess the diagnostic accuracy of these three modalities for the diagnosis of HCC, as well as the early stage of HCC (when the cancer may still be resectable), in people with chronic liver disease.

Target condition being diagnosed

Hepatocellular carcinoma is the most common primary liver cancer which occurs mostly in people with chronic liver disease. The incidence of HCC increases in individuals with hepatitis B and C, alcohol use, and non-alcoholic fatty liver disease, and in those with liver cirrhosis of various aetiologies (Bruix 2011). There is no definite threshold in the definition of lesion size, although the literature tends to classify lesions with a diameter equal to or less than 2 cm as 'small' (Hussain 2002; Choi 2014; Park 2017).

In clinical practice, and according to pertinent guidelines, multiphasic computed tomography (CT) or magnetic resonance imaging (MRI) with intravascular contrast allow for a highly accurate diagnosis of HCC, without an invasive biopsy (EASL 2018; Heimbach 2018). The diagnosis of HCC is usually obtained on the basis of cross-sectional CT or MRI features: focal liver lesions which show non-rim-like hyper enhancement in the arterial phase, subsequent non-peripheral washout appearance, and capsule appearance (LI-RADS 2018). Liver histology is required only for undefined lesions during CT and MRI (EASL-EORTC 2012; Omata 2017; Heimbach 2018).

A number of staging systems for HCC have been proposed and developed; however, there is no globally applicable staging system (Kinoshita 2015). Among different staging protocols, the Barcelona Clinic Liver Cancer (BCLC) classification system has a notable feature of treatment recommendations for each stage, based on the best treatment options currently available (Llovet 1999; Llovet 2003; Llovet 2008). The staging is based on four elements: tumour extension, liver functional reserve, physical status, and cancer-related symptoms. According to the BCLC classification

system, only patients with early-stage HCC are eligible for curative treatment, such as surgical resection or percutaneous treatment. Orthotopic liver transplantation is reserved for patients with decompensated cirrhosis.

Orthotopic liver transplantation is considered a definite curative treatment for HCC. When orthotopic liver transplantation for HCC was initially introduced in the 1980s, it was associated with poor five-year survival rates and high recurrence rates, which led to the treatment being contraindicated for HCC (Yokoyama 1990). In 1996, specific criteria, known as Milan criteria (Mazzaferro 1996), were developed for the selection of patients for liver transplantation. With the implementation of these criteria, the overall five-year survival rates for post-orthotopic liver transplantation patients exceeded 70% (Mazzaferro 2011). The criteria for patients eligible for orthotopic liver transplantation include: a single HCC lesion with a diameter equal to or less than 5 cm, or up to three HCC lesions, each with a diameter equal to or less than 3 cm; no vascular invasion; and no extrahepatic involvement (no metastasis). The same criteria are recommended for the selection of patients eligible for surgical resection.

Along with interferon-based treatment, a new direct-acting antiviral (DAA) therapy was developed for people with chronic hepatitis C; these therapies therefore acted against one of the major risk factors for developing HCC (Bourliere 2015; Charlton 2015; Leroy 2016). DAA therapy allowed the achievement of sustained virologic response (SVR) in more than 70% of patients, compared to less than 40% with interferon therapy (Jakobsen 2017; Calvaruso 2018). However, a consensus exists that even after achieving SVR, people with chronic hepatitis C should be surveyed closely, especially those with advanced fibrosis and those who received a recent treatment for HCC in order to detect HCC at an early stage (Butt 2018).

Index test(s)

Abdominal US is a safe, inexpensive, non-invasive, and real-time diagnostic technique with relatively low costs. A transducer transforms electrical energy into sound waves (two megahertz (mHz) to eight mHz) and transmits them into the body. Simultaneously, the transducer detects the sound waves reflected by the underlying tissue. The intensity of these reflected (echo) waves is based on several properties of the tissue, such as density, depth, and properties of adjacent tissues. The echo waves are converted into electrical energy and displayed as a cross-sectional tomography image.

According to the Liver Reporting and Data System (LI-RADS) for detection of HCC, there are three US categories for diagnosing suspected liver lesions: US-1 (negative), US-2 (subthreshold), and US-3 (positive). Since US is an operator-dependent imaging modality and limitations due to patient characteristics can occur, an US visualisation score is added: A (no or minimal limitations); B (moderate limitations); and C (severe limitations). A negative observation is reported when no liver lesions have been detected or the detected lesions are definitely benign. Subthreshold lesions of less than 10 mm are noted only when no definitely benign features have been observed. A positive observation is reported when a lesion of more than 10 mm with no definitely benign features is observed, or a new venous thrombus has been detected (LI-RADS 2018; Rodgers 2019).

Alpha-foetoprotein (AFP) is a glycoprotein of 591 amino acids and a carbohydrate moiety which is assessed in serum by enzyme immunoassays (Pucci 1991). In presence of HCC, high serum values of AFP are reported with variable accuracy (Colli 2006; Tateishi 2008; Singal 2009; Kansagara 2014; Singal 2014; Tzartzeva 2018).

Clinical pathway

For people with chronic liver disease, a surveillance programme is usually recommended. There are minimal variations among the surveillance programmes of the different scientific societies (Table 1).

American Association for the Study of Liver Disease (AASLD) guidelines

According to the AASLD guidelines, to increase overall survival, only adults with cirrhosis who are considered at risk of developing HCC need surveillance. It is suggested that surveillance be performed using abdominal US, with or without AFP, every six months. However, it is not possible to determine which type of surveillance test (ultrasound alone or ultrasound plus AFP) would lead to a greater improvement in survival. Surveillance is not suggested for those with Child-Pugh class C cirrhosis, unless they are on the liver transplant waiting list, because of low anticipated survival (Heimbach 2018).

European Association for the Study of the Liver with European Organization for Research and Treatment of Cancer (EASL-EORTC) guidelines

According to the EASL-EORTC guidelines, people at risk of developing HCC for which surveillance should be performed include: people with Child-Pugh stage A or stage B cirrhosis, people with Child-Pugh stage C cirrhosis awaiting liver transplantation, non-cirrhotic hepatitis B virus carriers with active hepatitis or family history of HCC, and people with chronic hepatitis C in the absence of cirrhosis but with advanced liver fibrosis stage 3 (F3). People on liver transplant waiting lists should be screened for HCC in order to detect and manage tumour progression. Surveillance should be performed using abdominal US every six months. A three- to four-month interval is recommended in people where a nodule of less than 1 cm has been detected, and in the follow-up strategy, after resection or loco-regional therapies. Serum biomarkers such as AFP, AFP-L3 (third electrophoretic form of lentil lectin-reactive AFP), and des-gamma-carboxy prothrombin are suboptimal for routine clinical practice, and therefore, not recommended for screening (EASL-EORTC 2012; EASL 2018).

Asian Pacific Association for the Study of the Liver (APASL) guidelines

According to the APASL guidelines, the following people are at risk of HCC development and therefore are eligible for HCC screening: those with cirrhosis, those who have chronic hepatitis B virus infection with cirrhosis, and those who have chronic hepatitis B virus infection in the absence of cirrhosis. The optimal surveillance strategy includes abdominal US with serum AFP measurement every six months. Measurement of AFP alone is not recommended for routine surveillance of people with HCC (Omata 2017).

Outside surveillance programmes

Ultrasound and AFP are usually performed in people with clinically suspected HCC, or liver cirrhosis, or both, or at the moment

of decompensation of chronic liver disease, or all these factors together.

Prior test(s)

The diagnosis of liver cirrhosis is usually based on clinical judgement derived from history, laboratory testing, physical examination, imaging, liver stiffness measurement, liver histology, or a combination of these. Due to the accuracy of non-invasive tests, liver histology is reserved to only a minority of patients with unclear diagnosis, and a non-invasive diagnosis of advanced chronic liver disease is considered equivalent to a histological diagnosis of cirrhosis (de Franchis 2015). No test is recommended by the above guidelines, prior to a surveillance programme for HCC detection.

Role of index test(s)

Abdominal US and AFP (independently, or in combination, or in sequence) are used as triage tests to exclude the presence of focal liver lesions suspected of being HCC. Further alternative testing is required to confirm the diagnosis as well as staging.

Alternative test(s)

Contrast-enhanced ultrasound (CEUS) is an advanced form of US examination in which images are acquired using intravenously injected microbubble contrast agent with optimised technology required for contrast visualisation. The CEUS exam consists of a 'bolus' administration of contrast media through a superficial peripheral vein. The sequence of blood entering the liver is first arterial (10 seconds to 40 seconds), then portal (40 seconds to 120 seconds after injection), and then late venous (more than 120 seconds). This vascular discrimination, similar to that obtained by contrast CT or MRI, allows for the collection of information regarding the circulatory system of a tumour (e.g. types of feeding vessels, tumour circulatory volume). Positivity criteria for HCC are based on arterial hyper enhancement and subsequent washout appearance. The advantages of US agent over CT and MRI agents include no adverse reactions, possible multiple injections of contrast in the same examination, safety, practicality, no risk of nephrotoxicity, and no ionising radiation (Chung 2015).

Contrast-enhanced multiphasic multi detector CT and contrast-enhanced MRI have been established as relevant non-invasive modalities for detection and evaluation of liver lesions (Lee 2012; O'Neill 2015). The ability to detect HCC rests on characterising the enhancement patterns in arterial, portal venous and subsequent phases relative to the surrounding liver tissue. The differences in blood flow and extracellular volume between HCC and normal liver tissue lead to main radiological hallmarks such as non-rim-like arterial phase hyper enhancement and subsequent non-peripheral washout with enhancing capsule in later phases (Hennedige 2012; Choi 2014; Shah 2014; LI-RADS 2018). CT is a commonly used modality for diagnosing HCC due to its short acquisition time and high spatial resolution. However, MRI offers several beneficial features such as absence of X-ray radiation and combination of various sequences (multiphasic T1- and T2-weighted sequences, diffusion-weighted imaging, and apparent diffusion coefficient) in combination with the use of extracellular or hepatocellular gadolinium-based contrast agent, or both (Arif-Tiwari 2014; Roberts 2018).

Apart from AFP, there are other potential serological tumour biomarkers for the detection of HCC. Des-gamma-carboxyprothrombin, also known as prothrombin induced by vitamin K absence-II (PIVKA-II), is an abnormal prothrombin protein that is increased in the serum of people with HCC. It is recognised as a specific marker for the detection and prognosis of HCC (Imamura 1999; Koike 2001), although contrary data exist on the benefit of using PIVKA-II over AFP (Nakamura 2006; Li 2014). AFP-L3 can differentiate an increase in AFP due to HCC from that in people with benign liver disease, and from a potential biomarker for early HCC detection (Kumada 2014). Glypican-3 (GPC3) is considered to be a promising biomarker for early detection of HCC and a potential epitope for HCC-targeted therapies (Zhou 2018). Other biomarkers include Golgi protein 73, osteopontin, circulating free DNA, and microRNAs. However, none of these have been introduced in daily practice (Omata 2017).

Rationale

Hepatocellular carcinoma is currently detected by liver ultrasound in people with chronic liver disease with normal or high AFP levels during surveillance programmes. Following ultrasound, the diagnosis is usually confirmed by high levels of AFP or by using contrast-enhanced ultrasound (CEUS) (or both), CT, or MRI. The diagnosis in people who are not in a surveillance programme is usually obtained at decompensation of chronic liver disease (i.e. detection of oesophageal varices, gastrointestinal haemorrhage, or ascites), or during the diagnosis of previously unrecognised chronic liver disease. In such patients, liver ultrasound or AFP (or both) are also the first test(s) of choice and, if positive, further testing is required with CEUS, CT, or MRI.

There is no clear evidence on the benefit of surveillance programmes in terms of overall survival: the conflicting results could be a consequence of inaccurate detection, ineffective treatment, or both. Assessing the diagnostic accuracy of abdominal US and AFP serum concentration may clarify whether the absence of benefit in surveillance programs might be related to under-diagnosis. Furthermore, an assessment of the accuracy of these two tests for diagnosing HCC is needed for either ruling out, diagnosing, or supporting further testing in people with chronic liver disease who are not included in surveillance programs.

People with previous diagnoses of, and who had previous treatments for, HCC make up a distinct group. The diagnostic accuracy for the recurrence of HCC after surgical or any other type of treatment is not the focus of this review.

This review represents the first part of a planned overall evaluation of diagnostic performances of the most commonly used modalities for diagnosing HCC in people with chronic liver disease. The present systematic review will assess the diagnostic accuracy of ultrasound and AFP serum concentration for the diagnosis of HCC. Another systematic review will focus on the diagnostic accuracy of CEUS in characterising suspected lesions as HCC as a second-line diagnostic modality (Fraquelli 2019), and a third systematic review will focus on the assessment of CT as another second- or third-line imaging modality (if CEUS was used as second-line test) in assessing focal liver lesions detected on ultrasound (Nadarevic 2019). A review assessing the accuracy of MRI for diagnosing HCC is also in progress (Nadarevic 2020). We are planning to produce an overview of the systematic reviews that assess abdominal US and AFP, CEUS, CT, and MRI for the diagnosis of HCC.

OBJECTIVES

To assess the diagnostic accuracy of abdominal ultrasound (US) and alpha-foetoprotein (AFP), alone or in combination, for the diagnosis of hepatocellular carcinoma (HCC) of any size and at any stage in adults with chronic liver disease, either in a surveillance programme or in a clinical setting.

Secondary objectives

- To assess the diagnostic accuracy of abdominal US and AFP, alone or in combination, for the diagnosis of resectable HCC in people with chronic liver disease, either in a surveillance programme or in a clinical setting. The definition of resectable HCC is a neoplasm amenable to surgical radical resection according to the current guidelines (EASL-EORTC 2012; Omata 2017; EASL 2018; Heimbach 2018), that is, a single lesion with a maximum diameter of less than 5 cm, or fewer than three lesions with a maximum diameter of 3 cm.
- To compare the diagnostic accuracy of individual tests versus the combination of both tests.
- To investigate the following predefined sources of heterogeneity:
  - study design (prospective compared to retrospective; case-control studies compared to cross-sectional cohort studies);
  - study date (studies published before the year 2000 compared to studies published after the year 2000, due to advancements in technology and changes in diagnostic criteria);
  - inclusion of participants without cirrhosis (studies including more than 10% participants without cirrhosis compared to studies including less than 10% participants without cirrhosis);
  - study location (population differences): studies conducted in North and South America compared to Europe compared to Asia and Africa;
  - prevalence of the target condition (studies with HCC prevalence more than 10% compared to studies with HCC prevalence less than 10%);
  - participant selection (participants recruited from planned surveillance programs compared to clinical cohorts);
  - different HCC stage (studies with more than 20% of participants with resectable HCC compared to studies with less than 20% of participants with resectable HCC);
  - different reference standard (histology of the explanted liver compared to liver biopsy compared to another reference standard);
  - different liver cirrhosis aetiology: studies with more than 80% participants with viral (hepatitis C or hepatitis B) chronic liver disease compared to studies with less than 80% of participants with viral chronic hepatitis;
  - different severity of the underlying chronic liver disease: studies with more than 50% of participants with MELD (model for end-stage liver disease) score less than 15 or with Child Pugh score A compared to studies with less than 50% of participants with MELD less than 15 or Child Pugh score A.

METHODS

Criteria for considering studies for this review

Types of studies

We aimed to include studies, irrespective of publication status and language, that have evaluated the diagnostic accuracy of abdominal ultrasound (US) and alpha-foetoprotein (AFP), independently or in combination, for the diagnosis of: hepatocellular carcinoma (HCC) in people with chronic liver disease. These studies should have used one of the acceptable reference standards (see below Reference standards).

We considered for inclusion studies of cross-sectional design including participants with clinical suspicion of HCC or cohort studies including high-risk participants in a surveillance programme, as well as studies with a case-control design that compared people with known HCC to a matched control (participants with chronic liver disease without evidence of HCC). We excluded studies that analysed data only per lesion, that is, those that considered the number of lesions rather than participants, unless participant data were made available by study authors.

Participants

Eligibility criteria

We included study participants aged 18 years and older, of any sex, who are diagnosed with a chronic liver disease, irrespective of the severity and duration of the disease. Study participants should have been treatment-naive for HCC when enrolled in the respective study.

Exclusion criteria

We excluded studies which had included participants treated for HCC unless they represented less than 5% of all the included participants, or if data were presented in such a way as to allow this group of participants to be isolated from the remaining included participants.

Index tests

We included abdominal US alone, AFP alone, and a combination of abdominal US and AFP for the detection of HCC in adults with chronic liver disease. For AFP, different cut-off values were used, ranging from 7 mg/mL to 400 mg/mL. For ultrasound (US), positive criteria include the minimum diameter of a detectable lesion and exclusion of benign criteria.

Target conditions

- Hepatocellular carcinoma of any size and at any stage.
- Resectable hepatocellular carcinoma (see Secondary objectives).

Reference standards

We accepted as a reference standard for the diagnosis of HCC one of the following.

- The pathology of the explanted liver in case of transplantation.
- The histology of resected focal liver lesion(s), or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months to exclude the presence of focal lesions not detected

by the index test and synchronous lesions from the parenchyma surrounding the resected or biopsied area.

- Typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months in order to allow the confirmation of an initial negative result on CT or on MRI.

We acknowledge that all these reference standards, even if commonly used in clinical practice, are not perfect. The pathology of the explanted liver is possible only in the case when all the included patients undergo liver transplantation; therefore, the setting does not correspond to the clinical question as only people with advanced and decompensated liver disease are candidates for orthotopic liver transplantation. In the case of histology of resected focal lesion, histology of biopsied liver lesions, CT or MRI examination, the negative result can be confirmed only with an adequate follow-up period. This would introduce an unavoidable differential verification bias. In addition, CT and MRI cannot be considered completely accurate.

Search methods for identification of studies

Electronic searches

We searched the Cochrane Hepato-Biliary Group (CHBG) Controlled Trials Register and the Cochrane Hepato-Biliary Group Diagnostic-Test-Accuracy Studies Register (both maintained and searched internally by the CHBG Information Specialist via the Cochrane Register of Studies Web; June 2020), the Cochrane Library (2020, Issue 6), MEDLINE Ovid (1946 to June 2020), Embase Ovid (1974 to June 2020), LILACS (Bireme; 1982 to June 2020), Science Citation Index Expanded (Web of Science; 1900 to June 2020), and Conference Proceedings Citation Index – Science (Web of Science; 1990 to June 2020; (Royle 2003)). Appendix 1 gives the search strategies with the time spans of the searches.

We applied no language or document type restrictions.

Searching other resources

We attempted to identify additional references by manually searching articles retrieved from digital databases and relevant review articles. We sought information on unpublished studies by contacting experts in the field. In addition, we handsearched abstract books from meetings of the American Association for the Study of Liver Diseases (AASLD), the European Association for the Study of the Liver (EASL), and Asia-Paciifc Association for the study of the Liver (APASL), held over the past 10 years. We also searched for other kinds of grey literature in the System for Information on Grey Literature in Europe "OpenGrey" (www.opengrey.eu/).

Data collection and analysis

We followed available guidance as provided in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (DTA Handbook 2013).

Selection of studies

Two review authors (AC and MF) independently scrutinised half of the titles and abstracts identified by electronic literature searching to identify potentially eligible studies, and two other review authors (TN and VG) independently scrutinised the other half. We recorded any citation, identified by one of the four review authors, as potentially eligible for full-text review. Then, two review authors (AC and TN) independently reviewed publications for eligibility. To determine eligibility, we assessed each publication to determine whether participants met the inclusion criteria detailed above. We included abstracts only if they provided sufficient data for analysis. We resolved disagreements by consensus.

Data extraction and management

We developed a standardised data extraction form and piloted the form on nine of the included studies. Based on the pilot, we finalised the form.
Then, two review authors (AC and TN) completed the data extraction form for each included study. Each review author independently retrieved study data. In cases of disagreement, we reached consensus through discussion with a third review author (GC).

We retrieved the following data.

- General information: title, journal, year, publication status, and study design (prospective versus retrospective), surveillance program or clinical cohorts.
- Sample size: number of participants meeting the criteria and total number of participants screened.
- Baseline characteristics: baseline diagnosis, age, sex, and presence of cirrhosis and mean diameter of HCC.
- Index tests with predefined positivity criteria and when appropriate all cut-off values.
- Target condition.
- Order of tests.
- Time between tests.
- Reference standard tests.
- Numbers of true-positive, true-negative, false-positive, false-negative, and uninterpretable index test results. We extracted these data for each presented cut-off value and for either HCC of any size, stage, and resectable HCC.

We summarised the data from each study in 2 × 2 tables (true positive, false positive, false negative, true negative), according to the index tests considered, and we entered the data into Review Manager 5.4 software (Review Manager 2020).

Missing data

We contacted primary authors by email to request missing data: number of AFP false-positive results (Baig 2009; Chen 2010; Abdelgawad 2013; El-Emshaty 2014; Dengler 2017), and results of per patients analyses as only per lesions were reported in Lim 2006. We received no reply and sent a second email after two weeks. No reply was received; therefore, we excluded the above-mentioned studies.

Assessment of methodological quality

Two review authors (AC and TN) independently assessed the risk of bias of included studies and applicability of their results using QUADAS-2 (revised tool for quality assessment of diagnostic accuracy studies; (Whiting 2011)). In cases of disagreement, we reached consensus through discussion. We addressed aspects of study quality involving the participant spectrum, index tests, target conditions, reference standards, and flow and timing. For studies that assessed ultrasound as the index test, the visualisation of the liver can often be sub optimal due to patient characteristics;

therefore, lack of reporting or exclusion of uninterpretable results from analyses could overestimate the accuracy of ultrasound. We considered the study to be at high risk of bias if uninterpretable results were excluded from the analysis. We classified a study at high risk of bias if at least one of the domains of QUADAS-2 was judged as being at high or unclear risk of bias (Appendix 2).

Statistical analysis and data synthesis

We provided a description of the included studies by calculating median values and interquartile ranges (IQR) across studies for some characteristics of our interest, defined at study level. In particular, we considered HCC mean diameter and the prevalence of participants with the following characteristics: HCC, Child-Pugh class A, liver cirrhosis, viral aetiology of cirrhosis, and resectable HCC.

We carried out statistical analyses according to recommendations provided in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (DTA Handbook 2013). We designed 2 × 2 tables (see Data extraction and management) for each primary study for the two index tests and for their combination. We planned the following strategy of analyses.

Alpha-foetoprotein

Alpha-foetoprotein (AFP) was considered positive when higher than a defined cut-off (threshold) value was noted (Colli 2006; Marrero 2009; Lok 2010). Firstly, we performed a graphical descriptive analysis of the included studies. We presented forest plots (sensitivity and specificity separately, with their 95% confidence intervals (CIs)), and we provided a graphical representation of the studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Secondly, we performed a meta-analysis. In the case that primary studies reported accuracy estimates of AFP using different cut-off values, we used the hierarchical summary ROC model (HSROC) in order to pool data (sensitivities and specificities) and to estimate a summary ROC (SROC) curve (Rutter 2001). When considering studies with a common cut-off value, we used the bivariate model, and we provided estimates of summary sensitivity and specificity. We used the pooled estimates obtained from the fitted models to calculate summary estimates of positive and negative likelihood ratios (LR + and LR-, respectively). For primary studies reporting accuracy results for more than one cut-off value, we reported sensitivities and specificities for all cut-off values, but we used a single cut-off value for each study in HSROC or bivariate analysis. The most common cut-off values were expected to be 10, 20, 200, or 400 nanograms per millilitre (ng/mL).

Abdominal ultrasound

Abdominal ultrasound (US) was considered positive when a lesion of more than 10 mm with no definitely benign features was observed, or a new venous thrombus was detected according to defined criteria (LI-RADS 2018). Subthreshold lesions of less than 10 mm were noted only when no definitely benign features were observed (LI-RADS 2018). Firstly, we performed a graphical descriptive analysis of the included studies. We presented forest plots (sensitivity and specificity separately, with their 95% CIs), and we provided a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Secondly, we performed a meta-analysis using the bivariate model, and we provided estimates of summary sensitivity and specificity. We used the pooled estimates obtained from the fitted models to calculate summary estimates of positive and negative likelihood ratios (LR+ and LR-, respectively).

Uninterpretable index test results

In case of uninterpretable index test results (especially relevant for US), we performed a further analysis according to the intention-to-diagnose (ITD) principle (Schuetz 2012). We classified participants with uninterpretable results as false-positive if they had a negative reference standard, or false-negative result on a positive reference standard.

Combination of abdominal ultrasound and alpha-foetoprotein

The index test obtained by the combination of US and AFP tests is considered positive when at least one of the two tests is positive. Firstly, we performed a graphical descriptive analysis of the included studies. We presented forest plot results (sensitivity and specificity separately, with their 95% CIs), and we provided a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Secondly, we performed a meta-analysis. In the case that primary studies reported accuracy estimates of the combination of tests using different cut-off values for AFP, we used the hierarchical summary ROC model (HSROC) to pool data (sensitivities and specificities) and to estimate a summary ROC (SROC) curve (Rutter 2001). When considering studies with a common cut-off value, we used the bivariate model and provided estimates of summary sensitivity and specificity. We used the pooled estimates obtained from the fitted models to calculate summary estimates of positive and negative likelihood ratios (LR+ and LR). For primary studies reporting accuracy results for more than one cut-off value, we reported sensitivities and specificities for all cut-off values, but we used a single cut-off value for each study in HSROC or bivariate analysis.

Comparisons

The combination of the two tests, US and AFP, was considered positive when at least one of the two tests was positive. We made pair-wise comparisons between individual tests, and between individual tests and the index test obtained by the combination of the two tests when both tests are used, by adding a covariate for the index test to the bivariate model. We assessed the significance of differences in test accuracy by using the log-likelihood ratio test for comparison of models with and without the index test covariate term. We included separate variance terms for sensitivity and specificity in the bivariate model for the two tests in comparison. We performed both indirect and direct comparisons when sufficient data were available. We calculated relative sensitivity (i.e. ratio between the sensitivities of the two index tests) and relative specificity (i.e. ratio between the two specificities).

We considered two-sided P values less than 0.05, as statistically significant. We performed all statistical analyses using SAS statistical software, release 9.4 (SAS Institute Inc., Cary, NC, USA) and macro METADAS (DTA Handbook 2013).

Investigations of heterogeneity

We investigated the effects of the following predefined sources of heterogeneity.

- Study design (case-control compared to cross-sectional studies, prospective compared to retrospective).
- Study date (studies before compared to after the year 2000 due to advancements in technology and change in diagnostic criteria).
- Inclusion of participants without cirrhosis (studies including more than 10% participants without cirrhosis compared to studies including less than 10% participants without cirrhosis).
- Study location (population differences): studies conducted in the USA compared to Europe compared to Asia and Africa.
- Prevalence of the target condition (studies with HCC prevalence of more than 10% compared to studies with HCC prevalence of less than 10%).
- Participant selection (participants recruited from planned surveillance programs compared to clinical cohorts).
- Different HCC stage (studies with more than 20% of participants with resectable HCC compared to studies with less than 20% of participants with resectable HCC).
- Different reference standard (histology of the explanted liver compared to liver biopsy compared to another reference standard).
- Different liver cirrhosis aetiology: studies with more than 80% participants with viral (hepatitis C or hepatitis B) chronic liver disease compared to studies with less than 80% of participants with viral chronic hepatitis.
- Different severity of the underlying chronic liver disease: studies with more than 50% of participants with MELD (model for end-stage liver disease) score less than 15 or with Child Pugh score A compared to studies with less than 50% of participants with MELD less than 15 or Child Pugh score A.

We estimated the above effects by adding covariates to the bivariate models. We assessed the statistical significance of the covariate effect by using the log-likelihood ratio test for comparison of models with and without the covariate term.

Sensitivity analyses

We assessed the effects of risk of bias of the included studies on diagnostic accuracy by performing a sensitivity analysis in which we exclude studies classified as having high or unclear risk of bias in at least one of the domains of QUADAS-2 (Appendix 2). In addition, we defined the following signalling questions as most relevant, and planned to conduct a sensitivity analyses in which we excluded studies with answers of 'no' or 'unclear'.

- “Was a case-control design avoided?” (i.e. was the study design clearly cross-sectional including a series of participants at risk of with a clinical suspicion of HCC?)
- For studies using AFP as index test: “if a threshold was used, was it pre-specified?”; or for ultrasound as index test: “were the positivity criteria defined?”.
- "Were all participants included in the analysis and analysed according to ITD principle (non-evaluable results considered as false)?”

We did not perform the planned analysis excluding studies using AFP without a pre-specified threshold as we chose to analyse the results of studies using the two most common cut-off values of 20 ng/mL and 200 ng/mL. We did not perform the planned analysis excluding studies not reporting results obtained with ITD principle for uninterpretable results due to lack of data because only two studies reported the number of uninterpretable results.

We also conducted, as planned, a sensitivity analysis in which studies published only in abstract or letter form are excluded.

Assessment of reporting bias

In order to reduce reporting bias, we did not plan to use a filter search strategy nor to implement any language or sample limitations. We did not plan to test for publication bias due to the lack of validated methods for diagnostic test accuracy reviews.

'Summary of findings' table

We prepared 'Summary of findings' tables to present the main results and key information regarding the certainty of evidence, We assessed the certainty of evidence as recommended using the GRADE approach (Schünemann 2008; Balshem 2011; Schünemann 2016; GRADEpro GDT). We rated the certainty of evidence as either high (when not downgraded), moderate (when downgraded by one level), low (when downgraded by two levels), or very low (when downgraded by more than two levels) based on five domains: risk of bias, indirectness, inconsistency, imprecision, and publication bias. For each outcome, the certainty of evidence started as high when there were high-quality observational studies (cross-sectional or cohort studies) that enrolled participants with diagnostic uncertainty. If we found a reason for downgrading, we used our judgement to classify the reason as either serious (downgraded by one level) or very serious (downgraded by two levels; (Schünemann 2020a; Schünemann 2020b)).

Five authors (AC, TN, MF, VG, and GC) discussed judgments and applied GRADE In the following way.

- Risk of bias: we used QUADAS-2 to assess risk of bias
- Indirectness: we assessed indirectness in relation to the population (including disease spectrum), setting, interventions, and outcomes (accuracy measures). We also used prevalence as a guide to whether there was indirectness in the population.
- Inconsistency: we carried out prespecified analyses to investigate potential sources of heterogeneity and downgraded when we could not explain inconsistency in the accuracy estimates
- Imprecision: we looked at the confidence intervals of sensitivity and specificity estimates and at the unexplained heterogeneity of the results
- Publication bias: we did not evaluate publication bias due to the lack of validated methods for diagnostic test accuracy reviews

RESULTS

Results of the search

We ran the search on 5 June 2020. We identified 45,837 records by searching the Cochrane Hepato-Biliary Group Controlled Trials Register (n = 31), the Cochrane Hepato-Biliary Group Diagnostic Test Accuracy Register (n = 3), the Cochrane Library (n = 958), MEDLINE Ovid (n = 12,856), Embase Ovid (n = 22,264), LILACS (n = 351), and Science Citation Index Expanded and Conference Proceedings Citation Index – Science (both Web of Science) (n = 9374). We retrieved seven additional records through handsearching. After exclusion of 11,347 duplicates, 34,497 records remained for possible eligibility. After reading the title and the

abstract of these records, we excluded 33,932 of them, as they did not meet the inclusion criteria. We retrieved full texts of the remaining 565 records, and after reading the full texts, we excluded 219 studies for various reasons (Figure 1; Characteristics of excluded studies). In particular, we excluded 109 studies not reporting data or reported only incomplete data on the accuracy of the index tests, 54 studies comparing participants with hepatocellular carcinoma (HCC) with healthy participants or including healthy participants in the control arm, and not reporting the results of the comparison of participants with HCC and participants with chronic liver disease, 31 reporting no original data on the index tests, 10 studies including participants with treated HCC and suspected recurrences, seven studies reporting only per lesion analyses, seven studies not conducted in people with chronic liver disease, and one study (Heyward 1985) reporting preliminary data fully reported in an included study (McMahon 2000). Fourteeen full-text articles were translated from non English languages, but then excluded (Del Vecchio-Blanco 1977; Aburano 1979; Mebazaa 1985; Salmi 1988; Luning 1991; Sakai 1991; Biwole Sida 1992; Bago 1993; Carriere 1993; Ding 1995; Beaugrand 2000; Baumgarten 2001; Ben Hassine 2007; Gao 2012).

Figure 1. Study flow diagram

Date of search: 5 June 2020

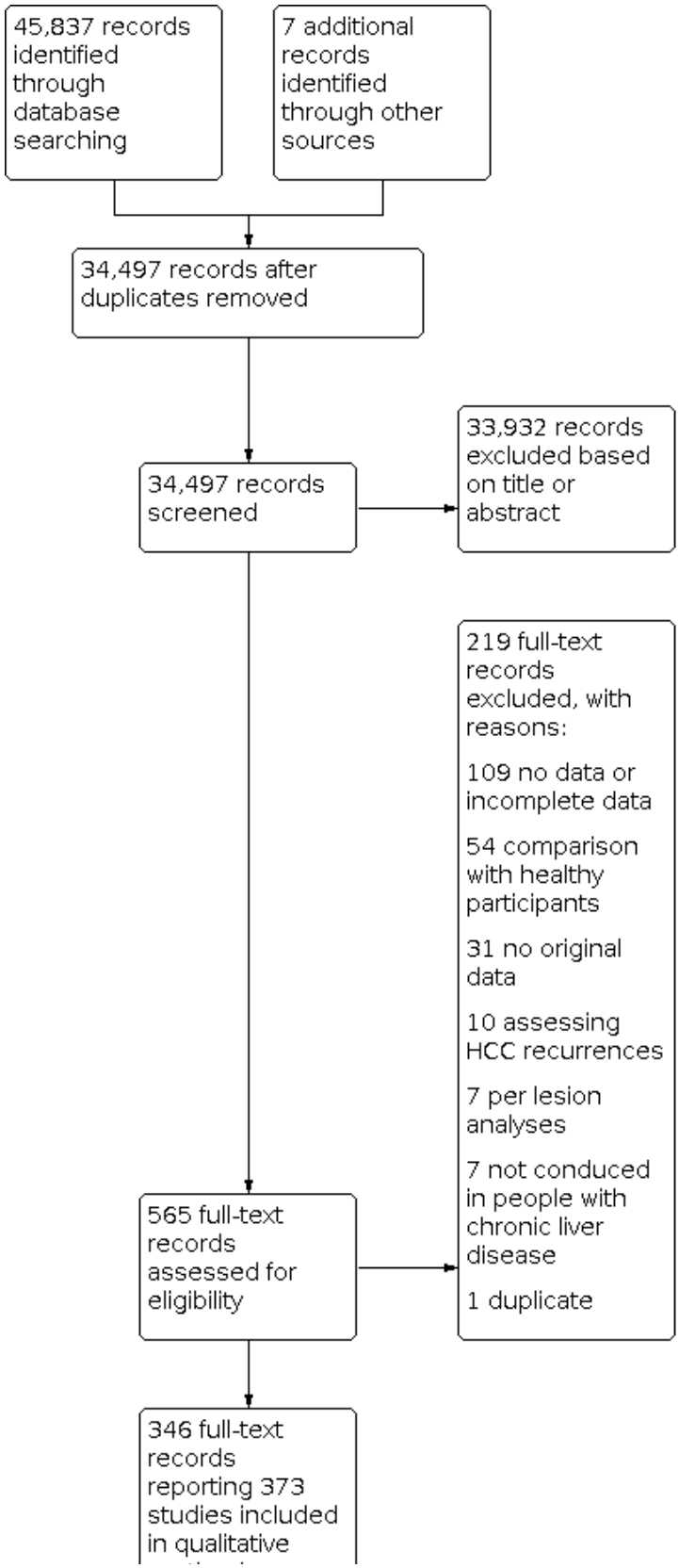
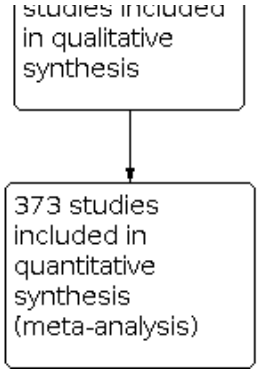


Figure 1. (Continued)



Finally, we included in our review 346 records reporting data on 373 studies (Figure 1), including as a whole 168,816 participants, with a percentage of males ranging from 40% to 100% and age ranging from 14 to 97 years. Thirteen papers reported multiple studies in different populations that we quoted and analysed separately as 22 studies (Wang 2013a; Wang 2013b; Wang 2014a; Wang 2014b; Wong 2014a; Wong 2014b; da Costa 2015a; da Costa 2015b: da Costa 2015c; da Costa 2015d; Li 2016b, Li 2016c; Tayob 2016a; Tayob 2016b; Wang 2016a, Wang 2016b; Wang 2016c; Wang 2016d; Wang 2016e; Luo 2018a; Luo 2018b; Luo 2018c). We translated six studies from non-English languages in order to include them in this review (Mauduit Astolfi 1987; Buffet 1988; Garretti 1988; Lee 2004; Kim 2006c; Kim 2006b). Concerning the direction of data collection, 77% (288/373) of the studies were retrospective.

We included 326 studies that assessed alpha-foetoprotein (AFP) as the index test in 144,570 participants; 39 studies that assessed abdominal ultrasound (US) in 18,792 participants; eight studies that assessed both AFP and abdominal US as the index tests in 5454 participants. The studies were conducted since 1971 for AFP, 1983 for abdominal US, and 1988 for the combination of AFP and US.

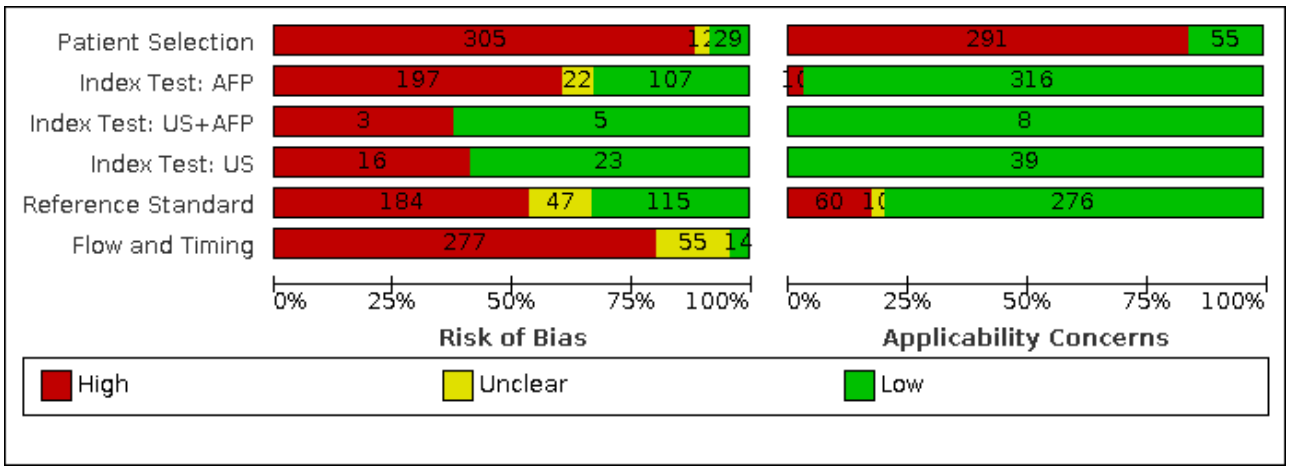
We reported in the Characteristics of included studies tables the main characteristics of the 373 studies. Investigators reported 19 studies only in abstract form, of which 17 with AFP as the index test (Song 2011; Cheng 2012; Kim 2012; Chan 2013; Unic 2013; Min 2014; Raff 2014; Khairy 2015; El-Serag 2017; Omar 2017; Park 2017b; Tsai 2017; Zheng 2017; Aboelfotoh 2018; Iyer 2018; Loglio 2018; Talkahn 2018), one with abdominal US as index test (Raff 2014), and one with both AFP and US as index tests (Raff 2014).

Of the 373 included studies, 190 were conducted in Asia, 66 in Europe, 57 in Africa, 55 in North and South America, and six were collaborative studies in two or three continents. Seventy-seven studies were conducted in the context of a surveillance program, and 297 studies in participants with clinical suspicion of having an HCC. Two hundred and eighty-eight studies were conducted retrospectively and 86, prospectively. Three hundred and eight studies used a mix of radiological imaging with or without histology as reference standard, 49 used only histology, and 17 used pathology of the explanted liver.

Methodological quality of included studies

We have reported in detail results of the quality assessment of included studies in the Characteristics of included studies tables, and we have summarised this information in Figure 2 and Appendix 3.

Figure 2. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies



Patient selection domain

Two hundred fifty-nine studies had a case-control design, 108 a cross-sectional design, and six a nested case-control design (Wong 2009; Lok 2010; Wang 2016d; Yu 2016; Choi 2019; Tayob 2019).

Alpha-foetoprotein (AFP)

Risk of bias: we judged 291 of 326 studies assessing the accuracy of AFP, with any cut-off, to be at high risk of bias. The most common reason was the case-control design (256 studies). Among the 70 cross-sectional studies,40 were judged to be at high risk of bias for inappropriate exclusion or for non-consecutive enrolment of participants. Seventeen studies were at low risk of bias in this domain (Arrigoni 1988; Cottone 1988; Sherman 1995; Chalasani 1999; Gambarin-Gelwan 2000; Ishii 2000; Tong 2001; Matievskaya 2003; Lee 2004; Sterling 2009; Song 2011; Singal 2012; Sterling 2012; Tayob 2016a; Tayob 2016b; Wang 2016b; Choi 2019). Among the 147 studies using 20 ng/mL as a cut-off value, we judged 129 studies to be at high risk of bias and 12 at unclear risk of bias; among the 56 studies using AFP with a cut-off value of 200 ng/mL, we judged 48 studies to be at high risk of bias.

Applicability: we judged 273 studies to be at high concern because study participants were highly selected on the basis of aetiology or severity of the liver diease and HCC characteristics. Among the 147 studies using 20 ng/mL as a cut-off value, we judged 115 studies to be at high concern; among the 56 studies using AFP with a cut-off value of 200 ng/mL, we judged 47 studies to be at high concern.

Abdominal ultrasound (US)

Risk of bias: 21 of the 39 studies assessing the accuracy of abdominal US were judged to be at high risk of bias: three studies were case-control studies (Powell-Jackson 1987; Jalli 2015; Yang 2019), and the remaining 18 were cross-sectional studies. The risk of bias was judged as high because of inappropriate exclusion or for non-consecutive enrolment of participants. Two studies were judged to be at unclear risk of bias for the latter domain as they did not report any exclusion criteria (Pateron 1994; Atiq 2017).

Applicability: we judged 22 studies at high concern as participants were highly selected on the basis of aetiology or severity of the liver diease and HCC characteristics.

Combination of AFP and abdominal (US)

Risk of bias: of the eight studies assessing the accuracy of the combination of AFP and abdominal US, three studies were judged at high risk of bias for inappropriate exclusion or for non-consecutive enrolment of participants (Buffet 1988; Chang 2015; Ungtrakul 2016). Chang 2015 and Ungtrakul 2016 used AFP with a cut-off of 20 ng/mL. All the eight studies were cross-sectional.

Applicability: we judged two studies to be at high concern, both of which with AFP cut-off value of 20 ng/mL, as only participants with severe liver disease on waiting list for orthotopic liver transplantation were included (Ungtrakul 2016; Gambarin-Gelwan 2000).

Index tests domain

Alpha-foetoprotein (AFP)

Risk of bias: we judged a total of 196 studies to be at high risk of bias. In 128 studies, no pre-definition of a cut-off value was reported. In 122 studies, the result of AFP measurement was interpreted knowing the result of the reference standard, and in 47 studies, it was unclear. Among the 147 studies using 20 ng/mL as a cut-off value, we judged 73 studies to be at high risk of bias; among the 56 studies using AFP wit a cut-off value of 200 ng/mL, we judged 54 studies to be at high risk of bias.

Applicabilty: we judged 10 studies to be at high concern due to variations in test technology, execution or interpretation (Alpert 1971; Giannelli 2005; Tan 2014; Wang 2014b; Wang 2016b; Wang 2016c; Wang 2016d; Wang 2016e; Wang 2019a; Sun 2020). All the studies using AFP with a cut-off value of 20 ng/mL or 200 ng/mL were at low concern.

Abdominal ultrasound (US)

Risk of bias: we judged 16 studies to be at high risk of bias as no definition of positivity criteria was reported (Okazaki 1984; Tanaka 1986; Cottone 1988; Garretti 1988; Tremolada 1989; Saada 1997; Yu 2011; Raff 2014; Chang 2015; Jalli 2015; Pinero 2015; Atiq 2017; Choi 2019; Kim 2019b; Kudo 2019; Yang 2019).

Applicability: we judged all the 39 studies to be at low concern.

Combination of AFP and abdominal US

Risk of bias: we judged three studies, two with a cut-off value of 20 ng/mL (Tremolada 1989; Kim 2019b), and one with a cut-off value of 5 ng/mL (Choi 2019) to be at high risk of bias as no definition of US positivity criteria was reported.

Applciability: we judged all eight studies to be at low concern.

Reference standard domain

Alpha-foetoprotein (AFP)

Risk of bias: we judged 174 studies to be at high risk of bias. In 105 studies with a case-control design, the reference standard was not adequate to exclude the presence of HCC, and in 24 studies, authors reported only how they assessed the presence of a chronic liver disease without any information concerning the target disease. In 100 studies, the reference standard was interpreted knowing the results of the index test, and in 43 studies we judged the available information to be insufficient.

Applicability: we judged 55 studies to be at high concern as pathological examination of explanted liver, or of surgical specimen, or necroscopy, or technologies that were no longer in use, were required to confirm the presence of HCC.

Abdominal ultrasound (US)

Risk of bias: we judged 23 studies to be at high risk of bias. In 20 studies, the reference standard was interpreted knowing the results of the index test, and in 11 studies the reference standard was judged to be inadequate to exclude the absence of HCC.

Applicability: we judged 13 studies to be at high concern as pathological examination of explanted liver, or of surgical specimen, or necroscopy, or technologies no longer in use, were required to confirm the presence of HCC.

Combination of AFP and abdominal US

Risk of bias: we judged five studies to be at high risk of bias, four using AFP with cut-off 20 ng/mL (Tremolada 1989; Singal 2012;

Ungtrakul 2016; Kim 2019b) and one with a cut-off of 250 ng/mL (Buffet 1988). In these studies, the reference standard was interpreted knowing the results of the index test and was judged inadequate to exclude the absence of HCC.

Applicability: we judged two studies to be at high concern as the reference standard was the pathological examination of explanted liver (Gambarin-Gelwan 2000) or histology and arteriography (Buffet 1988). Of these two studies, Gambarin-Gelwan 2000 used AFP with a cut-off value of 20 ng/mL.

Flow and timing domain

Alpha-foetoprotein (AFP)

Risk of bias: we judged 263 studies to be at high risk of bias. In 259 studies, participants did not receive the same reference standard. In six studies, the time interval between the index test and the reference standard was judged to be too long, whereas in other 305 studies, this information was not reported.

Abdominal ultrasound (US)

Risk of bias: we judged at high risk of bias 27 studies: in 22 studies participants did not receive the same reference standard. In six studies, the time interval between the index test and the reference standard was judged to be too long, whereas in other 25 studies, this information was not reported. Two studies reported the proportion of uninterpretable results (Atiq 2017, 56/523 and Maringhini 1988, 28/363), allowing an analysis according to the intention-to-diagnose principle, and another study included in the analyses uninterpretable results (Chang 2015).

Combination of AFP and abdominal US

Risk of bias: we judged six studies to be at high risk of bias (Buffet 1988; Tremolada 1989; Singal 2012; Chang 2015; Ungtrakul 2016; Kim 2019b). In five studies, participants did not receive the same reference standard,and in five studies, there was no information on the time interval between the index test and the reference standard. We judged one study to be at unclear risk of bias (Gambarin-Gelwan 2000), and one study to be at low risk of bias (Choi 2019). Of the six studies using AFP with cut-off 20 ng/mL, five were at high risk of bias and one at unclear risk of bias.

Overall assessment

As shown in Figure 2, we judged 304 studies at high risk of bias and 13 studies at unclear risk for the patient selection domain. For the index test domain, 196 studies with AFP were judged at high risk of bias and 23 at unclear risk; 16 studies with US were judged at high risk, and three studies with combination of AFP and US were judged at high risk. For the reference standard domain, 184 studies were judged at high risk of bias and 47 at unclear risk. For the flow and timing domain, 276 studies were judged at high risk of bias and 53 at unclear risk. We classified a study as having a high risk of bias if at least one of the domains of QUADAS-2 was judged as being at high or unclear risk of bias (Methods). We judged only one study to be at low risk of bias (Bennett 2002): this study was retrospectively conducted in a series of consecutive participants who underwent liver transplantation. The index test was abdominal US performed according to predefined positivity criteria and performed less than 90 days earlier, and the reference standard was the pathological examination of the explanted liver.

Concerning applicability, for the patient selection domain we judged at high concern 289 studies; for the index test domain 10 studies using AFP were judged at high concern, none using US or combination of AFP and US; for the reference standard domain 60 studies were judged at high concern and 10 at unclear concern.

Findings

Alpha-foetoprotein (AFP)

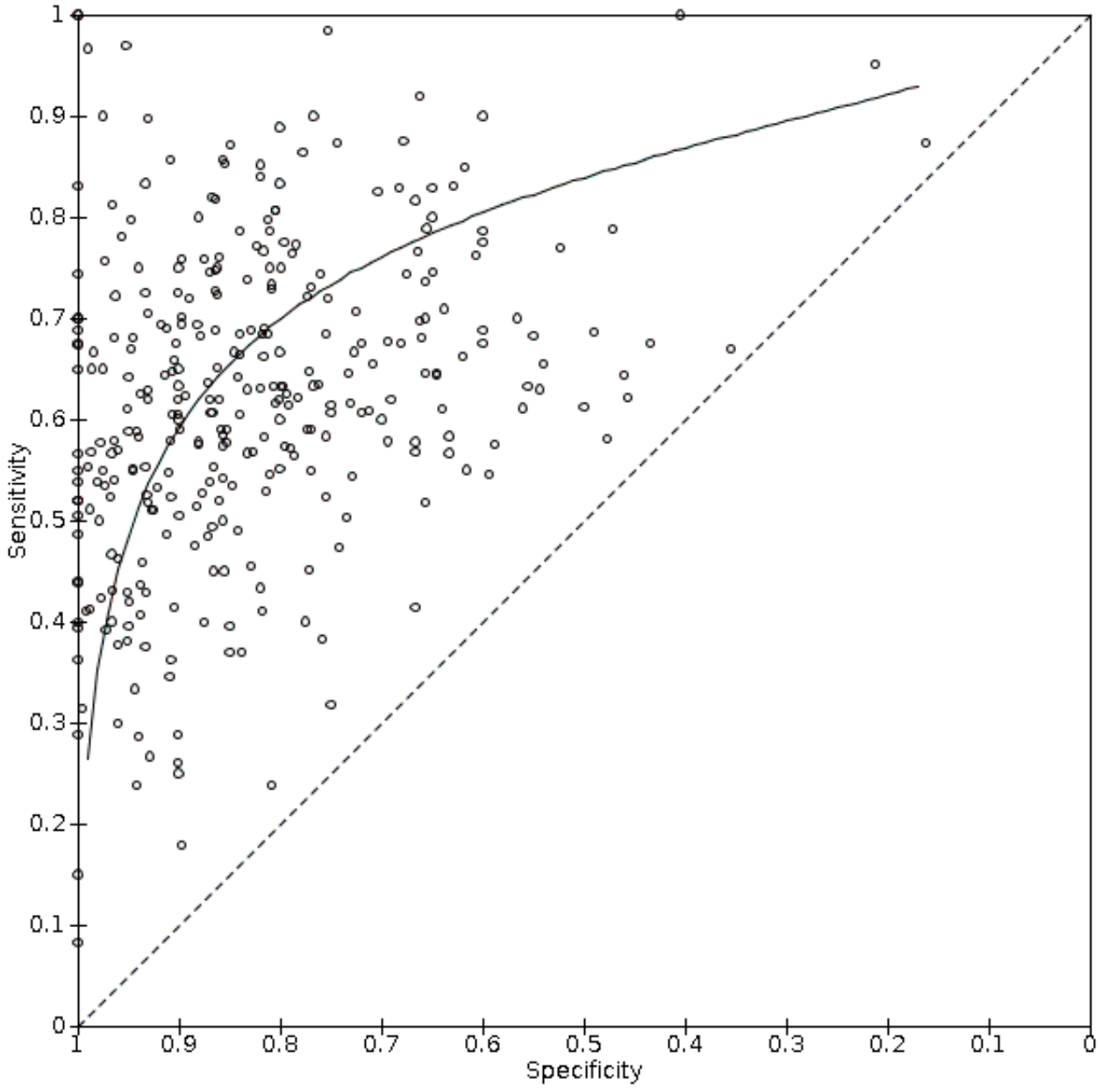
Description of the included studies

Three hundred and twenty-six studies with 144,570 participants provided data assessing serum alpha-foetoprotein (AFP) measurement for the diagnosis of HCC. The median prevalence of the target disease was 50% (interquartile range (IQR) 33% to 59%). When considering only the 70 cross-sectional studies, the median prevalence was 16% (IQR 9% to 33%). The cut-off values ranged from 5 ng/mL to 1000 ng/mL. The median prevalence of cirrhosis was 100% (IQR 73% to 100%). The median of the proportion of participants in Child-Pugh class A was 61% (IQR 38% to 82%) while the median proportion of participants with viral aetiology was 100% (IQR 76% to 100%). The median proportion of resectable HCC was 57% (IQR 34% to 91%) and the median of the mean HCC diameter across studies was 29.5 mm (IQR 20.5 mm to 46 mm). The studies were conducted from 1971 to 2020. Considering study location, 174 studies were conducted in Asia, 57 in Africa, 52 in Europe, 39 in North and South America, and four in more than one continent. Fifty studies were conducted in the context of a surveillance programme for HCC and 276 in a clinical setting.

Pooled results

Appendix 4 shows a forest plot of sensitivity and specificity with their 95% confidence intervals (CIs), and Figure 3 shows a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). We performed a meta-analysis using the hierarchical summary ROC model (HSROC) as the primary studies reported accuracy estimates of AFP using different cut-off values (Figure 3).

Figure 3. Summary receiver operating characteristic (ROC) comparing in 326 studies alpha-foetoprotein serum measurement with any cut-off value and different reference standards. Reference standards were: the pathology of the explanted liver in case of transplantation; the histology of resected focal liver lesions, or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months.



We then carried out two meta-analyses that included only studies that reported a cut-off value of 20 ng/mL or 200 ng/mL (the most used values).

AFP cut-off value around 20 ng/mL

Description of the included studies

One hundred forty seven studies with 52,144 participants provided data using a cut-off value of around 20 ng/mL (from 19 to 21 ng/mL). Five studies were published only in abstract form; 111 were case-control studies. The median prevalence of HCC across studies was 50% (IQR 33% to 63%). When considering only the 32 cross-sectional studies, the median prevalence was 11% (IQR 7% to 20%). The median proportion of participants with liver cirrhosis was 100% (data reported by 96 studies, IQR 75% to 100%), and the median prevalence of participants in Child-Pugh class A was 67% (51 studies, IQR 43% to 82%). The median proportion of participants with viral aetiology of cirrhosis was 97% (119 studies, IQR 78% to 100%) and the median of mean HCC across studies diameter was 27 mm (20 studies, IQR 22.5 to 46.5 mm). Finally, the median of

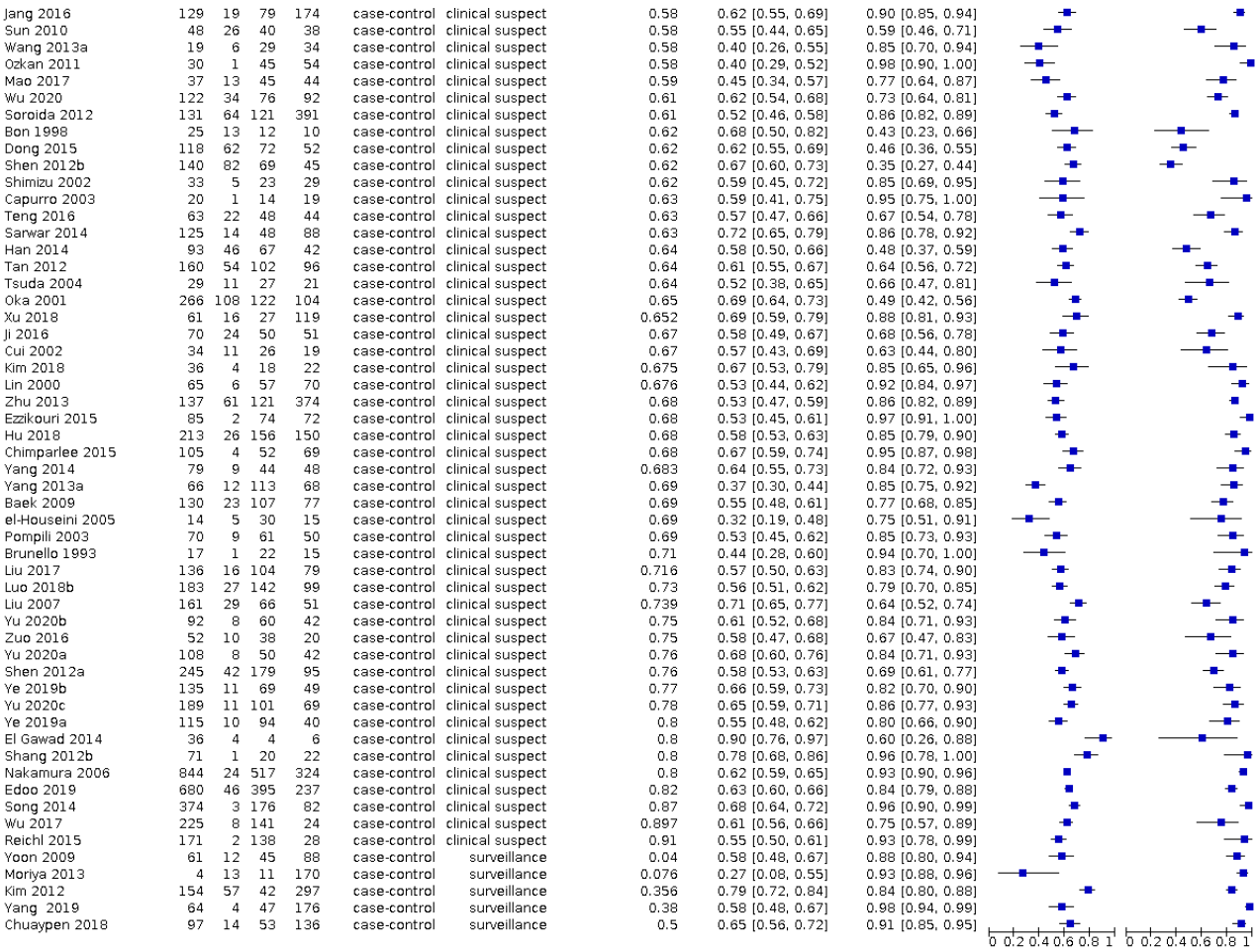
participants with resectable HCC was 59% (29 studies, IQR 42% to 87%). The studies were conducted from 1982 to 2020. Considering study location, 98 were conducted in Asia, 22 in Europe, 7 in Africa, 19 in North and South America, and one in three continents. Thirty studies were conducted in the context of a surveillance programme for HCC and 117 in a clinical setting. The sensitivity varied from 25% to 90% (IQR from 53% to 67%) and the specificity from 35% to 100% (IQR from 76% to 90%; Figure 4).

Figure 4. Forest plots of sensitivity and specificity of alpha-foetoprotein with a cut-off value around 20 ng/mL against different reference standards in 147 studies ordered by study design, setting and increasing HCC prevalence. Reference standards were: the pathology of the explanted liver in case of transplantation, the histology of resected focal liver lesions, or the histology of biopsied focal liver lesions with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets

are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line).

Study	TP	FP	FN	TN	design	setting	HCC prevalence	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Chalasani 1999	14	34	8	229	cross-sectional	clinical suspect	0.077	0.64 [0.41, 0.83]	0.87 [0.82, 0.91]		
Fabris 1991	23	38	4	173	cross-sectional	clinical suspect	0.13	0.85 [0.66, 0.96]	0.82 [0.76, 0.87]		
Gambarin-Gelwan 2000	11	8	8	79	cross-sectional	clinical suspect	0.179	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]		
Kim 2006c	26	26	16	159	cross-sectional	clinical suspect	0.185	0.62 [0.46, 0.76]	0.86 [0.80, 0.91]		
Lee 2004	34	91	20	108	cross-sectional	clinical suspect	0.213	0.63 [0.49, 0.76]	0.54 [0.47, 0.61]		
Yu 2016	22	2	29	136	cross-sectional	clinical suspect	0.269	0.43 [0.29, 0.58]	0.99 [0.95, 1.00]		
Paul 2007	68	29	33	165	cross-sectional	clinical suspect	0.36	0.67 [0.57, 0.76]	0.85 [0.79, 0.90]		
Maringhini 1988	114	46	32	171	cross-sectional	clinical suspect	0.402	0.78 [0.70, 0.84]	0.79 [0.73, 0.84]		
Wong 2009	24	0	13	37	cross-sectional	clinical suspect	0.5	0.65 [0.47, 0.80]	1.00 [0.91, 1.00]		
Iizuka 2010a	62	8	46	48	cross-sectional	clinical suspect	0.65	0.57 [0.48, 0.67]	0.86 [0.74, 0.94]		
Wong 2008	363	51	109	56	cross-sectional	clinical suspect	0.82	0.77 [0.73, 0.81]	0.52 [0.42, 0.62]		
Ungtrakul 2016	7	29	10	2247	cross-sectional	surveillance	0.007	0.41 [0.18, 0.67]	0.99 [0.98, 0.99]		
Sherman 1995	9	91	5	964	cross-sectional	surveillance	0.01	0.64 [0.35, 0.87]	0.91 [0.90, 0.93]		
El-Serag 2017	5	32	16	511	cross-sectional	surveillance	0.037	0.24 [0.08, 0.47]	0.94 [0.92, 0.96]		
Ishii 2000	18	153	11	552	cross-sectional	surveillance	0.04	0.62 [0.42, 0.79]	0.78 [0.75, 0.81]		
Raedle 1995	6	20	1	120	cross-sectional	surveillance	0.048	0.86 [0.42, 1.00]	0.86 [0.79, 0.91]		
Sterling 2012	25	219	21	590	cross-sectional	surveillance	0.057	0.54 [0.39, 0.69]	0.73 [0.70, 0.76]		
Okazaki 1984	11	44	3	187	cross-sectional	surveillance	0.057	0.79 [0.49, 0.95]	0.81 [0.75, 0.86]		
Li 2017b	11	10	8	271	cross-sectional	surveillance	0.063	0.58 [0.33, 0.80]	0.96 [0.94, 0.98]		
Chen 2003	142	466	115	2989	cross-sectional	surveillance	0.067	0.55 [0.49, 0.61]	0.87 [0.85, 0.88]		
Cheng 2012	71	229	60	1362	cross-sectional	surveillance	0.076	0.54 [0.45, 0.63]	0.86 [0.84, 0.87]		
Raff 2014	8	20	20	308	cross-sectional	surveillance	0.08	0.29 [0.13, 0.49]	0.94 [0.91, 0.96]		
Tremolada 1989	15	39	5	155	cross-sectional	surveillance	0.089	0.75 [0.51, 0.91]	0.80 [0.74, 0.85]		
Singal 2012	27	38	14	363	cross-sectional	surveillance	0.09	0.66 [0.49, 0.80]	0.91 [0.87, 0.93]		
Hallager 2018	63	71	41	469	cross-sectional	surveillance	0.1	0.61 [0.51, 0.70]	0.87 [0.84, 0.90]		
Sterling 2009	45	86	29	212	cross-sectional	surveillance	0.1	0.61 [0.49, 0.72]	0.71 [0.66, 0.76]		
Raedle 1998	52	52	23	584	cross-sectional	surveillance	0.11	0.69 [0.58, 0.79]	0.92 [0.89, 0.94]		
Atiq 2017	27	51	51	511	cross-sectional	surveillance	0.115	0.35 [0.24, 0.46]	0.91 [0.88, 0.93]		
Cottone 1988	10	39	5	104	cross-sectional	surveillance	0.157	0.67 [0.38, 0.88]	0.73 [0.65, 0.80]		
Kim 2019b	32	7	32	321	cross-sectional	surveillance	0.163	0.50 [0.37, 0.63]	0.98 [0.96, 0.99]		
da Costa 2015d	5	18	16	76	cross-sectional	surveillance	0.18	0.24 [0.08, 0.47]	0.81 [0.71, 0.88]		
Bolondi 2001	25	46	36	206	cross-sectional	surveillance	0.195	0.41 [0.29, 0.54]	0.82 [0.76, 0.86]		
Cedrone 2000	39	34	35	242	cross-sectional	surveillance	0.21	0.53 [0.41, 0.64]	0.88 [0.83, 0.91]		
Chang 2015	192	83	171	1151	cross-sectional	surveillance	0.227	0.53 [0.48, 0.58]	0.93 [0.92, 0.95]		
Oka 1994	21	48	34	150	cross-sectional	surveillance	0.24	0.38 [0.25, 0.52]	0.76 [0.69, 0.82]		
Lok 2010	24	15	15	62	cross-sectional	surveillance	0.34	0.62 [0.45, 0.77]	0.81 [0.70, 0.89]		
Bell 1982	12	10	2	100	case-control	clinical suspect	0.11	0.86 [0.57, 0.98]	0.91 [0.84, 0.96]		
Nomura 1996	17	17	10	84	case-control	clinical suspect	0.21	0.63 [0.42, 0.81]	0.83 [0.74, 0.90]		
Sultanik 2017	29	21	17	95	case-control	clinical suspect	0.28	0.63 [0.48, 0.77]	0.82 [0.74, 0.88]		
Chen 2018	99	70	103	371	case-control	clinical suspect	0.31	0.49 [0.42, 0.56]	0.84 [0.80, 0.87]		
Takikawa 1992	81	62	35	191	case-control	clinical suspect	0.31	0.70 [0.61, 0.78]	0.75 [0.70, 0.81]		
Weiss 2019	5	4	15	36	case-control	clinical suspect	0.33	0.25 [0.09, 0.49]	0.90 [0.76, 0.97]		
Luo 2018c	78	38	77	105	case-control	clinical suspect	0.34	0.50 [0.42, 0.58]	0.73 [0.65, 0.80]		
Shang 2012a	21	5	19	68	case-control	clinical suspect	0.35	0.53 [0.36, 0.68]	0.93 [0.85, 0.98]		
Passos-Castilho 2015	12	2	20	28	case-control	clinical suspect	0.37	0.38 [0.21, 0.56]	0.93 [0.78, 0.99]		
Zinkin 2008	29	14	12	37	case-control	clinical suspect	0.38	0.71 [0.54, 0.84]	0.73 [0.58, 0.84]		
Liao 2012	47	18	12	78	case-control	clinical suspect	0.38	0.80 [0.67, 0.89]	0.81 [0.72, 0.88]		
Ismail 2017a	45	12	21	87	case-control	clinical suspect	0.4	0.68 [0.56, 0.79]	0.88 [0.80, 0.94]		
Gopal 2014	316	69	135	603	case-control	clinical suspect	0.401	0.70 [0.66, 0.74]	0.90 [0.87, 0.92]		
Best 2016	166	24	119	378	case-control	clinical suspect	0.41	0.58 [0.52, 0.64]	0.94 [0.91, 0.96]		
Lin 2015	75	10	33	88	case-control	clinical suspect	0.42	0.69 [0.60, 0.78]	0.90 [0.82, 0.95]		
Nomura 1999	21	12	15	37	case-control	clinical suspect	0.42	0.58 [0.41, 0.74]	0.76 [0.61, 0.87]		
Tian 2017	66	56	54	90	case-control	clinical suspect	0.45	0.55 [0.46, 0.64]	0.62 [0.53, 0.70]		
Zhang 2020	47	26	16	48	case-control	clinical suspect	0.46	0.75 [0.62, 0.85]	0.65 [0.53, 0.76]		
Zhu 2020	55	42	46	179	case-control	clinical suspect	0.46	0.54 [0.44, 0.64]	0.81 [0.75, 0.86]		
Kim 2014	20	7	10	28	case-control	clinical suspect	0.462	0.67 [0.47, 0.83]	0.80 [0.63, 0.92]		
Kanmura 2007	12	11	17	22	case-control	clinical suspect	0.467	0.41 [0.24, 0.61]	0.67 [0.48, 0.82]		
Kim 2006b	40	25	15	37	case-control	clinical suspect	0.47	0.73 [0.59, 0.84]	0.60 [0.46, 0.72]		
Luo 2018a	22	18	14	23	case-control	clinical suspect	0.47	0.61 [0.43, 0.77]	0.56 [0.40, 0.72]		
Na 2013	36	13	21	51	case-control	clinical suspect	0.47	0.63 [0.49, 0.76]	0.80 [0.68, 0.89]		
Grazi 1995	61	3	50	113	case-control	clinical suspect	0.48	0.55 [0.45, 0.64]	0.97 [0.93, 0.99]		
Wang 2005	36	15	25	51	case-control	clinical suspect	0.48	0.59 [0.46, 0.71]	0.77 [0.65, 0.87]		
El-Sherif 2012	19	3	11	27	case-control	clinical suspect	0.5	0.63 [0.44, 0.80]	0.90 [0.73, 0.98]		
Fujii1995	40	6	10	44	case-control	clinical suspect	0.5	0.80 [0.66, 0.90]	0.88 [0.76, 0.95]		
Jeon 2016	86	14	71	142	case-control	clinical suspect	0.5	0.55 [0.47, 0.63]	0.91 [0.85, 0.95]		
Kumada 2014	43	10	61	94	case-control	clinical suspect	0.5	0.41 [0.32, 0.51]	0.90 [0.83, 0.95]		
Ahn 2016	185	45	181	321	case-control	clinical suspect	0.5	0.51 [0.45, 0.56]	0.88 [0.84, 0.91]		
da Costa 2015b	38	7	12	43	case-control	clinical suspect	0.5	0.76 [0.62, 0.87]	0.86 [0.73, 0.94]		
da Costa 2015c	37	10	38	65	case-control	clinical suspect	0.5	0.49 [0.38, 0.61]	0.87 [0.77, 0.93]		
Elnemr 2012	45	6	15	54	case-control	clinical suspect	0.5	0.75 [0.62, 0.85]	0.90 [0.79, 0.96]		
Tanglijvanich 2010	73	23	27	77	case-control	clinical suspect	0.5	0.73 [0.63, 0.81]	0.77 [0.68, 0.85]		
Trevisani 2001	102	16	68	154	case-control	clinical suspect	0.5	0.60 [0.52, 0.67]	0.91 [0.85, 0.95]		
Tsai 2017	253	58	240	435	case-control	clinical suspect	0.5	0.51 [0.47, 0.56]	0.88 [0.85, 0.91]		
Vongsuvanh 2016	37	6	49	166	case-control	clinical suspect	0.5	0.43 [0.32, 0.54]	0.97 [0.93, 0.99]		
Marrero 2009	247	42	172	375	case-control	clinical suspect	0.501	0.59 [0.54, 0.64]	0.90 [0.87, 0.93]		
Sanai 2010	118	34	88	165	case-control	clinical suspect	0.51	0.57 [0.50, 0.64]	0.83 [0.77, 0.88]		
Beneduce 2008	16	4	17	27	case-control	clinical suspect	0.52	0.48 [0.31, 0.66]	0.87 [0.70, 0.96]		
Chen 2015	77	13	26	82	case-control	clinical suspect	0.52	0.75 [0.65, 0.83]	0.86 [0.78, 0.93]		
Lin 2016	17	23	9	27	case-control	clinical suspect	0.52	0.65 [0.44, 0.83]	0.54 [0.39, 0.68]		
Nguyen 2002	103	30	60	119	case-control	clinical suspect	0.52	0.63 [0.55, 0.71]	0.80 [0.73, 0.86]		
Fang 2010	110	16	35	112	case-control	clinical suspect	0.53	0.76 [0.68, 0.83]	0.88 [0.80, 0.93]		
Lim 2015	205	48	156	228	case-control	clinical suspect	0.53	0.57 [0.52, 0.62]	0.83 [0.78, 0.87]		
Long 2011	76	13	35	56	case-control	clinical suspect	0.53	0.68 [0.59, 0.77]	0.81 [0.70, 0.90]		
da Costa 2015a	49	0	10	49	case-control	clinical suspect	0.54	0.83 [0.71, 0.92]	1.00 [0.93, 1.00]		
Wang 2014a	25	7	15	27	case-control	clinical suspect	0.54	0.63 [0.46, 0.77]	0.79 [0.62, 0.91]		
Beneduce 2004	26	9	34	41	case-control	clinical suspect	0.55	0.43 [0.31, 0.57]	0.82 [0.69, 0.91]		
Song 2002	18	8	20	23	case-control	clinical suspect	0.55	0.47 [0.31, 0.64]	0.74 [0.55, 0.88]		
Li 2016a	31	6	22	36	case-control	clinical suspect	0.558	0.58 [0.44, 0.72]	0.86 [0.71, 0.95]		
Gani 2015	43	9	16	38	case-control	clinical suspect	0.56	0.73 [0.60, 0.84]	0.81 [0.67, 0.91]		
Shu 2010	113	44	49	86	case-control	clinical suspect	0.56	0.70 [0.62, 0.77]	0.66 [0.57, 0.74]		
Cui 2003	70	33	50	57	case-control	clinical suspect	0.57	0.58 [0.49, 0.67]	0.63 [0.53, 0.73]		
Shaheen 2015	25	14	15	16	case-control	clinical suspect	0.57	0.63 [0.46, 0.77]	0.53 [0.34, 0.72]		
Jang 2016	129	19	79	174	case-control	clinical suspect	0.58	0.62 [0.55, 0.69]	0.90 [0.85, 0.94]		
Sun 2010	48	26	40	38	case-control	clinical suspect	0.58	0.55 [0.44, 0.65]	0.59 [0.46, 0.71]		
Wang 2013a	19	6	29	34	case-control	clinical suspect	0.58	0.40 [0.26, 0.55]	0.85 [0.70, 0.94]		

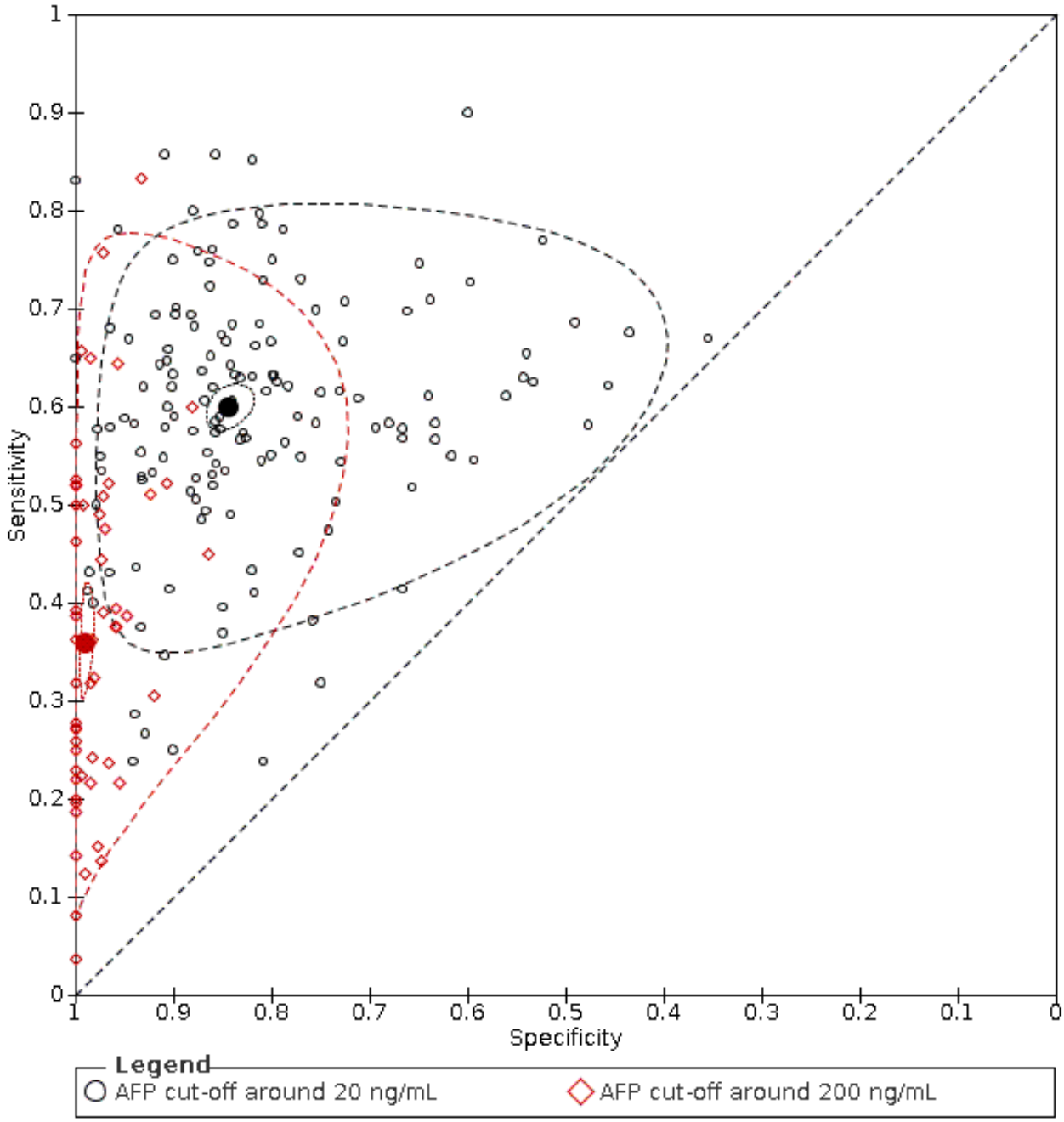
Figure 4. (Continued)



Pooled results

By using the bivariate model, we obtained the following pooled estimates: sensitivity 60% (95% CI 58% to 62%), specificity 84% (95% CI 82% to 86%), LR+ 3.84 (95% CI 3.39 to 4.33), LR- 0.48 (95% CI 0.45 to 0.50; Figure 5).

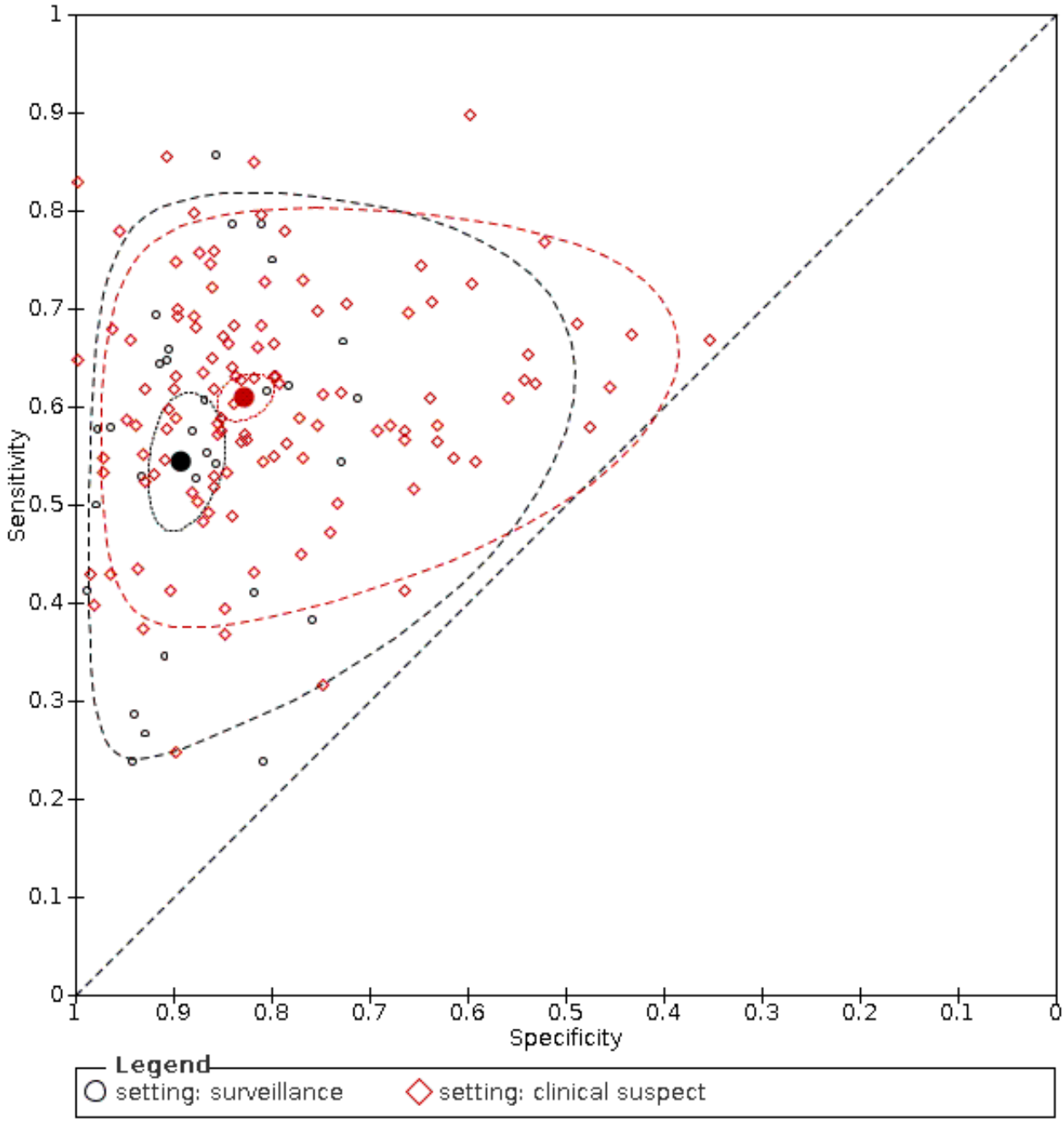
Figure 5. Summary receiver operating characteristic (ROC) comparing alpha-foetoprotein with a cut-off value around 20 ng/mL (black circles) and alpha-foetoprotein with a cut-off value around 200 ng/mL (red diamonds) against the same reference standards.
Reference standards were: the pathology of the explanted liver in case of transplantation;the histology of resected focal liver lesion(s), or the histology of biopsied focal liver lesions with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. The solid circles represent the summary estimates of sensitivity and specificity for AFP cut-off around 20 ng/ml (black circle) and AFP cut off 200 ng/ml (red circle). The dotted lines represent the 95% confidence regions. The dashed lines represent the 95% prediction regions.



In the 30 studies conducted in a surveillance programme, the pooled sensitivity was 54% (95% CI 59% to 63%) and specificity 83% (95% CI 84% to 85%); in the 117 studies conducted in a clinical

setting, the pooled sensitivity was 61% (95% CI 59% to 63%) and the specificity 83% (95% CI 84% to 85%; Figure 6).

Figure 6. Summary receiver operating characteristic (ROC) comparing the results of studies conducted in different settings, surveillance programs (black circles) and clinical setting (red diamonds) against the same reference standards. Reference standards were: the pathology of the explanted liver in case of transplantation; the histology of resected focal liver lesions, or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. The solid circles represent the summary estimates of sensitivity and specificity for surveillance setting (black circle) and clinical suspect setting (red circle). The dotted lines represent the 95% confidence regions. The dashed lines represent the 95% prediction regions.



We assessed the diagnostic accuracy for resectable HCC as a secondary objective. We found six studies with 1722 participants with more than 90% of participants with resectable HCC (Nomura 1996; Nomura 1999; Gambarin-Gelwan 2000; Shen 2012b; Tan 2012; Song 2014). By using the bivariate model, the sensitivity was 65% (95% CI 62% to 68%), the specificity 80% (95% CI 59% to 91%), LR+ 3.2 (95% CI 1.4 to 7.2) and LR- 0.44 (95% CI 0.34 to 0.56).

Heterogeneity analysis

We investigated heterogeneity while considering studies with AFP cut-off values around 20 ng/mL. Table 2 shows the comparisons of different predefined subgroups. The estimates of sensitivity and specificity were different only for the comparison of studies including participants recruited from planned surveillance programs compared to clinical cohorts (P = 0.005).

Sensitivity analysis

When considering only the 36 studies with a cross-sectional design, we obtained an AFP sensitivity of 57% (95% CI 52% to 62%) and specificity of 88% (95% CI 84% to 91%; Table 2). When considering the 142 studies published in full text, we obtained an AFP sensitivity of 60% (95% CI 58% to 62%) and specificity of 84% (95% CI 82% to 86%; Table 2). We did not perform the remaining sensitivity analyses as all studies were judged to be at high risk of bias, and no study reported uninterpretable results.

AFP cut-off value 200 ng/mL

Description of the included studies

Fifty-six studies with 20,452 participants provided data using a cut-off value of 200 ng/mL.

Two studies were published only in abstract form, 42 were case-control studies. The median prevalence of HCC was 51% (IQR 34% to 63%). When considering only the 14 cross-sectional studies, the median prevalence was 21% (IQR 9% to 34%). The median proportion of participants with liver cirrhosis was 100% (data reported by 41 studies, IQR 92% to 100%) and the median prevalence of Child-Pugh class A participants was 47% (24 studies, IQR 32% to 77%); the median proportion of participants with viral aetiology of cirrhosis was 100% (41 studies, IQR 79% to 100%); the median of the mean HCC diameter across studies was 31 mm (10 studies, IQR 20 mm to 42 mm). The median prevalence of resectable HCC was 51% (10 studies, IQR 36% to 73%). The studies were conducted from 1988 to 2018. Considering study location, 31 studies were conducted in Asia, nine in Africa, nine in North and South America, and eight in Europe, Seven studies were conducted in the context of a surveillance programme for HCC and 49 in a clinical setting. Sensitivity varied from 4% to 83% (IQR 23% to 50%) and specificity from 87% to 100% (IQR from 97% to 100%; Appendix 5).

Pooled results

By using the bivariate model, we obtained the following estimates: sensitivity 36% (95% CI 31% to 41%), specificity 99% (95% CI 98% to 100%), LR+ 35.9 (95% CI 22.2 to 57.9) LR- 0.64 (95% CI 0.60 to 0.695; Figure 5).

We assessed the diagnostic accuracy for resectable HCC as a secondary objective. We found only two studies with more than 90% of participants with resectable HCC, preventing a meta-analysis of their results: Nomura 1996, with 128 participants, reported a sensitivity of 4% (95% CI 0% to 19%) and a specificity of 100% (95% CI 96% to 100%) and Sassa 1999, with 195 participants, reported a sensitivity of 8% (95% CI 3% to 18%) and a specificity of 100% (95% CI 97% to 100%).

Heterogeneity analysis

We investigated heterogeneity while considering studies with AFP cut-off value of 200 ng/mL. Table 3 shows the comparisons of different predefined subgroups. The estimates of sensitivity and specificity were different for the comparison of studies conducted in different continents; and also for studies including more than 50% of participants in Chil-Pugh class A compared to studies including less than 50% in Child-Pugh class A.

Sensitivity analysis

When considering only the 14 studies with a cross-sectional design, we obtained an AFP sensitivity of 39% (95% CI 28% to 51%) and a specificity of 99% (95% CI 98% to 99%; Table 3).

When considering the 54 studies published in full text and excluding the two published in abstract form, we obtained an AFP sensitivity of 36% (95% CI 31% to 41%) and a specificity of 99% (95% CI 98% to 100%; Table 3),

We did not perform the remaining sensitivity analyses as all studies were judged to be at high risk of bias, and no study reported uninterpretable results.

Abdominal ultrasound (US)

Description of the included studies

Thirty-nine studies with 18,792 participants provided data assessing abdominal ultrasound (US) for the diagnosis of HCC.

The median prevalence of the target disease was 15% (interquartile range 8% to 31%). When considering the 36 cross-sectional studies, the median prevalence of HCC was 15% (IQR 9% to 25%). All included participants had hepatic cirrhosis. The median prevalence of Child-Pugh class A participants was 69% (14 studies, IQR 30% to 81%), and the median proportion of participants with viral aetiology was 60% (26 studies, IQR 40% to 84%). The median proportion of participants with resectable HCC was 76% (20 studies, IQR 40% to 95%) and the median of the mean diameter across studies was 24 mm (17 studies, IQR 20.5 mm to 31 mm). The studies were conducted from 1983 to 2020. Considering study location, 13 studies were conducted in North and South America, 13 in Asia, 12 in Europe, and one in three continents. Twenty studies were conducted in the context of a surveillance program for HCC and 19 in participants with clinical suspected HCC.

Pooled results

Figure 7 shows the forest plot of sensitivity and specificity with their 95% CIs, and Figure 8 shows a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Sensitivity ranged from 28% to 100% (IQR 44% to 89%) and specificity from 43% to 100% (IQR 86% to 96%). We performed a meta-analysis using the bivariate model, as the index test results are dichotomous (i.e. positive or negative) without a threshold. We obtained the following estimates: sensitivity 72%

(95% CI 63% to 79%), specificity 94% (95% CI 91% to 96%), LR+ 12.5 (95% CI 8.6 to 18.25), LR- 0.29 (95% CI 0.22 to 0.39).

Figure 7. Forest plots of sensitivity and specificity of ultrasound against different reference standards.in 39 studies. Reference standards were: the pathology of the explanted liver in case of transplantation.;the histology of resected focal liver lesions, or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line).The individual studies are ordered by study design (cross-sectional or case-control), study setting (clinical setting or surveillance program) and increasing sensitivity.

Study	TP	FP	FN	TN	design	setting	Sensitivity (95% CI)	Specificity (95% CI)
Bennett 2002	8	6	19	167	cross-sectional	clinical suspect	0.30 [0.14, 0.50]	0.97 [0.93, 0.99]
Saada 1997	2	0	4	33	cross-sectional	clinical suspect	0.33 [0.04, 0.78]	1.00 [0.89, 1.00]
Kim 2001	6	3	10	33	cross-sectional	clinical suspect	0.38 [0.15, 0.65]	0.92 [0.78, 0.98]
Libbrecht 2002	6	0	9	31	cross-sectional	clinical suspect	0.40 [0.16, 0.68]	1.00 [0.89, 1.00]
Dodd 1992	12	4	16	168	cross-sectional	clinical suspect	0.43 [0.24, 0.63]	0.98 [0.94, 0.99]
Gambarin-Gelwan 2000	11	5	8	82	cross-sectional	clinical suspect	0.58 [0.33, 0.80]	0.94 [0.87, 0.98]
Chalasani 1999	13	18	9	245	cross-sectional	clinical suspect	0.59 [0.36, 0.79]	0.93 [0.89, 0.96]
Van Thiel 2004	12	5	8	77	cross-sectional	clinical suspect	0.60 [0.36, 0.81]	0.94 [0.86, 0.98]
Yu 2011	88	11	50	281	cross-sectional	clinical suspect	0.64 [0.55, 0.72]	0.96 [0.93, 0.98]
Villacastin Ruiz 2016	86	6	21	157	cross-sectional	clinical suspect	0.80 [0.72, 0.87]	0.96 [0.92, 0.99]
Buffet 1988	20	16	4	168	cross-sectional	clinical suspect	0.83 [0.63, 0.95]	0.91 [0.86, 0.95]
Maringhini 1988	128	37	18	180	cross-sectional	clinical suspect	0.88 [0.81, 0.93]	0.83 [0.77, 0.88]
Teefey 2003	8	4	1	12	cross-sectional	clinical suspect	0.89 [0.52, 1.00]	0.75 [0.48, 0.93]
Cottone 1983	27	5	3	65	cross-sectional	clinical suspect	0.90 [0.73, 0.98]	0.93 [0.84, 0.98]
Mauduit Astolfi 1987	33	3	2	42	cross-sectional	clinical suspect	0.94 [0.81, 0.99]	0.93 [0.82, 0.99]
Garretti 1988	38	54	2	306	cross-sectional	clinical suspect	0.95 [0.83, 0.99]	0.85 [0.81, 0.89]
Wong 2008	463	61	9	46	cross-sectional	clinical suspect	0.98 [0.96, 0.99]	0.43 [0.33, 0.53]
Park 2020	12	56	31	958	cross-sectional	surveillance	0.28 [0.15, 0.44]	0.94 [0.93, 0.96]
Pinero 2015	19	2	37	514	cross-sectional	surveillance	0.34 [0.22, 0.48]	1.00 [0.99, 1.00]
Son 2019	11	27	17	352	cross-sectional	surveillance	0.39 [0.22, 0.59]	0.93 [0.90, 0.95]
Atiq 2017	33	119	45	326	cross-sectional	surveillance	0.42 [0.31, 0.54]	0.73 [0.69, 0.77]
Singal 2012	18	34	23	367	cross-sectional	surveillance	0.44 [0.28, 0.60]	0.92 [0.88, 0.94]
Choi 2019	17	6	18	162	cross-sectional	surveillance	0.49 [0.31, 0.66]	0.96 [0.92, 0.99]
Raff 2014	3	9	3	148	cross-sectional	surveillance	0.50 [0.12, 0.88]	0.94 [0.89, 0.97]
Kim 2019b	36	0	28	328	cross-sectional	surveillance	0.56 [0.43, 0.69]	1.00 [0.99, 1.00]
Tanaka 1986	31	6	22	5286	cross-sectional	surveillance	0.58 [0.44, 0.72]	1.00 [1.00, 1.00]
Kudo 2019	17	9	9	240	cross-sectional	surveillance	0.65 [0.44, 0.83]	0.96 [0.93, 0.98]
Cottone 1988	10	3	5	140	cross-sectional	surveillance	0.67 [0.38, 0.88]	0.98 [0.94, 1.00]
Sherman 1995	5	33	2	498	cross-sectional	surveillance	0.71 [0.29, 0.96]	0.94 [0.91, 0.96]
Pateron 1994	11	7	3	97	cross-sectional	surveillance	0.79 [0.49, 0.95]	0.93 [0.87, 0.97]
Tremolada 1989	17	18	3	176	cross-sectional	surveillance	0.85 [0.62, 0.97]	0.91 [0.86, 0.94]
Mok 2004	18	15	3	67	cross-sectional	surveillance	0.86 [0.64, 0.97]	0.82 [0.72, 0.89]
Okazaki 1984	12	3	2	228	cross-sectional	surveillance	0.86 [0.57, 0.98]	0.99 [0.96, 1.00]
Chang 2015	334	318	29	916	cross-sectional	surveillance	0.92 [0.89, 0.95]	0.74 [0.72, 0.77]
Ungtrakul 2016	16	397	1	1879	cross-sectional	surveillance	0.94 [0.71, 1.00]	0.83 [0.81, 0.84]
Sutherland 2017	6	19	0	167	cross-sectional	surveillance	1.00 [0.54, 1.00]	0.90 [0.85, 0.94]
Jalli 2015	18	9	12	57	case-control	clinical suspect	0.60 [0.41, 0.77]	0.86 [0.76, 0.94]
Powell-Jackson 1987	24	5	3	79	case-control	clinical suspect	0.89 [0.71, 0.98]	0.94 [0.87, 0.98]
Yang 2019	80	37	10	136	case-control	surveillance	0.89 [0.81, 0.95]	0.79 [0.72, 0.84]

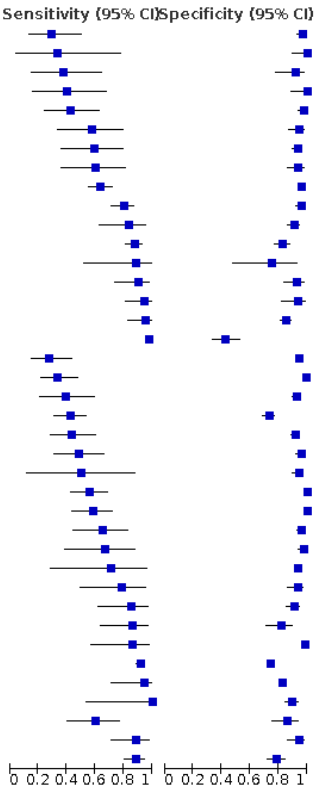
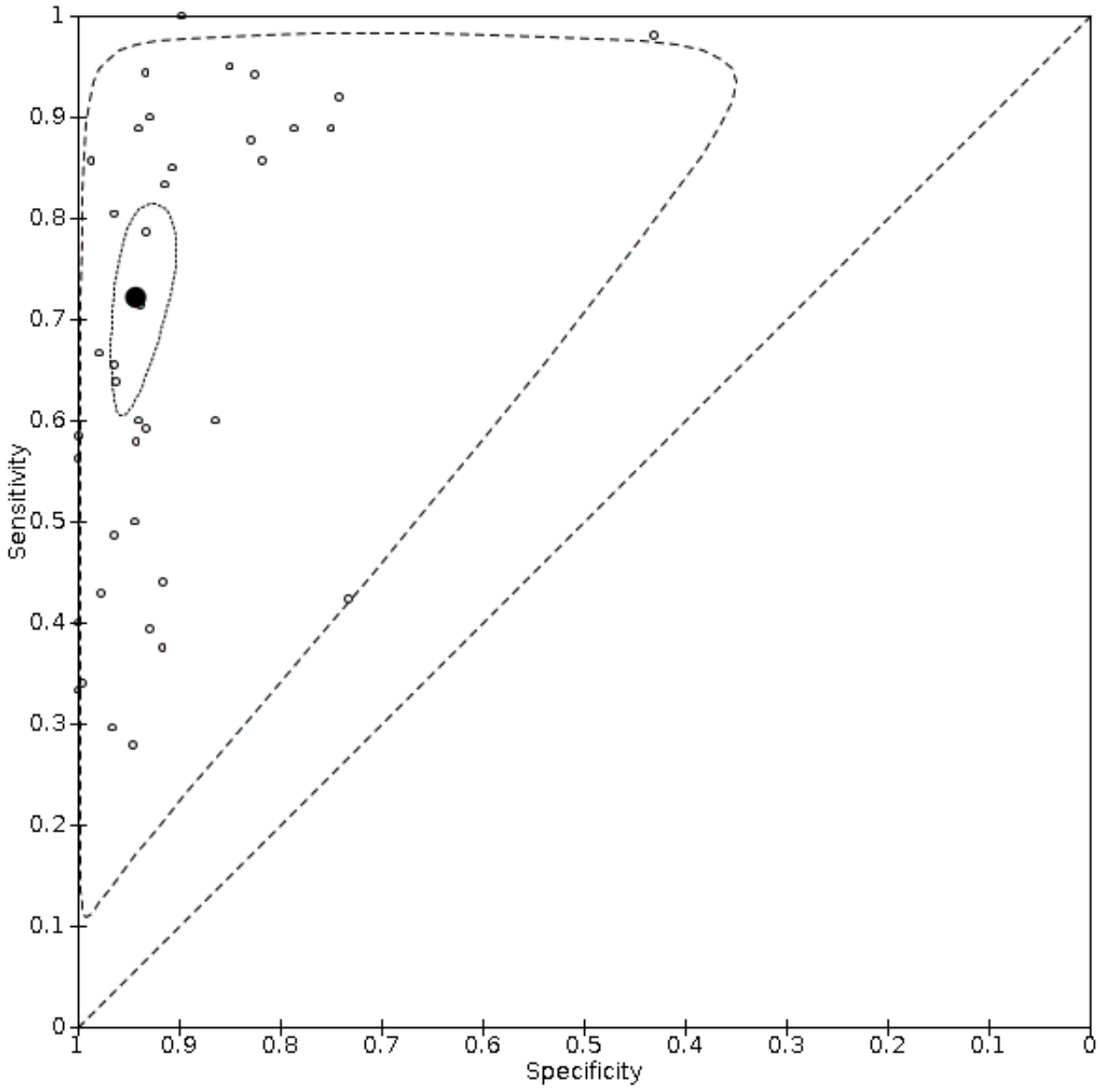


Figure 8. Summary receiver operating characteristic (ROC) comparing, in 39 studies, ultrasound and different reference standards. Reference standards were: the pathology of the explanted liver in case of transplantation; the histology of resected focal liver lesions, or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. The solid circle represents the summary estimate of sensitivity and specificity.The dotted lines represent the 95% confidence regions. The dashed lines represent the 95% prediction regions.



We assessed the diagnostic accuracy for resectable HCC as a secondary objective. We found seven studies (2163 participants) with more than 90% with resectable HCC (Dodd 1992; Gambarin-Gelwan 2000; Kim 2001; Villacastin Ruiz 2016; Choi 2019; Kudo 2019; Park 2020). By using the bivariate model, the pooled sensitivity was 53% (95% CI 38% to 67%), specificity 96% (95% CI 94% to 97%), LR + 12.3 (95% CI 7.7 to 19.5), LR- 0.5 95% CI 0.36 to 0.68).

Heterogeneity analysis

We investigated heterogeneity while considering studies using US as the index test and found no difference between the prespecified subgroups (Table 4).

Sensitivity analysis

When considering only the 36 studies with a cross-sectional design, we obtained a pooled sensitivity of 71% (95% CI 62% to 79%) and a specificity of 95% (95% CI 92% to 97%; Table 4).

When considering only the 25 studies that prespecified the positivity criteria, we obtained a pooled sensitivity of 74% (95% CI 63% to 83%) and a specificity of 93% (95% CI 89% to 96%; Table 4).

When considering only the three studies reporting uninterpretable results with intention-to-diagnose analysis, we obtained a sensitivity of 80% (95% CI 71% to 81%) and a specificity of 76% (95% CI 71% to 81%).

When considering the 38 studies published in full text and excluding the two published studies in abstract form, we obtained sensitivity of 72% (95% CI 64% to 80%) and specificity of 94% (95% CI 91% to 96%).

Combination of AFP and US

Description of the included studies

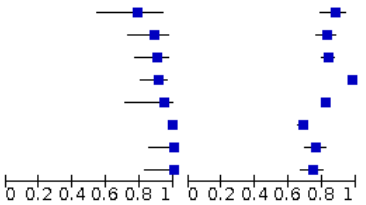
Eight studies with 5454 participants provided data assessing the combination of measurement of serum AFP and abdominal US for the diagnosis of HCC.

All studies considered positive the combination of the two tests when at least one was positive. The median prevalence of the target disease was 16% (IQR 9% to 17%). The median proportion of participants with liver cirrhosis was 100% (data reported by eight studies: in six studies it was 100%, in one study it was 93%, and in another study it was 53%). The median prevalence of participants with Child-Pugh class A was 86% (data reported by four studies, IQR 60% to 96%) and the median prevalence of participants with viral aetiology was 84% (six studies, IQR 44% to 88%). The median proportion of resectable HCC was 76% (six studies, IQR 59% to 91%), and the mean diameter was 24 mm (four studies, IQR 18.5 to 31.5 mm). The studies were conducted from 1988 to 2019. Considering study location, three studies were conducted in North and South America, three in Asia, one in Europe, and one in three continents. Seven studies were conducted in the context of a surveillance programme for HCC and two studies in participants with the clinical suspected HCC.

Figure 9 shows the forest plot of sensitivity and specificity with their 95% CIs, and Appendix 6 shows a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Considering only the six studies (5,044 participants) which used for AFP a cut-off value of 20 ng/mL, we performed a meta-analysis using the bivariate model and we obtained the following pooled estimates: sensitivity 96% (95% CI 88% to 98%), specificity 85% (95% CI 73% to 93%), LR+ 6.5 (95% CI 3.5 to 12.0) and LR- 0.05 (95% CI 0.02 to 0.14; (Tremolada 1989; Gambarin-Gelwan 2000; Singal 2012; Chang 2015; Ungtrakul 2016; Kim 2019b)).

Figure 9. Forest plots of sensitivity and specificity of the combination of alpha-foetoprotein and ultrasound against different reference standards in 8 studies ordered by increasing sensitivity.Rreference standards were: the pathology of the explanted liver in case of transplantation.;the histology of resected focal liver lesion(s), or the histology of biopsied focal liver lesions with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line)

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Gambarin-Gelwan 2000	15	11	4	76	20.0	0.79 [0.54, 0.94]	0.87 [0.79, 0.94]
Choi 2019	31	29	4	139	5.0	0.89 [0.73, 0.97]	0.83 [0.76, 0.88]
Singal 2012	37	67	4	334	20.0	0.90 [0.77, 0.97]	0.83 [0.79, 0.87]
Kim 2019b	58	7	6	321	20.0	0.91 [0.81, 0.96]	0.98 [0.96, 0.99]
Ungtrakul 2016	16	404	1	1872	20.0	0.94 [0.71, 1.00]	0.82 [0.81, 0.84]
Chang 2015	360	391	3	843	20.0	0.99 [0.98, 1.00]	0.68 [0.66, 0.71]
Buffet 1988	23	44	0	140	250.0	1.00 [0.85, 1.00]	0.76 [0.69, 0.82]
Tremolada 1989	20	50	0	144	20.0	1.00 [0.83, 1.00]	0.74 [0.67, 0.80]



We assessed the diagnostic accuracy for resectable HCC as a secondary objective. We found only two studies with more than 90% of participants with resectable HCC, preventing a meta-analysis of their results: Choi 2019 with 203 participants, reported a sensitivity of 89% (95% CI 73% to 97%) and specificity of 83% (95% CI 76% to 88%) and Gambarin-Gelwan 2000 with 106 participants, reported a sensitivity of 79% (95% CI 54% to 94%) and a specificity of 87% (95% CI 79% to 94%).

Heterogeneity analysis

We investigated heterogeneity while considering studies using the combination of AFP 20 ng/mL and US as the index test and found no difference between some prespecified subgroups: prospective compared to retrospective studies, studies conducted before 2000 compared to those conducted after 2000, studies with HCC prevalence lower than 10% compared to studies with HCC prevalence higher than 10%, studies conducted in surveillance programmes compared to studies conducted in people with suspected HCC. We could not assess the remaining comparisons because of the small number of included studies (Table 5).

Sensitivity analysis

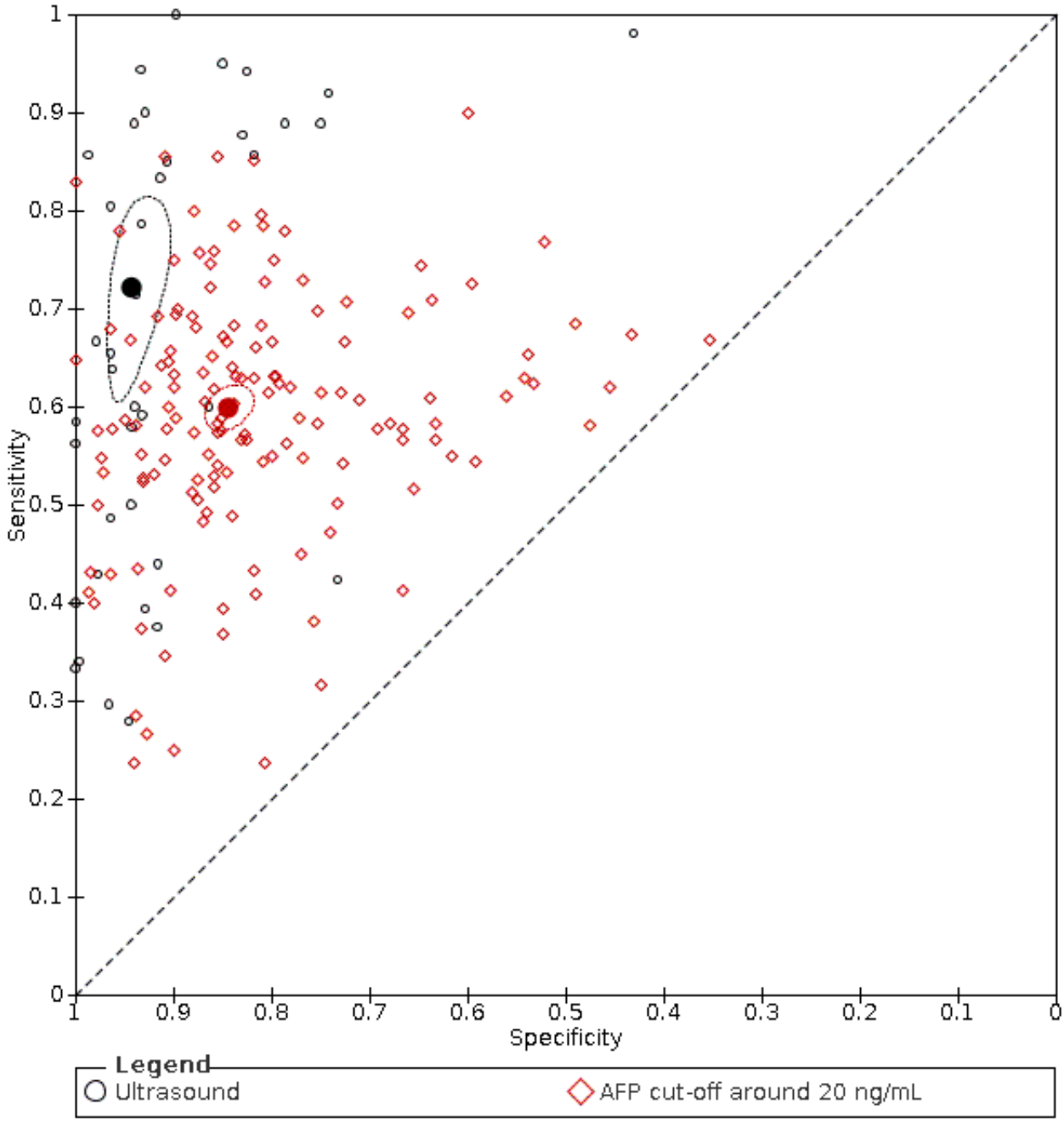
We did not perform the sensitivity analyses as all the studies were judged to be at high risk of bias; all the studies were cross-sectional; no study reported uninterpretable results; all the studies were published as full text, and only two studies reported predefined US positivity criteria (Singal 2012; Ungtrakul 2016; Table 5).

Comparative analyses

The indirect comparison between the 147 studies with AFP at a cut-off value of around 20 ng/mL showed an AFP sensitivity of 60%

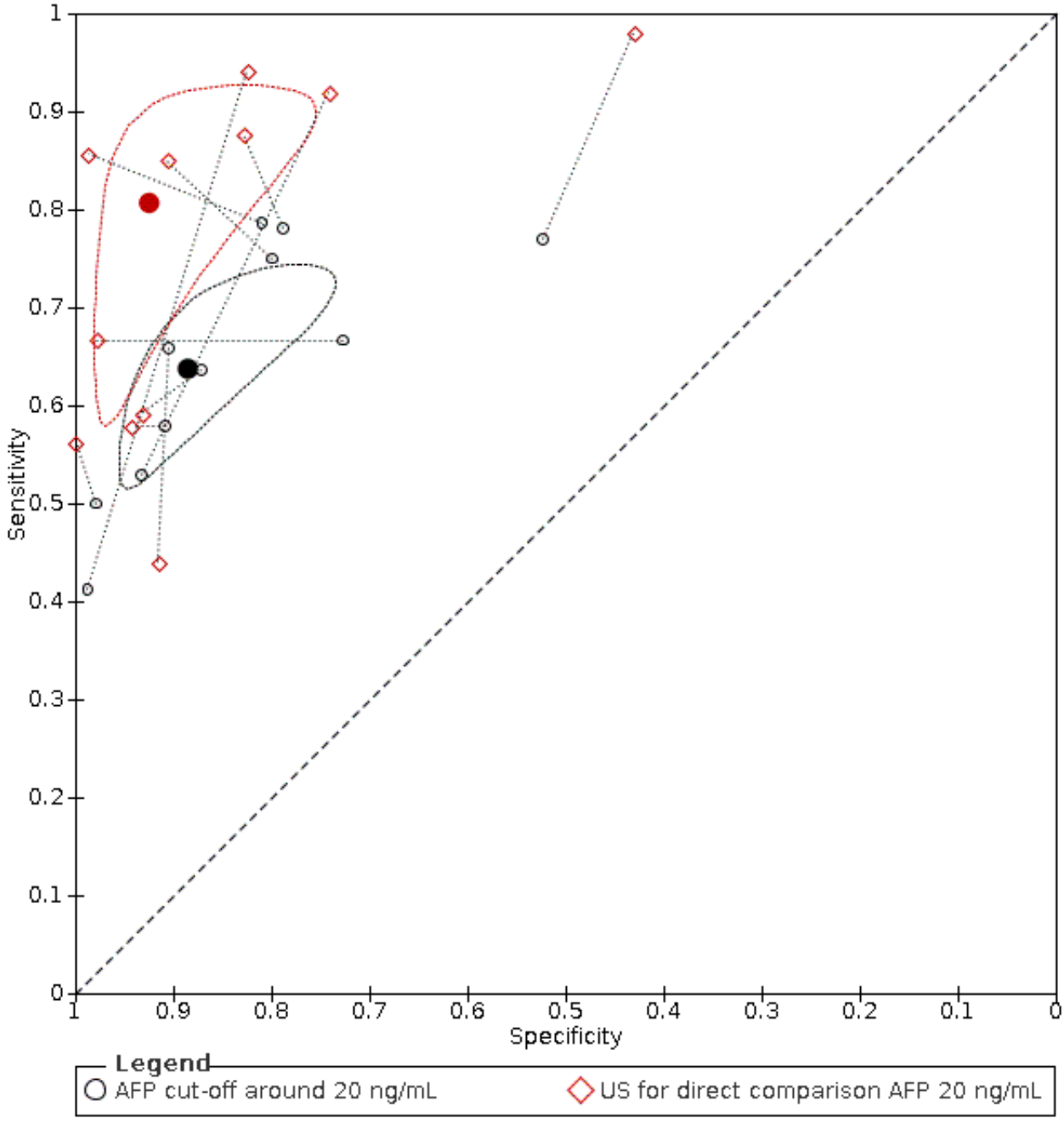
(95% CI 58% to 62%) and specificity of 84% (95% CI 82% to 86%) compared to the 39 studies with US showing a sensitivity of 72% (95% CI 63% to 79%) and specificity of 94% (95% CI 91% to 96%). Both US sensitivity (P = 0.0011) and specificity (P < 0.0001) were higher than those of AFP (Figure 10).

Figure 10. Summary receiver operating characteristic (ROC) showing the indirect comparison (between study) of the results of two different index tests, ultrasound (black circles) and alpha-foetoprotein with a cut-off value around 20 ng/mL (red diamonds) against the same reference standards (the pathology of the explanted liver in case of transplantation.;the histology of resected focal liver lesion(s), or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months). The solid circles represent the summary estimates of sensitivity and specificity for ultrasound (black circle) and AFP cut-off 20 ng/ml (red circle)." The dotted lines represent the 95% confidence regions.



For the direct comparison between the two tests, 15 studies provided data assessing AFP measurement with a cut-off value of 20 ng/mL and abdominal US (Okazaki 1984; Cottone 1988; Maringhini 1988; Tremolada 1989; Sherman 1995; Chalasani 1999; Gambarin-Gelwan 2000; Wong 2008; Singal 2012; Raff 2014; Chang 2015; Ungtrakul 2016; Atiq 2017; Kim 2019b; Yang 2019). We found that four studies (Sherman 1995; Raff 2014; Atiq 2017; Yang 2019) reported data obtained in different participants for the two index tests. For this reason, we excluded them from the direct comparison analysis. Thus, we included 11 studies with 6674 participants allowing a direct comparison (Figure 11). By using the bivariate model, we obtained the following pooled estimates: for AFP (cut-off value 20 ng/mL), sensitivity 64% (95% CI 56% to 71%) and specificity 89% (95% CI 79% to 94%); for US, sensitivity 81% (95% CI 66% to 90%) and specificity 92% (95% CI 83% to 97%). The sensitivity of US was higher (P = 0.0044; relative sensitivity 1.27, 95% CI 1.06 to 1.49) while the specificities did not differ (P = 0.3861; relative specificity 1.04, 95% CI 0.95 to 1.12).

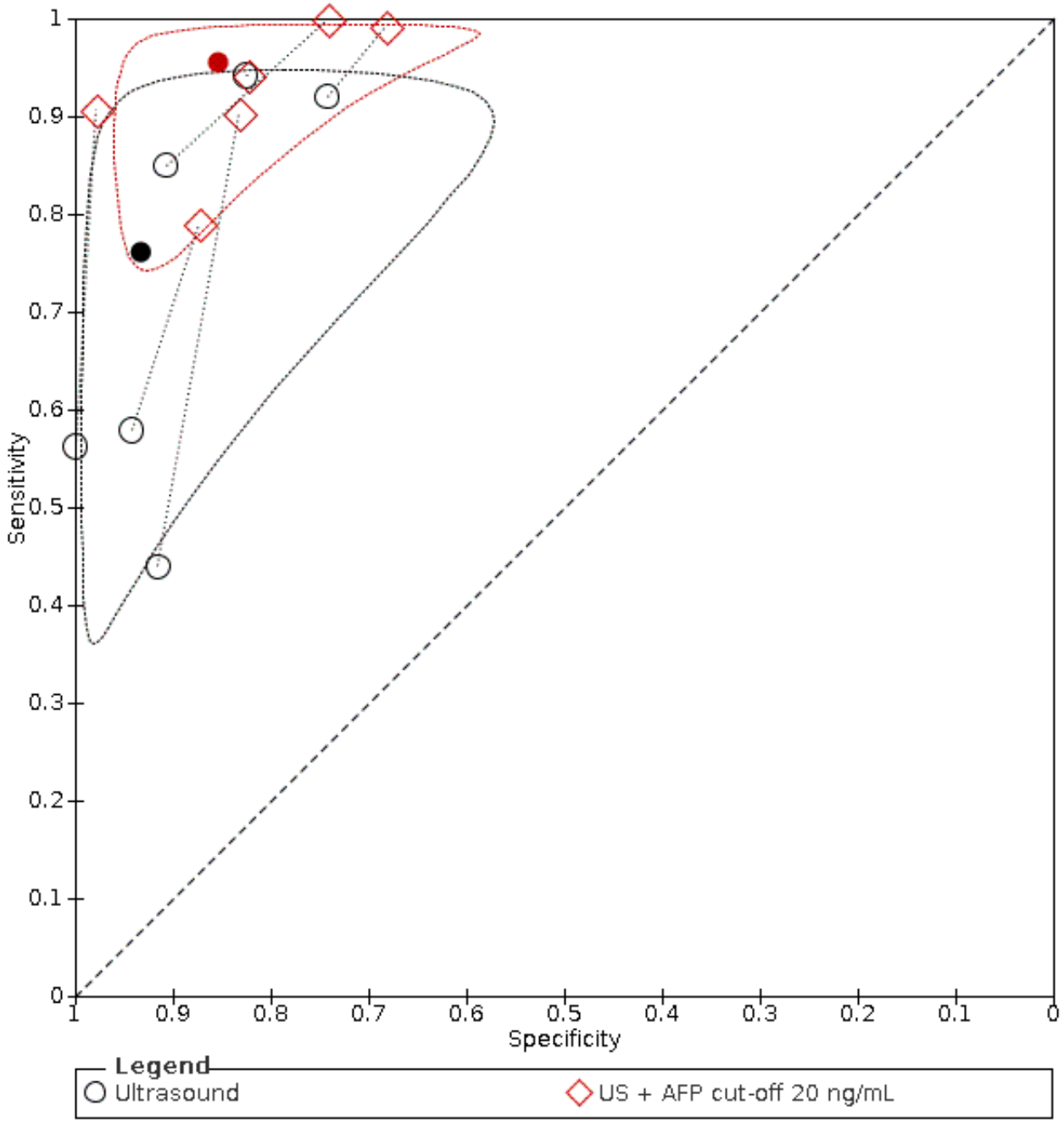
Figure 11. Summary receiver operating characteristic (ROC) showing the direct comparison (within study) of the results of two different index tests, alpha-foetoprotein with a cut-off value around 20 ng/mL (black circles) and ultrasound (red diamonds)in the same participants against the same reference standards (the pathology of the explanted liver in case of transplantation.;the histology of resected focal liver lesion(s), or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months). The solid circles represent the summary estimates of sensitivity and specificity for AFP, with cut-off around 20 ng/ml (black circle) and for US, for direct comparison (red circle)." The dotted lines represent the 95% confidence regions. .



Seven studies provided data assessing either US or the combination of US and AFP with a cut-off 20 ng/mL. After excluding the Raff 2014 study which reported data obtained in different participants, six studies with 5044 participants allowed a direct comparison (Figure 12). By using the bivariate model, we obtained for US a sensitivity of 76% (95% CI 56% to 89%) and a specificity

of 93% (95% CI 80% to 98%); for the combination of US and AFP, a sensitivity of 96% (95% CI 88% to 98%) and a specificity of 85% (95% CI 73% to 92%). The sensitivity of the combination of US and AFP was higher (P = 0.0141; relative sensitivity 1.28, 95% CI 1.03 to 1.53) while the specificity did not differ (P = 0.1024; relative specificity 0.94, 95% CI 0.87 to 1.01) compared with US alone.

Figure 12. Summary receiver operating characteristic (ROC) showing the direct comparison (within study) of the results of two different index tests, ultrasound (black circles) and the combination of alpha-foetoprotein with a cut of value around 20 ng/mL and ultrasound (red diamonds) in the same participants against the same reference standards. Reference standards were: the pathology of the explanted liver in case of transplantation,the histology of resected focal liver lesions, or the histology of biopsied focal liver lesion(s) with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. The solid circles represent the summary estimates of sensitivity and specificity for ultrasound (black circle) and US + AFP cut-off 20 ng/ml (red circle). The dotted lines represent the 95% confidence regions.



Summaryof findings tables

The main results are shown in the Summary of findings 1 and Summary of findings 2.

DISCUSSION

Summary of main results

This review aimed to assess the diagnostic accuracy of abdominal ultrasound (US) and alpha-foetoprotein (AFP), alone or in combination, for the diagnosis of hepatocellular carcinoma (HCC) of any size and at any stage in people with chronic liver disease, either in a surveillance programme or in a clinical setting. The main results are shown in the Summary of findings 1 and Summary of findings 2 tables.

We included 373 studies: 326 studies assessed AFP as the index test in 144,570 participants; 39 studies assessed abdominal US in 18,792 participants; and eight studies assessed both AFP and abdominal US as the index tests in 5454 participants.

We judged only one study (US as the index test) to be at low risk of bias for all four QUADAS-2 domains (Bennett 2002); all the remaining studies were considered to be at high or unclear risk of bias in at least one domain. We also judged most studies (323/373) to be at high concern for the applicability of the results, mainly because of the patient selection domain, as only people with viral aetiology or decompensated liver disease were included, or participants were selected according to volume or other characteristics of the target disease, and because of the reference standard domain, as to confirm the presence of HCC, pathological examination of explanted liver, or of surgical specimen, or necroscopy, or technologies no longer in use, were required.

We summarised the main results of analyses in the Summary of findings 1 and Summary of findings 2. We considered the following consequences of test results: people with true-positive results, i.e. with HCC and positive test results, will receive appropriate further testing and possibly treatment; people with true-negative results, i.e. without HCC and negative test results, will appropriately avoid further testing; people with false-negative results, i.e. with HCC and negative test results, are misdiagnosed and will not receive the appropriate treatment; people with false-positive results, i.e. without HCC and positive test results, will undergo inappropriately further testing with computed tomography (CT), contrast-enhanced ultrasound (CEUS), magnetic resonance imaging (MRI), or biopsy.

The prevalence of HCC varied widely, from 1% to 82%, according to the study design and the different settings. For exemplification, we considered in the 'Summary of findings' tables two different populations: a population at low risk of HCC, with an HCC prevalence of 5%, a value close to that reported by most epidemiological studies (Lok 2009; EASL 2018; Forner 2018); a population at high risk of HCC, with a prevalence of 30%, that is the median of the prevalence in the included cross-sectional studies conducted in clinical cohorts.

Alpha-foetoprotein (AFP)

There was a wide variation in the used cut-off values in the studies with AFP as the index test, and, therefore, we performed a meta-analysis with the hierarchical summary ROC model (HSROC) (Figure 3). There was a considerable heterogeneity in the accuracy estimates, which could in some degree be attributable to the different cut-off values. In order to obtain a pooled estimate of the sensitivity and the specificity, we chose the two most used cut-off values: around 20 ng/mL reported in 147 of 326 studies, and 200 ng/mL reported in 56 studies.

AFP cut-off around 20 ng/mL

For AFP with a cut-off of around 20 ng/mL, performing the meta-analysis with the bivariate model, we obtained the following pooled estimates: sensitivity of 60% (95% CI 58% to 62%) and specificity of 84% (95% CI 82% to 86%). Considering a hypothetical cohort of 1000 people with an HCC prevalence of 5%, we can expect 20 false-negative and 148 false-positive results; with a prevalence of 30%, we can expect 121 false-negative and 109 false-positive results (Summary of findings 1).

These results were also consistent with those obtained in a sensitivity analysis considering the studies with a cross-sectional design alone. We found the setting of the studies as a possible source of heterogeneity: we found different results in studies with enrolment from surveillance programmes compared to studies with enrolment from clinical series. We observed some heterogeneity of accuracy estimates between studies (sensitivity, IQR from 53% to 67%; specificity, IQR from 76% to 90%). Altogether, the heterogeneity of the results remained unexplained despite the exploration of many other possible sources. We did not find any difference between studies with cross-sectional and case-control design. Moreover, the results seem consistent in different geographical areas, along the time, according to HCC prevalence and volume, and according to viral or non viral aetiology and severity of the underlying chronic liver disease. The pooled estimates are quite precise with narrow 95% CIs, but all the studies were at high risk of bias and at high concern for applicability, and with a wide inconsistency that could not be explained by the investigation of potential sources. We judged the certainty of evidence as very low.

AFP cut-off value of 200 ng/mL

For the 56 studies on AFP with a cut-off value of 200 ng/mL, performing the meta-analysis with bivariate model, we obtained sensitivity of 36% (95% CI 31% to 41%) and specificity of 99% (95% CI 98% to 99%). Considering a hypothetical cohort of 1000 people with a HCC prevalence of 5%, we can expect 32 false-negative and 10 false-positive results; with a prevalence of 30%, we can expect 195 false-negative and 7 false-positive results (Summary of findings 1). These results were consistent also in a sensitivity analysis of the studies with cross-sectional design alone. We observed some heterogeneity of accuracy estimates between studies (sensitivity, IQR, 23% to 50%; specificity, IQR 97% to 100%). As possible sources of heterogeneity, we found geographical location (studies conducted in different continents) and severity of the underlying liver disease, according to Child-Pugh classification (Table 3). The pooled estimates are quite precise with narrow 95% CIs, but all studies were at high risk of bias and at high concern for applicability, and with a wide inconsistency that could not be

explained by the investigation of potential sources. We judged the certainty of evidence as very low.

Abdominal ultrasound

For the 39 studies using US as the index test, performing the meta-analysis with bivariate model, we obtained the following pooled estimates: sensitivity of 72% (95% CI 63% to 79%) and specificity of 94% (95% CI 91% to 96%). Considering a hypothetical cohort of 1000 people with an HCC prevalence of 5%, we can expect 2 false-negative and 143 false-positive results; with a prevalence of 30%, we can expect 143 false-negative and 42 false-positive results (Summary of findings 1). We observed some heterogeneity of accuracy estimates between studies (sensitivity, IQR 44% to 89%; specificity, IQR 86% to 96%). Our investigation of the potential sources cannot explain this inconsistency of the results. Most studies are at high risk of bias and many at high concern for applicability. The pooled estimates of accuracy have narrow 95% CI. We judged the certainty of evidence as very low.

Combination of AFP and abdominal ultrasound

For the six studies, using a combination of AFP with cut-off value 20 ng/mL and US as index test, the meta-analysis with the bivariate model produced the following pooled estimates: sensitivity of 96% (95% CI 88% to 98%) and specificity of 85% (95% CI 73% to 93%). Considering a hypothetical cohort of 1000 people with a HCC prevalence of 5%, we can expect 2 false-negative and 143 false-positive results; with a prevalence of 30%, we can expect 2 false-negative and 105 false-positive results (Summary of findings 1). All studies are at high risk of bias and many at high concern for applicability. We did not find a considerable inconsistency of the results and imprecision of the estimates with wide confidence intervals 95% CI. We judged the certainty of evidence as low.

Comparisons

We compared the results of the two index tests: AFP and US. We performed a direct (within-study) comparison in 11 studies using US and AFP with a cut-off value of around 20 ng/mL and showing a higher sensitivity of US with similar specificities (Figure 11). An indirect comparison between 147 AFP studies, with a cut-off value of around 20 ng/mL, and 39 US studies showed a higher sensitivity and specificity of US (Figure 10). The direct comparison considering only the six studies, reporting both US and the combination of AFP (cut-off 20 ng/mL) as index test and US performed in the same participants, showed a higher sensitivity of the combination of AFP and US (relative sensitivity 1.28, 95% CI 1.03 to 1.53, P = 0.0141), while the specificities did not differ (relative specificity 0.94, 95% CI 0.87 to 1.01; Figure 12). All studies were at high risk of bias and many at high concern for applicability. We judged the certainty of evidence as low (Summary of findings 2).

Strengths and weaknesses of the review

Strengths and weaknesses of included studies

Overall, the included studies cover a vast time span and a wide geographical distribution including areas with high and low prevalence of chronic liver disease and HCC.

We found more studies using AFP (n = 326) than using US (n = 39), or the combination of AFP and US (n = 8) as the index test. As we anticipated, many studies with biomarkers were conducted with a case-control design, and in order to improve the completeness of our review, we included studies that compared people with known HCC to matched control. The large number of studies allowed us to obtain precise summary estimates of sensitivity and specificity with narrow confidence intervals. On the other hand, we found only 11 studies providing data for a direct (within study) comparison of AFP and US.

An overall quality assessment of the studies showed some common methodological weaknesses. We considered only one study to be at low risk of bias (Bennett 2002). In most studies with AFP as the index test, the design was case-control and the risk of bias was high for patient selection. Furthermore, different cut-off values were used, ranging from 5 ng/mL to 1000 ng/mL, and these were rarely predefined. The choice of the reference standard was also a major concern for all studies, either with AFP or US, or the combination of AFP and US as index test. The most used reference standard was CT or MRI, or their combination (as also recommended by most clinical guidelines; (Omata 2017; EASL 2018; Heimbach 2018)), but these tests cannot be regarded as absolutely accurate. Another choice of a reference standard was the histology of focal lesion, which is highly specific, but not sensitive, especially for small lesions, and cannot be obtained in the participants with a negative index test. Lastly, another reference standard is the pathology of the explanted liver which is possible only in studies conducted on participants with advanced and decompensated liver disease on a waiting list for transplantation which does not match the review question. In some studies, an AFP value, higher than 200 ng/mL, 400 ng/mL, or 500 ng/mL was one of the criteria for the reference standard. Moreover, in case-control studies, it was often unclear how the target disease was excluded in control participants. Reporting the time interval between the index test and the reference standard was very rare, and often participants underwent different reference standards according to the results of the index test. Furthermore, US is also considered associated with frequent technical failure and with uninterpretable results: interferences due to extrinsic factors such as interposed bowel, ribs, lung, or ascites, as well as patient factors such as obesity or inability to comply with breathing instructions, severe steatosis or severe parenchymal heterogeneity from advanced cirrhosis may impair visualisation of the liver (Rodgers 2019). Up to 14% of US examination were retrospectively judged as inadequate and only 66.5% as definitely adequate in a study of US quality in a HCC surveillance programme in people with liver cirrhosis (Simmons 2017). We found only three studies that addressed this problem reporting the number of uninterpretable results. Not reporting these technical failures of US examination and excluding them from analyses could have produced an overestimation of test accuracy.

Using QUADAS-2, we judged more than 85% of the included studies at high concern for applicability. The case-control design, adopted in most AFP studies, results in an artefactual mixing of affected and non-affected participants which impairs applicability. However, even in cross-sectional studies, as were most US and combination of AFP and US studies, the inclusion/exclusion criteria and different settings make the included participants different from those targeted by the review question. On the contrary, we judged at low concern most studies for the other two domains, i.e. index test and reference standard.

Finally, many studies did not report all the covariates we planned to assess as possible source of heterogeneity, and this might have impaired both their and our analyses.

Strengths and weaknesses of the review process

Limitations of the search strategy

Our search strategy allowed us to obtain a large number of studies that were conducted in various countries, showing a widespread implementation globally of the index tests, and confirming the clinical relevance of the review question. In order to improve the completeness of our review, we planned to include even studies with case-control design that are considered to be at high risk of bias due to inflated accuracy estimates and could have been excluded. Most studies on biomarkers, such as AFP, are conducted with case-control design and indeed, almost 80% of the included AFP studies were case-control studies. Interestingly, their results were not different from those obtained by cross-sectional studies. Furthermore, we included many studies in which AFP was not used as the index test but as the comparator to some other biomarker, and this choice might arguably make publication bias less probable. We identified seven studies through manual searching of the references of the included studies or of previous reviews, and we are confident that we have included most, if not all, of the includable published studies. We applied no language restrictions in the inclusion criteria, and we retrieved 20 full-text studies published in non-English languages, of which we included six studies.

Quality assessment and data extraction

We considered our attempts to reduce subjectivity in our judgments and to minimise errors and miscalculations in data extraction as a strength of this review. According to the protocol plan, two review authors independently assessed the risk of bias of included studies and applicability of their results, using QUADAS-2, and completed the data extraction for each included study using a proper form. In case of disagreement, we reached consensus through discussion. Disagreement was more frequent for the assessment of two QUADAS-2 domains: patient selection (19 studies) and reference standard (15 studies). For data extraction, most of the discordances were due to simple miscalculations or typos and easily solved. For 27 studies a discussion was needed. The agreement obtained through discussion by two review authors was further discussed and approved by a third review author. Then the same authors assessed the certainty of evidence using the GRADE approach and the level of agreement was very high.

Limitations in the review analyses

Despite the large number of included studies and participants, and the consequent precision of accuracy estimates, the results of included studies were not consistent. The use of different cut-off values and different setting (surveillance programme compared to clinical series) could explain heterogeneity only in part. Considering only studies with the same AFP cut-off values, the most frequent cut-off values of 20 ng/mL and 200 ng/mL allowed obtaining more consistent estimates.

In studies with AFP with a cut-off of 20 ng/mL only, we found that study setting was another source of heterogeneity: studies conducted in a surveillance programme compared to those conducted in a clinical setting showed different pooled estimates, with a lower sensitivity and higher specificity in the former. We expected that studies conducted in a surveillance programme would obtain more consistent results: inclusion and exclusion criteria were clear and standardised, such as the index test, reference standard, and timing, whereas, in a clinical setting more variability was expected as participants may have different concurrent disease, different severity of the underlying chronic liver disease, and different stage of the detected HCC. Arguably, in a surveillance programme the underlying liver disease is less severe, and HCCs are smaller. Despite these considerations, we did not plan a separate analysis for the two settings as they are not so clearly distinct in the actual clinical practice (Poustchi 2011; Forner 2018). The two index tests, particularly US, are part of the routine evaluation of people with liver disease; HCC, the target disease, induces no symptom and is usually asymptomatic, thus the clinical suspect of HCC is based only on the presence of a chronic advance liver disease. On the other hand, we found no difference according to the study settings in studies with AFP cut-off value of 200 ng/mL, or with US.

As 80% of hepatocellular carcinoma occurrences occur in sub-Saharan Africa and eastern Asia, we expected that study geographical location could be a source of heterogeneity (Bray 2018). The sensitivity was different in studies conducted across continents in studies with AFP cut-off of 200 mg/mL. The severity of the underlying liver disease, as expressed by the percentage of participants with Child-Pugh class A, could also provide explanation of the heterogeneity of results. The sensitivity was lower in studies with AFP cut-off of 200 ng/mL and including more than 50% of patients with Chid-Pugh class A, i.e. participants with less severe liver disease.

Despite the availability of an adequate number of studies, we were unable to demonstrate any role of aetiology of the underlying chronic liver disease and of the HCC characteristics (volume, resectability). Most studies, conducted either in a surveillance programme or in a clinical setting included inconsistent mixture of participants at different risk of HCC, as shown by the large variability of the prevalence, and we were unable to show the role of the individual characteristics of participants. We could investigate only characteristics that could be assessed at a study level whereas patients' factors or HCC characteristics can be assessed only by aggregate statistics with the inherent risk of ecological bias. Thus, some important relationship such as that with the HCC volume could have been missed. In addition, many of the included studies did not report data on the covariates of our interest. Also, we could not evaluate variability associated to test interpretation, particularly for US which is considered dependent on a subjective judgment. We checked the presence of a definition of US positivity criteria in single studies but not their stringency, apart from their subjective interpretation. We were also unable to assess the effect of uninterpretable results which should be relevant for US due to frequent technical failures. We found only two studies reporting the number of uninterpretable results and could not conduct the planned analysis according to the intention-to-diagnose principle. Moreover, we cannot exclude that most of the studies did not report uninterpretable results and excluded them from analyses, thus inflating the accuracy estimates.

In any case, the sensitivity analyses show that the obtained results are arguably robust, with no variation, after excluding studies published in abstract form or studies with case-control design. As we conducted the analyses of AFP studies using the two most frequent cut-off values of 20 ng/mL and 200 ng/mL, we considered

unnecessary to conduct the planned analysis excluding studies without a predefinition of a cut-off value.

Within- and between-study comparisons

In order to assess any difference in the accuracy of the three index tests (AFP, US, and the combination of AFP and US) ,we planned and performed a direct (or within-study) comparison. After the exclusion of the four studies that reported data for two or three index tests obtained in different numbers of participants (Sherman 1995; Raff 2014; Atiq 2017; Yang 2019), we could do a direct comparison with 11 primary studies with AFP and US, and with six studieswith US, and a combination of AFP and US. The US sensitivity was higher than that of AFP at a cut-off of 20 ng/mL, with comparable specificity. Also, with the combination of AFP (cut-off 20 ng/mL) and US, the sensitivity increased, in comparison to US alone, from 74% to 93% with comparable specificity. These results were confirmed by the indirect (between-study) comparisons which were possible in a greater number of studies (146 with AFP and 39 with US). This between-study comparison, including a greater number of studies, and hence with more power to detect any difference and with more precise results, has a high risk of confounding due to differences in population characteristics, reference standards, and study design.

Comparison with previous research

We found seven reviews on the same topic (Colli 2006; Tateishi 2008; Singal 2009; Kansagara 2014; Singal 2014; Chou 2015; Tzartzeva 2018). Two of these compared imaging techniques for the diagnosis of HCC (Colli 2006; Chou 2015), one assessed only AFP (Tateishi 2008), and four reviews focused mainly on the effectiveness of surveillance programmes with US and AFP (Singal 2009; Kansagara 2014; Singal 2014; Tzartzeva 2018). With our search, we could include many more studies for each index test, and differently from the other reviews, we explored the accuracy of AFP, US, and the combination of AFP and US in the clinical pathway as the first diagnostic step, either in clinical setting or surveillance programme. Due to differences in the methodologic approach, in the inclusion/exclusion criteria, and in the statistical analyses, the results are not comparable to each other and to our results. Colli 2006 reports the results from 14 studies published before January 2005, and the summary estimate of US sensitivity was 60% and specificity 97%. Both Chou 2015 and Tzartzeva 2018, pooling the results of more recent 15 studies, found a US sensitivity higher than 75% and specificity higher than 90%, more similar to our findings. According to Tzartzeva 2018, the accuracy of combining US and AFP improves the diagnostic accuracy with a sensitivity of 97%, but for the detection of early HCC it remains close to 60%.

Applicability of findings to the review question

The review question has broad inclusion criteria, and the consequent large heterogeneity of the results allows exploration of variation in accuracy across various settings, different patient groups or variations in index test, and reference standard application. Using the QUADAS-2 tool, we judged many studies at high concern for applicability in the participant selection domain. In fact, most AFP studies (77%) were case-control studies with an artefactual mixing of affected and non-affected participants. However, even in cross-sectional studies, the prevalence of the target disease ranged from 1% to 82%, as consequences of different settings and variable inclusion criteria often did not match the review question. On the other hand, we judged all studies to be at low concern for applicability in the index test domain. For the reference standard domain, we judged the studies using as reference standard the pathology of the explanted liver to be at high concern. This reference standard, even if perfectly accurate, cannot match the review question as it is applicable only to participants in a waiting list for a liver transplantation.

AUTHORS' CONCLUSIONS

Implications for practice

Hepatocellular carcinoma (HCC) is a frequent complication of chronic liver disease. The detection of a tumour amenable to surgical resection, thermal ablation, or liver transplantation could improve the prognosis which in the absence of indications to radical treatment is severe. Being the fourth leading cause of death from cancer worldwide, accurate tests are needed to diagnose HCC, either in a surveillance programme or in a clinical setting. In the clinical pathway for the diagnosis of HCC in people with chronic liver disease, AFP and US are the first step investigations. Both tests, in separate or in combination, can be considered as triage tests. Ideally, they should ensure a low proportion of false-negative results because people with undetected HCC cannot receive proper treatment. False-positive results would have less severe consequences as misclassified people would undergo unnecessary further testing with CT, MRI, or rarely biopsy.

In surveillance programmes for HCC in high risk patients, the pooled sensitivity of alpha-foetoprotein (AFP) measurement, with a cut-off value of 20 mg/mL, suggests that using this test alone, a relevant number of HCC occurrences would be missed. The estimated sensitivity of ultrasound (US) is higher, but again more than a quarter of HCC occurrences would be missed. The combination of the two tests, considered positive when at least one is positive, reduces the false-negative ratio to around 5%, sparing further testing in case of negative results. The cost of the improvement of the sensitivity is an increased number of false-positive results from 6% to 15%**.** Moreover, our findings suggest that US sensitivity decreases for the diagnosis of potentially resectable HCC.

In a clinical setting, where the pre-test probability of having an HCC is expected to be higher than in surveillance programmes, both US and AFP, with a cut-off value of 20 ng/mL, have an estimated specificity higher than 80%, AFP with a cut-off value of 200 ng/mL, allows confirmation of the diagnosis with a specificity even around 99%. In any case, further testing is required for staging the disease and planning appropriate treatment. However, the role of these two tests is mainly as triage tests, but they individually do not ensure an adequate sensitivity. In particular, AFP is higher than 200 mg/mL only in 36% of patients with HCC. Therefore, clinicians cannot avoid further testing in case of negative results. In this context, the role of the combination of AFP and US cannot be assessed as we found only one study with pathology of explanted liver as reference standard.

Overall, caution is needed in interpreting our review results as we found large heterogeneity which is not due to a few outliers, and despite the investigation of multiple potential factors, heterogeneity remains unexplained. Furthermore, all studies were at high risk of bias, and most of them with high concern regarding their applicability, mainly due to participant selection domain.

Implications for research

As the evidence of the accuracy of AFP, US, and especially of the combination of AFP and US is not conclusive, further studies are needed. In order to obtain more consistent and applicable results; these studies should assess the sensitivity and specificity of AFP and US in people with chronic liver disease at a definite risk for HCC, with a cross-sectional design, evaluating either participants with positive or negative results of the index test with computed tomography (CT) or magnetic resonance imaging (MRI) as the reference standard. This reference standard, even if not absolutely accurate, should be chosen as in the clinical pathway both AFP and US tests play the role of a triage test, just before CT and MRI tests. The time interval between the index test and the reference standard should be clearly reported and should not exceed three months. The number of uninterpretable results should be reported at least for US due to their not negligible frequency. Moreover, no further study with a case-control design can be expected to be informative.

To explore the possible role of these tests on patient relevant outcomes, beyond their accuracy, studies with different designs are needed (Colli 2014). Only randomised clinical trials assessing the overall mortality in different surveillance programmes including these tests in separate or in combination could properly answer this question.

ACKNOWLEDGEMENTS

The review authors thank:

Dimitrinka Nikolova for continuous help during the review process
Sarah Louise Klingenberg for assistance with the search strategies and electronic searches

Jaewoo Cha Ph.D student Department of Preventive Medicine, College of Medicine, Korea University Institute for Evidence-based Medicine, Cochrane Korea Inchon-ro 73, Seongbuk-gu, Seoul, 02841, Korea for his help in translating some studies from Korean

Naoya Sakamoto, MD/PhD Department of Gastroenterology and Hepatology/ Faculty of Medicine/Graduate School of Medicine Hokkaido University North 15, West 7, Kita-ku, Sapporo, Hokkaido 060-8638, Japan, for his help in translating some studies from Japanese

De Zhao Kong MD, Liaoning University of Traditional Chinese Medicine Shenyang China for her help in translating some studies from Chinese

Cochrane Review Group funding acknowledgement: the Danish State is the largest single funder of the Cochrane Hepato-Biliary Group through its investment in the Copenhagen Trial Unit, Centre for Clinical Intervention Research, the Capital Region of Denamrk, Rigshospitalet, Copenhagen, Denmark. Disclaimer: the views and opinions expressed in this review are those of the authors and do not necessarily reflect those of the Danish State or the Copenhagen Trial Unit.

Cochrane Diagnostic Accuracy Reviews Editorial Team
Peer reviewers: Mia Schmidt-Hansen, UK; Sophie Robinson, UK; Francesca Chappell, UK
Contact editors: Sue Mallett, UK; Karen Steingart, UK

Cochrane Hepato-Biliary Group Editorial Team
Contact Editor: Christian Gluud, Denmark
Sign-off Editor: Christian Gluud, Denmark

REFERENCES

References to studies included in this review

Abdel-Aziz 2016 *{published data only}*

Abdel-Aziz MM, Elshal MF, Abass AT, El-Kafrawy S, Ezzat S, Abdel-Wahab. Comparison of AFP-L3 and p53 antigen concentration with alpha-fetoprotein as serum markers for hepatocellular carcinoma. *Clinical Laboratory* 2016;**62**(6):1121-9.

Abdelghany 2018 *{published data only}*

Abdelghany AM, Rezk NS, Osman MM, Hamid AI, Al-Breedy AM, Abdelsattar HA. Using Lamin B1 mRNA for the early diagnosis of hepatocellular carcinoma: a cross-sectional diagnostic accuracy study. *F1000 Research* 2018;**7**:1339.

Abdel-Hamid 2014 *{published data only}*

Abdel-Hamid NM, Wahid AM, Anbar NH, Helaly T. Estimation of des-gamma-carboxy prothrombin with alpha fetoprotein elevates the diagnostic performance of hepatocellular carcinoma among upper Egyptian hepatitis C patients. *Research Journal of Pharmaceutical, Biological and Chemical Sciences* 2014;**5**(1):821-31.

Abdel-Razik 2016 *{published data only}*

Abdel-Razik A, Elhelaly R, Elzehery R, El-Diasty A, Abed S, Elhammady D, et al. Could serotonin be a potential marker for hepatocellular carcinoma? A prospective single-center observational study. *European Journal of Gastroenterology & Hepatology* 2016;**28**(56):599-605.

Aboelfotoh 2018 *{published data only}*

Aboelfotoh AO, Foda EM, Elghandour AM, Teama NM , Abouzein RA, Mohamed GA. Talin-1, other than a potential marker for hepatocellular carcinoma diagnosis. *United European Gastroenterology Journal* 2018;**6**(8 Suppl):A566.

Abu El Makarem 2011 *{published data only}*

Abu El Makarem MA, Abdel-Aleem A, Ali A, Saber R, Shatat M, Rahem DA, et al. Diagnostic significance of plasma osteopontin in hepatitis C virus-related hepatocellular carcinoma. *Annals of Hepatology* 2011;**10**(3):296-305.

Ahmed Mohamed 2016 *{published data only}*

Ahmed Mohamed A, El-Toukhy N, Alkhalegy AA Boraii S. Osteopontin as a tumor marker for hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology Research* 2016;**5**(4):2140-6.

Ahn 2016 *{published data only}*

Ahn DG, Kim HJ, Kang H, Lee HW, Bae SH, Lee JH, et al. Feasibility of alpha-fetoprotein as a diagnostic tool for hepatocellular carcinoma in Korea. *Korean Journal of Internal Medicine* 2016;**31**(1):45-53.

Alexander 1978 *{published data only}*

Alexander MG, Purves LR, Kirsch RE, Bass NM, Gitlin N, Terblanche J, et al. Alpha-Fetoprotein in liver disease. *South African Medical Journal* 1978;**53**(12):433-6.

Ali 2019 *{published data only}*

Ali LH, Higazi AM, Moness HM, Farag NM, Saad ZM, Moukareb HA, et al. Clinical significances and diagnostic utilities of both miR-215 and squamous cell carcinomaantigen-IgM versus alpha-fetoprotein in Egyptian patients with hepatitis C virus-induced hepatocellular carcinoma. *Clinical and Experimental Gastroenterology* 2019;**12**:51-66.

Almani 2004 *{published data only}*

Almani SA, Memon AS, Ghori RA, Memon J, Qureshi AF. Alpha-fetoprotein as a diagnostic tool in differentiating hepatocellular carcinoma from benign hepatic disorders. *Liaquat University of Medical & Health Sciences* 2004;**3**(1):7-12.

Alpert 1971 *{published data only}*

Alpert E, Hershberg R, Shur PH, Isselbacher KJ. Alpha - fetoprotein in human hepatoma: improved detection in serum, and quantitative studies using a new sensitive technique. *Gastroenterology* 1971;**61**(2):137-43.

Alsebaey 2016 *{published data only}*

Alsebaey A, Ahmedy EA. Talin-1 and non-invasive fibrosis models in the assessment of patients with hepatocellular carcinoma. *Asian Pacific Journal of Cancer Prevention* 2016;**17**(8):4077-82.

Al-Zoubi 2017 *{published data only}*

Al-Zoubi S, Wassouf A, Zetoune AB. Measuring levels of osteopontin as potential biomarker for hepatocellular carcinoma in Syrian patients. *Gastroenterology and Hepatology from Bed to Bench* 2017;**10**(2):97-101.

Amuro 1988 *{published data only}*

Amuro Y, Nakaoka H, Shimomura S, Fujikura M, Yamamoto T, Tamura S, et al. Serum pseudouridine as a biochemical marker in patients with hepatocellular carcinoma. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 1988;**178**(2):151-8.

Arrieta 2007 *{published data only}*

Arrieta O, Cacho B, Morales-Espinosa D, Ruelas-Villavicencio A, Flores-Estrada D, Hernández-Pedro N. The progressive elevation of alpha fetoprotein for the diagnosis of hepatocellular carcinoma in patients with liver cirrhosis. *BMC Cancer* 2007;**7**:28-36.

Arrigoni 1988 *{published data only}*

Arrigoni A, Andrulli A, Gindro T, Piantino P, Capussotti L, Rizzetto M. Pattern analysis of serum alpha-fetoprotein in the early diagnosis of hepatocellular carcinoma in liver cirrhosis. *International Journal of Biological Markers* 1988;**3**(3):172-6.

Atiq 2017 *{published data only}*

Atiq O, Tiro J, Yopp AC, Muffler A, Marrero JA, Parikh ND, et al. An assessment of benefits and harms of hepatocellular carcinoma surveillance in patients with cirrhosis. *Hepatology (Baltimore, Md.)* 2017;**65**(4):1196-205.

Attallah 2011 *{published data only}*

Attallah AM, El-Far M, Abdel Malak CA, Zahran F, Farid K, Omran MA, et al. Evaluation of cytokeratin-1 in the diagnosis of hepatocellular carcinoma. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2011;**412**(23-24):2310-5.

Attallah 2013 *{published data only}*

Attallah AM, Omran MM, Attallah AA, Abdallah SO, Farid K, Darwish H, et al. HCC-ART score, a simple, highly sensitive and specific test for early diagnosis of hepatocellular carcinoma: a large-scale, multicentre study. *British Journal of Cancer* 2013;**109**:1657-65.

Attallah 2017 *{published data only}*

Attallah AM, Omran MM, Attallah AA, Abdelrazek MA, Farid K, El-Dosoky I. Simplified HCC-ART score for highly sensitive detection of small-sized and early-stage hepatocellular carcinoma in the widely used Okuda, CLIP, and BCLC staging systems. *International Journal of Clinical Oncology* 2017;**22**:332-9.

Attallah 2018 *{published data only}*

Attallah A, Abd M, El-Far E-H, Omran M, Mohamed A, Elbendary M, et al. Comparison between glypican-3 and alpha-fetoprotein in discrimination of hepatocellular carcinoma from cirrhotic patients. *Journal of Bioscience and Applied Research* 2018;**4**(4):459-68.

Attallah 2020 *{published data only}*

Attallah AM, Albannan MS, El-Deen MS, Farid K, Khedr FM, Attallah KA, et al. Diagnostic role of collagen-III and matrixmetalloproteinase-1 for early detection of hepatocellular carcinoma. *British Journal of Biomedical Science* 2020;**77**(2):58-63.

Bachtiar 2009 *{published data only}*

Bachtiar I, Santoso JM, Atmanegara B, Gani RA, Hasan I, Lesmana LA, et al. Combination of alpha-1-acid glycoprotein and alpha-fetoprotein as an improved diagnostic tool for hepatocellular carcinoma. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2009;**399**:97-101.

Badr 2014 *{published data only}*

Badr EA, Korah TE, Ghani AA, El-Sayed S, Badr S. Role of serum glypican-3 in the diagnosis and differentiation of small hepatocellular carcinoma from hepatitis-C virus cirrhosis. *Alexandria Journal of Medicine* 2014;**50**:221-6.

Baek 2009 *{published data only}*

Baek YH, Lee JH, Jang JS, Lee SW, Han JY, Jeong JS, et al. Diagnostic role and correlation with staging systems of PIVKA-II compared with AFP. *Hepato-gastroenterology* 2009;**56**(91-92):763-7.

Bell 1982 *{published data only}*

Bell H. Alpha-fetoprotein and carcinoembryonic antigen in patients with primary liver carcinoma, metastatic liver disease, and alcoholic liver disease. *Scandinavian Journal of Gastroenterology* 1982;**17**(7):897-903.

Beneduce 2004 *{published data only}*

Beneduce L, Castaldi F, Marino M, Tono N, Gatta A, Pontisso P, et al. Improvement of liver cancer detection with simultaneous assessment of circulating levels of free alpha-fetoprotein (AFP) and AFP-IgM complexes. *International Journal of Biological Markers* 2004;**19**(2):155-9.

Beneduce 2008 *{published data only}*

Beneduce L, Pesce G, Gallotta A, Zampieri F, Biasiolo A, Tono N, et al. Tumour-specific induction of immune complexes: DCP-IgM in hepatocellular carcinoma. *European Journal of Clinical Investigation* 2008;**38**(8):571-7.

Bennett 2002 *{published data only}*

Bennett GL, Krinsky GA, Abitbol RJ, Kim SY, Theise ND, Teperman LW. Sonographic detection of hepatocellular carcinoma and dysplastic nodules in cirrhosis:Correlation of pre transplantation sonography and liver explant pathology in 200 patients. *AJR. American Journal of Roentgenology* 2002;**179**:75-80.

Bessa 2010 *{published data only}*

Bessa SS, Elwan NM, Suliman GA, El-Shourbagyd SH. Clinical significance of plasma osteopontin level in Egyptian patients with hepatitis C virus-related hepatocellular carcinoma. *Archives of Medical Research* 2010;**41**:541-7.

Best 2016 *{published data only}*

Best J, Bilgi H, Heider D, Schotten C, Manka P, Bedreli S, et al. The GALAD scoring algorithm based on AFP, AFP-L3, and DCPsignificantly improves detection of BCLC early stage hepatocellular carcinoma [Der GALAD-Score, ein AFP-, AFP-L3- und DCP-basierter Diagnosealgorithmusverbessert die Detektionsrate des hepatozellulärenKarzinoms im BCLC-Frühstadium signifikant]. *Zeitschrift fur Gastroenterologie* 2016;**54**:1296-305.

Best 2020 *{published data only}*

Best J, Bechmann LP, Sowa JP, Sydor S, Dechêne A, Pflanz K, et al. GALAD score detects early hepatocellular carcinoma in an international cohort of patients with nonalcoholic steatohepatitis. *Clinical Gastroenterology and Hepatology* 2020;**18**(3):728-35.

Biselli 2015 *{published data only}*

Biselli M, Conti F, Gramenzi A, Frigerio M, Cucchetti A, Fatti G, et al. A new approach to the use of a-fetoprotein as surveillance test for hepatocellular carcinoma in patients with cirrhosis. *British Journal of Cancer* 2015;**112**(1):69-76.

Bolondi 2001 *{published data only}*

Bolondi L, Sofia S, Siringo S, Gaiani S, Casali A, Zironi G, et al. Surveillance programme of cirrhotic patients for early diagnosis and treatment of hepatocellular carcinoma: a cost effectiveness analysis. *Gut* 2001;**48**:251-9.

Bon 1998 *{published data only}*

Bon C, Brillard B. Decarboxyprothrombin: importance in the diagnosis of hepatocellular carcinoma [La decarboxyprothrombine: interert dans la diagnostic du

carcinoma hepatocellular]. *Annales de Biologie Clinique* 1998;**56**(2):175-81.

Brunello 1993 *{published data only}*

Brunello F, Marcarino C, Pasquero P, Gastaldi P, Gonella S, Martini S, et al. The des-gamma-carboxyprothrombin for the diagnosis of hepatocellular carcinoma. *Italian Journal of Gastroenterology* 1993;**25**(1):9-12.

Buffet 1988 *{published data only}*

Buffet G, Prades P, Hagege H, Caquil P, Ink O, Etienne JP. Focal ultrasonographic lesion on the liver in cirrhotic patients always reflects hepatocellular carcinoma? 217 cases [La lesion echographique focalisee du foie chez le cirrhotique correspond-t-elle toujours a un carcinome hepatocellulaire? 217 observations]. *La Presse Medicale* 1988;**17**(32):1629-32.

Cabrera 2012 *{published data only}*

Cabrera R, Fitian AI, Ararat M, Xu Y, Brusko T, Wasserfall C, et al. Serum levels of soluble CD25 as a marker for hepatocellular carcinoma. *Oncology Letters* 2012;**4**:840-6.

Capurro 2003 *{published data only}*

Capurro M, Wanless IR, Sherman M, Deboer G, Shi W, Miyoshi E, et al. Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology* 2003;**125**(1):89-97.

Caviglia 2016 *{published data only}*

Caviglia GP, Abate ML, Petrini E, Gaia S, Rizzetto M, Smedile A. Highly sensitive alpha-fetoprotein, Lens culinaris agglutinin reactive fraction of alpha-fetoprotein and des-gammacarboxyprothrombin for hepatocellular carcinoma detection. *Hepatology Research* 2016;**46**:E130-5.

Caviglia 2017 *{published data only}*

Caviglia GP, Abate ML, Gaia S, Petrini E, Bosco C, Olivero A, et al. Risk of hepatocellular carcinoma in HBV cirrhotic patients assessed by the combination of miR-122, AFP and PIVKA-II. *Panminerva Medica* 2017;**59**(4):283-9.

Cedrone 2000 *{published data only}*

Cedrone A, Covino M, Caturelli E, Pompili M, Lorenzelli G, Villani MR, et al. Utility of alpha-fetoprotein (AFP) in the screening of patients with virus-related chronic liver disease: does different etiology influence AFP levels in HCC? A study in 350 western patients. *Hepato-gastroenterology* 2000;**47**(36):1654-8.

Chalasani 1999 *{published data only}*

Chalasani N, Horlander JC Sr, Said A, Hoen H, Kopecky KK, Stockberger SM Jr, et al. Screening for hepatocellular carcinoma in patients with advanced cirrhosis. *American Journal of Gastroenterology* 1999;**94**(10):2988-93.

Chan 2013 *{published data only}*

Chan HLY, Wong GL, Lo Angeline OS, Wong VW. Revisit of alpha-fetoprotein as a tumor marker for hepatocellular carcinoma in chronic hepatitis B patients receiving entecavir. *Hepatology (Baltimore, Md.)* 2013;**58**(4 Suppl 1):1256A.

Chan 2014 *{published data only}*

Chan SL, Mo F, Johnson PJ, Siu DY, Chan MHM, Lau WY, et al. Performance of serum a-fetoprotein levels in the diagnosis of hepatocellular carcinoma in patients with a hepatic mass. *HPB : the Official Journal of the International Hepato Pancreato Biliary Association* 2014;**16**:366-72.

Chang 1988 *{published data only}*

Chang WY, Chuang WL. Complements as new diagnostic tools of hepatocellular carcinoma in cirrhotic patients. *Cancer* 1988;**62**(2):227-32.

Chang 2015 *{published data only}*

Chang TS, Wu YC, Tung SY, Wei KL, Hsieh YY, Huang HC, et al. Alpha-fetoprotein measurement benefits hepatocellular carcinoma surveillance in patients with cirrhosis. *American Journal of Gastroenterology* 2015;**110**(6):836-44.

Chayvialle 1977 *{published data only}*

Chayvialle JA, Brissot P, Pelletier MJ, de Nercy JH, Lambert R, Bourel M. Alpha-Fetoprotein screening in patients with idiopathic hemochromatosis and liver cirrhosis. *Digestion* 1977;**16**(1-2):118-27.

Chen 1977 *{published data only}*

Chen DS, Sung JL. Serum alphafetoprotein in hepatocellular carcinoma. *Cancer* 1977;**40**:779-83.

Chen 2003 *{published data only}*

Chen JG, Parkin DM, Chen QG, Lu JH, Shen QJ, Zhang BC, et al. Screening for liver cancer: results of a randomised controlled trial in Qidong, China. *Journal of Medical Screening* 2003;**10**:204-9.

Chen 2015 *{published data only}*

Chen L, Chu F, Cao Y, Shao J, Wang F. Serum miR-182 and miR-331-3p as diagnostic and prognostic markers in patients with hepatocellular carcinoma. *Tumour Biology* 2015;**36**:7439-47.

Chen 2018 *{published data only}*

Chen H, Zhang Y, Li S, Li N, Chen Y, Zhang B, et al. Direct comparison of five serum biomarkers in early diagnosis of hepatocellular carcinoma. *Cancer Management and Research* 2018;**10**:1947-58.

Cheng 2012 *{published data only}*

Cheng J, Gholami P, Kappus M, Sterling R. Evaluation of diagnostic performance of alpha-fetoprotein (AFP), AFP-l3 and des-gamma-carboxy prothrombin (DCP) for HCV and non-HCV related hepatocellular carcinoma (HCC) after long-term follow-up. *American Journal of Gastroenterology* 2012;**107**:S145-6.

Chimparlee 2015 *{published data only}*

Chimparlee N, Chuaypen N, Khlaiphuengsin A, Pinjaroen N, PayungpornS, Poovorawan Y, et al. Diagnostic and prognostic roles of serum osteopontin and osteopontin promoter polymorphisms in hepatitis b-related hepatocellular carcinoma. *Asian Pacific Journal of Cancer Prevention* 2015;**16**(16):7211-7.

Choi 2019 *{published data only}*

Choi J, Kim GA, Han S, Lee W, Chun S, Lim YS. Longitudinal assessment of three serum biomarkers to detect very early-stage hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2019;**69**(5):1983-94.

Chuaypen 2018 *{published data only}*

Chuaypen N, Chittmittraprap S, Pinjaroen N, Sirichindakul B, Poovorawan Y, Tanaka Y, et al. Serum Wisteria floribunda agglutinin-positive Mac-2 binding protein level as a diagnostic marker of hepatitis B virus related hepatocellular carcinoma. *Hepatology Research* 2018;**48**:872-81.

Cottone 1983 *{published data only}*

Cottone M, Marcenò MP, Maringhini A, Rinaldi F, Russo G, Sciarrino E, et al. Ultrasound in the diagnosis of hepatocellular carcinoma associated with cirrhosis. *Radiology* 1983;**147**:517-9.

Cottone 1988 *{published data only}*

Cottone M, Turri M, Caltagirone M, Maringhini A, Sciarrino E, Virdone R, et al. Early detection of hepatocellular carcinoma associated with cirrhosis by ultrasound and alpha fetoprotein: a prospective study. *Hepato-gastroenterology* 1988;**35**:101-3.

Cui 2002 *{published data only}*

Cui R, Wang B, Ding H, Shen H, Li Y, Chen X. Usfulness of determining a protein induced by vitamin K absence in detection of hepatocellular carcinoma. *Chinese Medical Journal* 2002;**115**(1):42-5.

Cui 2003 *{published data only}*

Cui R, He J, Zhang F, Wang B, Ding H, Shen H, et al. Diagnostic value of protein induced by vitamin K absence (PIVKAII) and hepatoma-specific band of serum gamma-glutamyl transferase (GGTII) as hepatocellular carcinoma markers complementary to a-fetoprotein. *British Journal of Cancer* 2003;**88**:1878-82.

da Costa 2015a *{published data only}*

da Costa AN, Plymoth A, Santos-Silva D, Ortiz-Cuaran S, Camey S, Guilloreau P, et al. Osteopontin and latent-TGF β binding-protein 2 as potential diagnostic markers for HBV-related hepatocellular carcinoma. *International Journal of Cancer. Journal International du Cancer* 2015;**136**:172-81.

da Costa 2015b *{published data only}*

da Costa AN, Plymoth A, Santos-Silva D, Ortiz-Cuaran S, Camey S, Guilloreau P, et al. Osteopontin and latent-TGF β binding-protein 2 as potential diagnostic markers for HBV-related hepatocellular carcinoma. *International Journal of Cancer. Journal International du Cancer* 2015;**136**:172-81.

da Costa 2015c *{published data only}*

da Costa AN, Plymoth A, Santos-Silva D, Ortiz-Cuaran S, Camey S, Guilloreau P, et al. Osteopontin and latent-TGF β binding-protein 2 as potential diagnostic markers for HBV-related hepatocellular carcinoma. *International Journal of Cancer. Journal International du Cancer* 2015;**136**:172-81.

da Costa 2015d *{published data only}*

da Costa AN, Plymoth A, Santos-Silva D, Ortiz-Cuaran S, Camey S, Guilloreau P, et al. Osteopontin and latent-TGF β binding-protein 2 as potential diagnostic markers for HBV-related hepatocellular carcinoma. *International Journal of Cancer. Journal International du Cancer* 2015;**136**:172-81.

Ding 2020 *{published data only}*

Ding Y, Liu K, Xu Y, Zhao Q, Lou S, Xiang X, et al. Combination of inflammatory score/liver function and AFP improves the diagnostic accuracy of HBV-related hepatocellular carcinoma. *Cancer Medicine* 2020;**9**(9):3057-69.

Dodd 1992 *{published data only}*

Dodd GD, Miller WJ, Baron LR, Sclonick ML, Campbell WL. Detection of malignant tumors in end-stage cirrhotic livers: efficacy of sonography as a screening technique. *AJR. American Journal of Roentgenology* 1992;**159**:727-33.

Dong 2015 *{published data only}*

Dong X, He H, Zhang W, Yu D, Wang X, Chen Y. Combination of serum RASSF1A methylation and AFP is a promising non-invasive biomarker for HCC patient with chronic HBV infection. *Diagnostic Pathology* 2015;**10**:133.

Durazo 2008 *{published data only}*

Durazo FA, Blatt LM, Corey WG, Lin JH, Han S, Saab S, et al. Des-gamma-carboxyprothrombin, alpha-fetoprotein and AFP-L3 in patients with chronic hepatitis, cirrhosis and hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology* 2008;**23**(10):1541-8.

Edis 1998 *{published data only}*

Edis C, Kähler C, Klotz W, Herold M, Feichtinger H, Königsreiner A, et al. A comparison between-fetoprotein and p53 antibodies in the diagnosis of hepatocellular carcinoma. *Transplantation Proceedings* 1998;**30**:780-1.

Edoo 2019 *{published data only}*

Edoo MIA, Chutturghoon VK, Wusu-Ansah GK, Zhu H, Zhen TY, Xie HY, et al. Serum biomarkers AFP, CEA and CA19-9 combined detection for early diagnosis of hepatocellular carcinoma. *Iranian Journal of Public Health* 2019;**48**(2):314-22.

Eissa 2013 *{published data only}*

Eissa LA, Eisa NH, Ebrahim MA, Ragab M, El-Gayar AM. Nitric oxide is a potential diagnostic marker for hepatocellular carcinoma. *Scientia Pharmaceutica* 2013;**81**:763-75.

El-Abd 2015 *{published data only}*

El-Abd NE, Fawzy NA, El-Sheikh SM, Soliman ME. Circulating miRNA-122, miRNA-199a, and miRNA-16 as biomarkers for early detection of hepatocellular carcinoma in Egyptian patients with chronic hepatitis C virus infection. *Molecular Diagnosis & Therapy* 2015;**19**(4):213-20.

El-Abd 2016 *{published data only}*

El-Abd N, Fawzy A, Elbaz T, Hamdy S. Evaluation of annexin A2 and as potential biomarkers for hepatocellular carcinoma. *Tumour Biology* 2016;**37**:211-6.

El Gawad 2014 *{published data only}*

El Gawad IA, Mossallam GI, Radwan NH, Elzawahry HM, Elhifnawy NM. Comparing prothrombin induced by vitamin

Kabsence-II (PIVKA-II) with the oncofetal proteins glypican-3, alpha feto protein and carcinoembryonic antigen in diagnosing hepatocellular carcinoma among Egyptian patients. *Journal of the Egyptian National Cancer Institute* 2014;**26**:79-85.

el-Houseini 2005 *{published data only}*

el-Houseini ME, Mohammed MS, Elshemey WM, Hussein TD, Desouky OS, Elsayed AA. Enhanced detection of hepatocellular carcinoma. *Cancer Control* 2005;**12**(4):248-53.

El Mahdy 2019 *{published data only}*

El Mahdy HA, Abdelhamid IA, Amen AI, Abdelsameea E, Hassouna MM. MicroRNA-215 as a diagnostic marker in Egyptian patients with hepatocellular carcinoma. *Asian Pacific Journal of Cancer Prevention* 2019;**20**(9):2723-31.

El Moety 2011 *{published data only}*

El Moety AA, El Moety HA. Evaluation of nitric oxide as a novel diagnostic marker for hepatocellular carcinoma. *Alexandria Journal of Medicine* 2011;**47**:31–5.

Elnemr 2012 *{published data only}*

Elnemr M, Abdel-Azeez H, Labib H, Abo-Taleb F. Clinical relevance of serum endoglin level in Egyptian hepatocellular carcinoma patients. *Clinical Laboratory* 2012;**9-10**:1023-8.

El-Serag 2017 *{published data only}*

El-Serag HB, White D, Tayob N, Alsarray A, Mori Y, Yamada H, et al. Phase 3 biomarker study for HCC surveillance using AFP, AFP-l3 and DCP: a prospective collection with retrospective blinded evaluation. *Hepatology (Baltimore, Md.)* 2017;**66**(Suppl 1):754A.

El Shafie 2012 *{published data only}*

El Shafie MA, Fawzy AM, Al Monem EA, Abbass S, Zakaria DM, El Baz S. Golgi protein 73 (GP73) as a novel serum marker for early detection of hepatocellular carcinoma in Egyptian patients. *Life Science Journal* 2012;**9**(2):823-30.

El-Shenawy 2012 *{published data only}*

El-Shenawy SZ, El-Sabawy MM, El-Razik EA, Allam MA. Detection of serum KL-6 as a tumor marker in hepatocellular carcinoma. *Life Science Journal* 2012;**9**(1):667-73.

El-Sherif 2012 *{published data only}*

El-Sherif WT, Makhlouf NA, El-Gendi SS, Hassan HI, Herdan OM. Evaluation of transforming growth factor alpha and vascular endothelial growth factor in diagnosis of hepatocellular carcinoma. *Egyptian Journal of Immunology / Egyptian Association of Immunologists* 2012;**19**(2):53-65.

Eltaher 2016 *{published data only}*

Eltaher SM, El-Gil R, Fouad N, Mitwali R, El-Kholy H. Evaluation of serum levels and significance of solubleCD40 ligand in screening patients with hepatitis C virus-related hepatocellular carcinoma. *Eastern Mediterranean Health Journal* 2016;**22**(8):603-10.

El-Tayeh 2012 *{published data only}*

El-Tayeh SF, Hussein TD, El-Houseini ME, Amer MA, El-Sherbini M, Elshemey WM. Serological biomarkers of hepatocellular carcinoma in Egyptian patients. *Disease Markers* 2012;**32**:255-63.

El Zefzafy 2015 *{published data only}*

El Zefzafy W, Eltokhy H, El Gaffar Mohamed NA, Abu Zahab Z. Significance of serum cytokeratin-18 in prediction of hepatocellular carcinoma in chronic hepatitis C infected Egyptian patients. *Macedonian Journal of Medical Sciences* 2015;**3**(1):117-23.

Erdal 2016 *{published data only}*

Erdal H, Gül Utku Ö, Karatay E, Çelik B, Elbeg Ş, Doğan İ. Combination of DKK1 and AFP improves diagnostic accuracy of hepatocellular carcinoma compared with either marker alone. *Turkish Journal of Gastroenterology* 2016;**27**:375-81.

Ertle 2013 *{published data only}*

Ertle JM, Heider D, Wichert M, Keller B, Kueper R, Hilgard P, et al. A combination of alpha-fetoprotein and des-gamma-carboxy prothrombin is superior in detection of hepatocellular carcinoma. *Digestion* 2013;**87**(2):121-31.

Ette 2015 *{published data only}*

Ette AI, Ndububa DA, Adekanle O, Ekrikpo U. Utility of serum des-gamma-carboxyprothrombinin the diagnosis of hepatocellular carcinoma among Nigerians, a case–control study. *BMC Gastroenterology* 2015;**15**:113.

Ezzikouri 2015 *{published data only}*

Ezzikouri S, Kimura K, Sunagozaka H, Kaneko S, Inoue K, Nishimura T, et al. Serum DHCR24 auto-antibody as a new biomarker for progression of hepatitis C. *EBioMedicine* 2015;**2**:604-12.

Fabris 1991 *{published data only}*

Fabris C, Basso DA, Leandro G, Meggiato T, Elba S, Panozzo MP, et al. Serum CA 19-9 and alpha-fetoprotein levels in primary hepatocellular carcinoma and liver cirrhosis. *Cancer* 1991;**68**(8):1795-8.

Fang 2010 *{published data only}*

Fang M, Zhao YP, Zhou FG, Lu LG, Qi P, Wang H, et al. N-glycan based models improve diagnostic efficacies in hepatitisB virus-related hepatocellular carcinoma. *International Journal of Cancer. Journal International du Cancer* 2010;**127**:148-59.

Farid 2014 *{published data only}*

Farid IM, Hamza IM, El-Abd DM, Mohyi AM, Abdul Latif MM, Aref AT, et al. Transforming growth factor-b1 gene expression in hepatocellular carcinoma: a preliminary report. *Arab Journal of Gastroenterology* 2014;**15**:142-7.

Feng 2016 *{published data only}*

Feng X, Song P, Bie P, Jiang P, Ma K, Li X, et al. Des-g-carboxyprothrombin plasma level in diagnosis of hepatocellular carcinoma in a Chinese population undergoing surgery. *Medical Science Monitor* 2016;**22**:1663-72.

Fujii1995 *{published data only}*

Fujii T, Horie Y, Ikuta Y, Nishimuki E, Murawaki Y, Suou T, et al. Clinical evaluation of a monoclonal antibody to serum KM01 for

the diagnosis of hepatocellular carcinoma. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 1995;**236**:71-9.

Gad 2005 *{published data only}*

Gad A, Tanaka E, Matsumoto A, Wahab MA, Serwah Ael-H, Attia F, et al. Assessment of KL-6 as a tumor marker in patients with hepatocellular carcinoma. *World Journal of Gastroenterology* 2005;**11**(42):6607-12.

Gambarin-Gelwan 2000 *{published data only}*

Gambarin-Gelwan M, Wolf DC, Shapiro R, Schwartz ME, Min AD. Sensitivity of commonly available screening tests in detecting hepatocellular carcinoma in cirrhotic patients undergoing liver transplantation. *American Journal of Gastroenterology* 2000;**95**(6):1535-8.

Gani 2015 *{published data only}*

Gani RA, Suryamin M, Hasan I, Lesmana CR, Sanityoso A. Performance of alpha fetoprotein in combination with alpha-1-acid glycoprotein for diagnosis of hepatocellular carcinoma among liver cirrhosis patients. *Acta Medica Indonesiana* 2015;**47**(3):216-22.

Garretti 1988 *{published data only}*

Garretti L, Fauciglietti P, Regge D, Bonino F, Brunetto M, Gandini G. Evaluation of the usefulness of ultrasonics in the diagnosis of cancerous cirrhosis. Comparison with CT [Valutazione dell'ultilità dell'ecotomografia nella diagnosi della cancro-cirrosi]. *La Radiologia Medica* 1988;**76**:187-92.

Ge 2015 *{published data only}*

Ge T, Shen Q, Wang N, Zhang Y, Ge Z, Chu W, et al. Diagnostic values of alpha-fetoprotein, dickkopf-1, and osteopontin for hepatocellular carcinoma. *Medical Oncology (Northwood, London, England)* 2015;**32**(3):59.

Gentile 2017 *{published data only}*

Gentile I, Buonomo AR, Scotto R, Zappulo E, Carriero C, Piccirillo M, et al. Diagnostic accuracy of PIVKA-II, alpha-fetoprotein and a combination of both in diagnosis of hepatocellular carcinoma in patients affected by chronic HCV infection. *In Vivo (Athens, Greece)* 2017;**31**(4):695-700.

Giannelli 2005 *{published data only}*

Giannelli G, Marinosci F, Trerotoli P, Volpe A, Quaranta M, Dentico P, et al. SCCA antigen combined with alpha-fetoprotein as serologic markers of HCC. *International Journal of Cancer. Journal International du Cancer* 2005;**117**(3):506-9.

Giannelli 2007 *{published data only}*

Giannelli G, Fransvea E, Trerotoli P, Beaugrand M, Marinosci F, Lupo L, et al. Clinical validation of combined serological biomarkers for improved hepatocellular carcinoma diagnosis in 961 patients. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2007;**383**(1-2):147-52.

Gopal 2014 *{published data only}*

Gopal P, Yopp AC, Waljee AK, Chiang J, Nehra M, Kandunoori P, et al. Factors that affect accuracy of α-fetoprotein test in detection of hepatocellular carcinoma in patients with cirrhosis. *Clinical Gastroenterology and Hepatology* 2014;**12**(5):870–7.

Grazi 1995 *{published data only}*

Grazi GL, Mazziotti A, Legnani C, Jovine E, Miniero R, Gallucci A, et al. The role of tumor markers in the diagnosis of hepatocellular carcinoma, with special reference to the des-gamma-carboxy prothrombin. *Liver Transplantation and Surgery* 1995;**1**(4):249-55.

Guan 2020 *{published data only}*

Guan W, Gao Z, Huang C, Fang M, Feng H, Chen S, et al. The diagnostic value of serum DSA-TRF in hepatocellular carcinoma. *Glycoconjugate Journal* 2020;**37**(2):231-40.

Hallager 2018 *{published data only}*

Hallager S, Ladelund S, Kjaer M, Madsen LG, Belard E, Laursen AL, et al. Hepatocellular carcinoma in patients with chronic hepatitis C and cirrhosis in Denmark: a nationwide cohort study. *Journal of Viral Hepatitis* 2018;**25**(1):47-55.

Han 2014 *{published data only}*

Han LY, Fan YC, Mu NN, Gao S, Li F, Ji XF, et al. Aberrant DNA methylation of G-protein-coupled bile acid receptor Gpbar1 (TGR5) is a potential biomarker for hepatitis B virus associated hepatocellular carcinoma. *International Journal of Medical Sciences* 2014;**11**(2):164-71.

Han 2018 *{published data only}*

Han C, Gao L, Bai H, Dou XG. Identification of a role for serum `aldo-keto` reductase family 1 member B10 in early detection of hepatocellular carcinoma. *Oncology Letters* 2018;**16**:7123-30.

Hu 2018 *{published data only}*

Hu J, Wang N, Yang Y, Ma L, Han R, Zhang W, et al. Diagnostic value of alpha-fetoprotein combined with neutrophil-to-lymphocyte ratio for hepatocellular carcinoma. *BMC Gastroenterology* 2018;**18**(1):186.

Hu 2019 *{published data only}*

Hu Z, Chen H, Chen S, Huang Z, Qin S, Zhong J, et al. The value of neutrophil to lymphocyte ratio and gamma-glutamyl transpeptidase to platelet ratio in patients with hepatocellular carcinoma. *Medicine (Baltimore)* 2019;**98**(9):e14749.

Huo 2007 *{published data only}*

Huo TI, Hsia CY, Chu CJ, Huang YH, Lui WY, Wu JC, et al. The predictive ability of serum alpha-fetoprotein for hepatocellular carcinoma is linked with the characteristics of the target population at surveillance. *Journal of Surgical Oncology* 2007;**15**(8):645-51.

Ibrahim 2013 *{published data only}*

Ibrahim AM, Hashem ME, Mostafa EF, Refaey MM, Hamed EF, Ibrahim I, et al. Annexin A2 versus AFP as an efficient diagnostic serum marker for hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology Reserch* 2013;**2**(9):780-5.

Iizuka 2010a *{published data only}*

Iizuka N, Oka M, Sakaida I, Moribe T, Miura T, Kimura N, et al. Efficient detection of hepatocellular carcinoma by a hybrid blood test of epigenetic and classical protein markers. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2010;**412**(1-2):152-8.

Iizuka 2010b *{published data only}*

Iizuka N, Oka M, Sakaida I, Moribe T, Miura T, Kimura N, et al. Efficient detection of hepatocellular carcinoma by a hybrid blood test of epigenetic and classical protein markers. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2010;**412**(1-2):152-8.

Ishii 2000 *{published data only}*

Ishii M, Gama H, Chida N, Ueno Y, Shinzawa H, Takagi T, et al. Simultaneous measurements of serum alpha-fetoprotein and protein induced by vitamin K absence for detecting hepatocellular carcinoma. *American Journal of Gastroenterology* 2000;**95**(4):1036-40.

Ismail 2017a *{published data only}*

Ismail MM, Morsi HK, Abdulateef NA, Noaman MK, Abou El-Ella GA. Evaluation of prothrombin induced by vitamin K absence, macrophage migration inhibitory factor and Golgi protein-73 versus alpha fetoprotein for hepatocellular carcinoma diagnosis and surveillance. *Scandinavian Journal of Clinical and Laboratory Investigation* 2017;**77**(3):175-83.

Ismail 2017b *{published data only}*

Ismail SA, El Saadany Sh, Ziada DH, Zakaria SS, Mayah WW, Elashry H, et al. Cytokeratin-18 in diagnosis of HCC in patients with liver cirrhosis. *Asian Pacific Journal of Cancer Prevention* 2017;**18**(4):1105-11.

Iyer 2018 *{published data only}*

Iyer A, Devdas K, Sreesh S, Sathar S, Sandesh K, Singh AP. To assess the diagnostic accuracy of HCC-alpha fetoprotein-routine test (HCC-ART) score based on age, AFP, aspartate aminotransferase (AST)/alanine aminotransferase (ALT) ratio, alkaline phosphatase (ALP), and albumin (Alb) for early detection of hepatocellular carcinoma in cirrhotic patients. *Hepatology International* 2018;**12**(2):S396.

Izzo 1999 *{published data only}*

Izzo F, Cremona F, Delrio P, Leonardi E, Castello G, Pignata S, et al. Soluble interleukin-2 receptor levels in hepatocellular cancer: a more sensitive marker than alfa fetoprotein. *Annals of Surgical Oncology* 1999;**6**(2):178-85.

Jalli 2015 *{published data only}*

Jalli R, Jafari SH, Sefidbakht S, Kazemi K. Comparison of the accuracy of DWI and ultrasonography in screening hepatocellular carcinoma in patients with chronic liver disease. *Iran Journal of Radiology* 2015;**12**(1):e12708.

Jang 2016 *{published data only}*

Jang ES, Jeong SH, Kim JW, Choi YS, Leissner P, Brechot C. Diagnostic performance of alpha-fetoprotein, protein induced by vitamin K absence, osteopontin, dickkopf-1 and its combinations for hepatocellular carcinoma. *PLOS One* 2016;**17**(11):e0151069.

Jeon 2016 *{published data only}*

Jeon Y, Jang ES, Choi YS, Kim JW, Jeong SH. Glypican-3 level assessed by the enzyme-linked immunosorbent assay is inferior to alpha-fetoprotein level for hepatocellular carcinoma diagnosis. *Clinical and Molecular Hepatology* 2016;**22**(3):359-65.

Ji 2016 *{published data only}*

Ji J, Wang H, Li Y, Zheng L, Yin Y, Zou Z, et al. Diagnostic evaluation of des-gamma-carboxy prothrombin versus α-fetoprotein for hepatitis B virus-related hepatocellular carcinoma in china: a large-scale, multicentre study. *PLOS One* 2016;**11**(4):e0153227.

Jiao 2018 *{published data only}*

Jiao C, Cui L, Piao J, Qi Y, Yu Z. Clinical significance and expression of serum Golgi protein 73 in primary hepatocellular carcinoma. *Journal of Cancer Research and Therapeutics* 2018;**14**(6):1239-44.

Johnson 1978 *{published data only}*

Johnson PJ, Portmann B, Williams R. Alpha-fetoprotein concentrations measured by radioimmunoassay in diagnosing and excluding hepatocellular carcinoma. *British Medical Journal* 1978;**2**(2 (6138)):661-3.

Kanmura 2007 *{published data only}*

Kanmura S, Uto H, Kusumoto K, Ishida Y, Hasuike S, Nagata K, et al. Early diagnostic potential for hepatocellular carcinoma using the SELDI ProteinChip system. *Hepatology (Baltimore, Md.)* 2007;**45**(4):948-56.

Khairy 2015 *{published data only}*

Khairy A, Soliman D, Zayed N, Zaky M, El-Akel W, Medhat E, et al. α-fetoprotein-L3 and simplified hepatocellular carcinoma α-fetoprotein routine test for early detection of hepatocellular carcinoma in Egyptians with liver cirrhosis. *Journal of Gastroenterology and Hepatology* 2015;**30**(Suppl 4):323-407.

Kim 2001 *{published data only}*

Kim CK, Lim JH, Lee WJ. Detection of hepatocellular carcinomas and dysplastic nodules in cirrhotic liver accuracy of ultrasonography in transplant patients. *Journal of Ultrasound in Medicine* 2001;**20**:99-104.

Kim 2006a *{published data only}*

Kim J, Ki SS, Lee SD, Han CJ, Kim YC, Park SH, et al. Elevated plasma osteopontin levels in patients with hepatocellular carcinoma. *American Journal of Gastroenterology* 2006;**101**(9):2051-9.

Kim 2006b *{published data only}*

Kim KA, Lee JS, Jung ES, Kim JY, Bae WK, Kim NH, et al. Usefulness of serum alpha-fetoprotein (AFP) as a marker for hepatocellular carcinoma (HCC) in hepatitis C virus related cirrhosis: analysis of the factors influencing AFP elevation without HCC development. *Korean Journal of Gastroenterololgy* 2006;**48**(5):321-6.

Kim 2006c *{published data only}*

Kim MJ, Bae KW, Seo PJ, Jong KI, Kim JH, Lee BH, et al. Optimal cut off value of PIVKAII for the diagnosis of hepatocellular carcinoma--using ROC curve. *Korean Journal of Hepatology* 2006;**12**(3):404-11.

Kim 2012 *{published data only}*

Kim M, Gu GG, Kim SY, Min WK. AFP and PIVKA II as surveillance biomarkers for hepatocellular carcinoma. *American Journal of Clinical Pathology* 2012;**138**:A190.

Kim 2014 *{published data only}*

Kim H, Kim K, Jin J, Park J, Yu SJ, Yoon JH, et al. Measurement of glycosylated alpha-fetoprotein improves diagnostic power over the native form in hepatocellular carcinoma. *PLOS One* 2014;**9**(10):e110366.

Kim 2016 *{published data only}*

Kim Y, Park YH, Hwang S, Kim KH, Ahn CS, Moon DB, et al. Diagnostic role of blood tumor markers in predicting hepatocellular carcinoma in liver cirrhosis patients undergoing liver transplantation. *Annals of Transplantation* 2016;**21**:660-7.

Kim 2018 *{published data only}*

Kim KH, Lee SY, Hwang H, Lee JY, Ji ES, An HJ, et al. Direct monitoring of fucosylated glycopeptides of alpha-fetoprotein in human serum for early hepatocellular carcinoma by liquid chromatography–tandem mass spectrometry with immunoprecipitation. *Proteomics. Clinical Applications* 2018;**12**(6):e1800062.

Kim 2019 *{published data only}*

Kim DJ, Cho EJ, Yu KS, Jang IJ, Yoon JH, Park T, et al. Comprehensive metabolomic search for biomarkers to differentiate early stage hepatocellular carcinoma from cirrhosis. *Cancers* 2019;**11**(10):1497.

Kim 2019a *{published data only}*

Kim MN, Kim BK, Kim SU, Park JY, Ahnb SH, Hanb KH, et al. Longitudinal assessment of alpha-fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis. *Scandinavian Journal of Gastroenterology* 2019;**54**(10):1283-90.

Kim 2019b *{published data only}*

Kim MN, Kim BK, Kim SU, Park JY, Ahnb SH, Hanb KH, et al. Longitudinal assessment of alpha-fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis. *Scandinavian Journal of Gastroenterology* 2019;**54**(10):1283-90.

Krygier 2011 *{published data only}*

Krygier R, Stefaniuk P, Mikuła T, Dusza M, Jabłonowska M, Cianciara J, et al. Usefulness of alpha-fetoprotein immuno-complex (AFP-IC) in the diagnosis of hepatocellular carcinoma in cirrhotic patients. *Experimental & Clinical Hepatology* 2011;**7**(1-2):44-8.

Kudo 2019 *{published data only}*

Kudo M, Ueshima K, Osaki Y, Hirooka M, Imai Y, Aso K, et al. B-mode ultrasonography versus contrast-enhanced ultrasonography for surveillance of hepatocellular carcinoma: a prospective multicenter randomized controlled trial. *Liver Cancer* 2019;**8**(4):271-80.

Kumada 2014 *{published data only}*

Kumada T, Toyoda H, Tada T, Kiriyama S, Tanikawa M, Hisanaga Y, et al. High-sensitivity Lens culinaris agglutinin-reactive alpha-fetoprotein assay predicts early detection of hepatocellular carcinoma. *Journal of Gastroenterology* 2014;**49**(3):555-63.

Lee 2004 *{published data only}*

Lee HY, Jung JH, Kang YS, Kim YS, Moon HS, Park KO, et al. Clinical significance of transiently elevated serum AFP level in developing hepatocellular carcinoma in HBsAg positive-liver cirrhosis. *Korean Journal of Gastroenterology* 2004;**43**(4):252-9.

Lee 2014 *{published data only}*

Lee HJ, Yeon JE, Suh SJ, Lee SJ, Yoon EL, Kang K, et al. Clinical utility of plasma glypican-3 and osteopontin as biomarkers of hepatocellular carcinoma. *Gut and Liver* 2014;**8**(2):177-85.

Li 2016a *{published data only}*

Li JW, Ling WW, Lu Q, Lu CL, He D, Luo Y. Liver stiffness and serum alpha-fetoprotein in discriminating small hepatocellular carcinoma from cirrhotic nodule. *Ultrasound Quarterly* 2016;**32**(4):319-26.

Li 2016b *{published data only}*

Li L, Chen J, Chen X, Tang J, Guo H, Wang X, et al. Serum miRNAs as predictive and preventive biomarker for pre-clinical hepatocellular carcinoma. *Cancer Letters* 2016;**373**(2):234-40.

Li 2016c *{published data only}*

Li L, Chen J, Chen X, Tang J, Guo H, Wang X, et al. Serum miRNAs as predictive and preventive biomarker for pre-clinical hepatocellular carcinoma. *Cancer Letters* 2016;**373**(2):234-40.

Li 2017a *{published data only}*

Li B, Li B, Guo T, Sun Z, Li X, Li X, et al. The clinical values of serum markers in the early prediction of hepatocellular carcinoma. *BioMed Research International* 2017;**2017**:5358615.

Li 2017b *{published data only}*

Li GJ, Chen QY, Harrison TJ, Wang XY, Hu LP, Yang QL, et al. Des-γ carboxyprothrombin may not be a good biomarker for hepatocellular carcinoma in those chronically infected with hepatitis B virus with basal core promoter double mutations (T^{1762}, A^{1764}), a prospective study. *Cancer Biomarkers* 2017;**18**(3):241-8.

Li 2019a *{published data only}*

Li T, Li H, Wang A, Su X, Zhao J, Cui Y, et al. Development and validation of a simple model for detection of early hepatocellular carcinoma in a liver cirrhosis cohort. *Cancer Management and Research* 2019;**11**:9379-86.

Liao 2012 *{published data only}*

Liao J, Zhang R, Qian H, Cao L, Zhang Y, Xu W, et al. Serum profiling based on fucosylated glycoproteins for differentiating between chronic hepatitis B and hepatocellular carcinoma. *Biochemical and Biophysical Research Communications* 2012;**420**(2):308-14.

Libbrecht 2002 *{published data only}*

Libbrecht L, Bielen D, Verslype C, Vanbeckevoort D, Pirenne J, Nevens F, et al. Focal lesions in cirrhotic explant livers: pathological evaluation and accuracy of pretransplantation imaging examinations. *Liver Transplantation* 2002;**8**(9):749-61.

Lim 2015 *{published data only}*

Lim TS, Kim DY, Han KH, Kim HS, Shin SH, Jung KS, et al. Combined use of AFP, PIVKA-II, and AFP-L3 as tumor markers enhances diagnostic accuracy for hepatocellular carcinoma in cirrhotic patients. *Scandinavian Journal of Gastroenterology* 2015;**51**(3):344-53.

Lin 2000 *{published data only}*

Lin ZY, Wang LY, Yu ML, Chen SC, Chuang WL, Hsieh MY, et al. Role of serum C-reactive protein as a marker of hepatocellular carcinoma in patients with cirrhosis. *Journal of Gastroenterology and Hepatology* 2000;**15**(4):417-21.

Lin 2015 *{published data only}*

Lin XJ, Chong Y, Guo ZW, Xie C, Yang XJ, Zhang Q, et al. A serum microRNA classifier for early detection of hepatocellular carcinoma: a multicentre, retrospective, longitudinal biomarker identification study with a nested case-control study. *Lancet Oncology* 2015;**16**(7):804-15.

Lin 2016 *{published data only}*

Lin L, Lu B, Yu J, Liu W, Zhou A. Serum miR-224 as a biomarker for detection of hepatocellular carcinoma at early stage. *Clinics and Research in Hepatology and Gastroenterology* 2016;**40**(4):397-404.

Liu 2007 *{published data only}*

Liu XE, Desmyter L, Gao CF, Laroy W, Dewaele S, Vanhooren V, et al. N-glycomic changes in hepatocellular carcinoma patients with liver cirrhosis induced by hepatitis B virus. *Hepatology (Baltimore, Md.)* 2007;**46**(5):1426-35.

Liu 2010a *{published data only}*

Liu H, Li P, Zhai Y, Qu CF, Zhang LJ, Tan YF, et al. Diagnostic value of glypican-3 in serum and liver for primary hepatocellular carcinoma. *World Journal of Gastroenterology* 2010;**16**(35):4410-5.

Liu 2017 *{published data only}*

Liu XY, Fan YC, Gao S, Zhao J, Chen LY, Li F, et al. Methylation of SOX1 and VIM promoters in serum as potential biomarkers for hepatocellular carcinoma. *Neoplasma* 2017;**64**(5):745-53.

Liu 2018 *{published data only}*

Liu XF, Thin KZ, Ming XL, Shuo-Li, Ping-Luo, Man-Zhu, et al. Small nucleolar RNA host gene 18 acts as a tumor suppressor and a diagnostic indicator in hepatocellular carcinoma. *Technology in Cancer Research & Treatment* 2018;**17**:1533033818794494.

Liu 2019 *{published data only}*

Liu X, Meng J, Xu H, Niu J. Alpha-fetoprotein to transaminase ratio is related to higher diagnostic efficacy for hepatocellular carcinoma. *Medicine* 2019;**98**(17):e15414.

Liu 2020 *{published data only}*

Liu HH, Fang Y, Wang JW, Yuan XD, Fan XC, Gao S, et al. Hypomethylation of the cyclin D1 promoter in hepatitis B virus-associated hepatocellular carcinoma. *Medicine* 2020;**99**:e20326.

Loglio 2018 *{published data only}*

Loglio A, Iavarone M, Facchetti F, Di Paolo D, Perbellini R, Lunghi G, et al. The combination of PIVKA-II and AFP levels improves the diagnostic accuracy of HCC diagnosis in long-term NUC suppressed HBV Caucasian cirrhotics. *Hepatology (Baltimore, Md.)* 2018;**68**(Suppl 1):517A.

Loglio 2019 *{published data only}*

Loglio A, Iavarone M, Viganò M, Orenti A, Facchetti F, Cortinovis I, et al. Minimal increases of serum alpha-foetoprotein herald HCC detection in Caucasian HBV cirrhotic patients under long-term oral therapy. *Liver International* 2019;**39**(10):1964-74.

Lok 2010 *{published data only}*

Lok AS, Sterling RK, Everhart JE, Wright EC, Hoefs JC, Di Bisceglie AM, et al, HALT-C Trial Group. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology* 2010;**138**:493-502.

Long 2011 *{published data only}*

Long J, Wang H, Lang Z, Wang T, Long M, Wang B. Expression level of glutamine synthetase is increased in hepatocellular carcinoma and liver tissue with cirrhosis and chronic hepatitis B. *Hepatology International* 2011;**5**(2):698-706.

Luo 2018a *{published data only}*

Luo P, Yin P, Hua R, Tan Y, Li Z, Qiu G, et al. A large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2018;**67**(2):662-75.

Luo 2018b *{published data only}*

Luo P, Yin P, Hua R, Tan Y, Li Z, Qiu G, et al. A large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2018;**67**(2):662-75.

Luo 2018c *{published data only}*

Luo P, Yin P, Hua R, Tan Y, Li Z, Qiu G, et al. A large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2018;**67**(2):662-75.

Ma 2018 *{published data only}*

Ma XL, Jiang M, Zhao Y, Wang BL, Shen MN, Zhou Y, et al. Application of serum annexin a3 in diagnosis, outcome prediction and therapeutic response evaluation for patients with hepatocellular carcinoma. *Annals of Surgical Oncology* 2018;**25**(6):1686-94.

Mao 2017 *{published data only}*

Mao L, Wang Y, Wang D, Han G, Fu S, Wang J. TEMs but not DKK1 could serve as complementary biomarkers for AFP in diagnosing AFP-negative hepatocellular carcinoma. *PLOS One* 2017;**12**(9):e0183880.

Maringhini 1988 *{published data only}*

Maringhini A, Cottone M, Sciarrino E, Marcenó MP, La Seta F, Fusco G, et al. Ultrasonography and alpha-fetoprotein in

diagnosis of hepatocellular carcinoma in cirrhosis. *Digestive Diseases and Sciences* 1988;**33**(1):47-51.

Marrero 2003 *{published data only}*

Marrero JA, Su GL, Wei W, Emick D, Conjeevaram HS, Fontana RJ, et al. Des-gamma carboxyprothrombin can differentiate hepatocellular carcinoma from nonmalignant chronic liver disease in American patients. *Hepatology (Baltimore, Md.)* 2003;**37**(5):1114-21.

Marrero 2005 *{published data only}*

Marrero JA, Romano PR, Nikolaeva O, Steel L, Mehta A, Fimmel CJ, et al. GP73, a resident Golgi glycoprotein, is a novel serum marker for hepatocellular carcinoma. *Journal of Hepatology* 2005;**43**(6):1007-12.

Marrero 2009 *{published data only}*

Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, et al. Alpha-fetoprotein, des-gamma carboxyprothrombin,and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 2009;**137**:110-8.

Mashaly 2018 *{published data only}*

Mashaly AH, Anwar R, Ebrahim MA, Eissa LA, El Shishtawy MM. Diagnostic and prognostic value of talin-1 and midkine as tumor markers in hepatocellular carcinoma in Egyptian patients. *Asian Pacific Journal of Cancer Prevention* 2018;**19**(6):1503-8.

Matievskaya 2003 *{published data only}*

Matievskaya NV, Warwzynowicz-Syzweska M, Tsyrkunov VM, Boropn-Kaczmarska A. Serum alphafetoprotein (AFP) level in patients with chronic HBV and HCV infection. *Gastroenerologia Polska* 2003;**10**(1):35-40.

Matsuda 2008 *{published data only}*

Matsuda Y, Yamagiwa Y, Fukushima K, Ueno Y, Shimosegawa T. Expression of galectin-3 involved in prognosis of patients with hepatocellular carcinoma. *Hepatology Research* 2008;**38**(11):1098-111.

Mauduit Astolfi 1987 *{published data only}*

Mauduit Astolfi L, Nogueira Soriano JM, Talegon Melendez A. Liver cirrhosis and hepatocellular carcinoma. Diagnostic accuracy of ultrasound [Cirrosis hepatica y hepatocarcinoma. Valor diagnostico de la ecografia]. *Revista Espanola de Las Enfermedades del Aparato Digestivo* 1987;**72**(6):701-4.

McMahon 2000 *{published data only}*

McMahon BJ, Bulkow L, Harpster A, Snowball M, Lanier A, Sacco F, et al. Screening for hepatocellular carcinoma in Alaska natives infected with chronic hepatitis B: a 16-year population-based study. *Hepatology (Baltimore, Md.)* 2000;**32**(4 Pt 1):842-6.

Mehinovic 2018 *{published data only}*

Mehinovic L, Islamagic E, Husic-Selimovic A, Kurtovic-Kozaric A, Vukobrat-Bijedic Z, Suljevic D. Evaluation of diagnostic efficiency of alpha-fetoprotein in patients with liver cirrhosis and hepatocellular carcinoma: single-center experience. *Macedonian Journal of Medical Sciences.* 2018;**6**(9):1668-73.

Min 2014 *{published data only}*

Min B, Moon I, Park H, Jang E, Jeong SH, Kim JW. Serial measurements do not increase performance of alpha-fetoprotein for hepatocellular carcinoma surveillance in hepatitis B virus-related liver cirrhosis. *Journal of Hepatology* 2014;**60**:S215–359.

Minami 2015a *{published data only}*

Minami T, Tateishi R, Kondo M, Nakagomi R, Fujiwara N, Sato M, et al. Serum alpha-fetoprotein has high specificity for the early detection of hepatocellular carcinoma after hepatitis C virus eradication in patients. *Medicine* 2015;**94**(23):e901.

Minami 2015b *{published data only}*

Minami T, Tateishi R, Kondo M, Nakagomi R, Fujiwara N, Sato M, et al. Serum alpha-fetoprotein has high specificity for the early detection of hepatocellular carcinoma after hepatitis C virus eradication in patients. *Medicine* 2015;**94**(23):e901.

Miura 2007 *{published data only}*

Miura N, Maruyama S, Oyama K, Horie Y, Kohno M, Noma E, et al. Development of a novel assay to quantify serum human telomerase reverse transcriptase messenger RNA and its significance as a tumor marker for hepatocellular carcinoma. *Oncology* 2007;**72**(Suppl 1):45-51.

Miura 2010 *{published data only}*

Miura N, Osaki Y, Nagashima M, Kohno M, Yorozu K, Shomori K, et al. A novel biomarker TERTmRNA is applicable for early detection of hepatoma. *BMC Gastroenterology* 2010;**10**:46.

Mohamed 2020a *{published data only}*

Mohamed AA, Omar AA, El-Awady RR, Hassan SM, Eitah WM, Ahmed R, et al. MiR-155 and MiR-665 role as potential non-invasive biomarkers for hepatocellular carcinoma in Egyptian patients with chronic hepatitis C virus infection. *Journal of Translational Medicine* 2020;**8**(1):32-40.

Mohamed 2020b *{published data only}*

Mohamed AM, El-Toukhy N, Ghaith DM, Badawy I, Abdo SM, Elkadeem M. Talin-1 gene expression as a tumor marker in hepatocellular carcinoma patients: a pilot study. *Open Biomarkers Journal* 2020;**10**:15-22.

Mok 2004 *{published data only}*

Mok TS, Yu SC, Lee C, Sung J, Leung N, Lai P, et al. False-negative rate of abdominal sonography for detecting hepatocellular carcinoma in patients with hepatitis B and elevated serum alpha-fetoprotein levels. *AJR. American Journal of Roentgenology* 2004;**183**(2):453-8.

Montaser 2012 *{published data only}*

Montaser FM, Sakr M A, Khalifa MO. Alpha-L-fucosidase as a tumour marker of hepatocellular carcinoma. *Arab Journal of Gastroenterology* 2012;**13**:9-13.

Moriya 2013 *{published data only}*

Moriya S, Morimoto M, Numata K, Nozaki A, Shimoyama Y, Kondo M, et al. Fucosylated fraction of alpha-fetoprotein as a serological marker of early hepatocellular carcinoma. *Anticancer Research* 2013;**33**(3):997-1001.

Moriyama 2000 *{published data only}*

Moriyama A, Tabaru A, Unoki H, Abe S, Masumoto A, Otsuki M. Plasma nitrite/nitrate concentrations as a tumor marker for hepatocellular carcinoma. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2000;**296**(1-2):181-91.

Mukozu 2013 *{published data only}*

Mukozu T, Nagai H, Matsui D, Kanekawa T, Sumino Y. Serum VEGF as a tumor marker in patients with HCV-related liver cirrhosis and hepatocellular carcinoma. *Anticancer Research* 2013;**33**(3):1013-21.

Mustika 2019 *{published data only}*

Mustika S, Wijaya H, Pratomo B. The expressions of CD44, CD90 and alpha fetoprotein biomarkers in Indonesian patients with advanced liver disease: an observational study. *Acta Medica Indonesiana* 2019;**51**(2):137-44.

Na 2013 *{published data only}*

Na K, Jeong SK, Lee MJ, Cho SY, Kim SA, Lee MJ, et al. Human liver carboxylesterase 1 outperforms alpha-fetoprotein as biomarker to discriminate hepatocellular carcinoma from other liver diseases in Korean patients. *International Journal of Cancer. Journal International du Cancer* 2013;**133**(2):408-15.

Nabih 2014 *{published data only}*

Nabih MI, Aref WM, Fathy MM. Significance of plasma osteopontin in diagnosis of hepatitis C virus-related hepatocellular carcinoma. *Arab Journal of Gastroenterology* 2014;**15**:103-7.

Nakamura 2006 *{published data only}*

Nakamura S, Nouso K, Sakaguchi K, Ito YM, Ohashi Y, Kobayashi Y, et al. Sensitivity and specificity of des-gamma-carboxy prothrombin for diagnosis of patients with hepatocellular carcinomas varies according to tumor size. *American Journal of Gastroenterology* 2006;**101**(9):2038-43.

Nguyen 2002 *{published data only}*

Nguyen MH, Garcia RT, Simpson PW, Wright TL, Keeffe EB. Racial differences in effectiveness of α-fetoprotein for diagnosis of hepatocellular carcinoma in hepatitis C virus cirrhosis. *Hepatology (Baltimore, Md.)* 2002;**36**(2):410-7.

Nomair 2019 *{published data only}*

Nomair AM, Madkour MA, Shamseya MM, Elsheredy HG, Shokr A. Profiling of plasma metabolomics in patients with hepatitis C-related liver cirrhosis and hepatocellular carcinoma. *Clinical and Experimental Hepatology* 2019;**5**(4):317-26.

Nomura 1996 *{published data only}*

Nomura F, Ishijima M, Horikoshi A, Nakai T, Ohnishi K. Determination of serum des-gamma-carboxy prothrombin levels in patients with small-sized hepatocellular carcinoma: comparison of the conventional enzyme immunoassay and two modified methods. *American Journal of Gastroenterology* 1996;**91**(7):1380-3.

Nomura 1999 *{published data only}*

Nomura F, Ishijima M, Kuwa K, Tanaka N, Nakai T, Ohnishi K. Serum des-gamma-carboxy prothrombin levels determined by a new generation of sensitive immunoassays in patients with small-sized hepatocellular carcinoma. *American Journal of Gastroenterology* 1999;**94**(3):650-4.

Nomura 2012 *{published data only}*

Nomura F, Sogawa K, Noda K, Seimiya M, Matsushita K, Miura T, et al. Serum anti-Ku86 is a potential biomarker for early detection of hepatitis C virus-related hepatocellular carcinoma. *Biochemical and Biophysical Research Communications* 2012;**421**(4):837-43.

Oka 1994 *{published data only}*

Oka H, Tamori A, Kuroki T, Kobayashi K, Yamamoto S. Prospective study of alpha-fetoprotein in cirrhotic patients monitored for development of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 1994;**19**(1):61-6.

Oka 2001 *{published data only}*

Oka H, Saito A, Ito K, Kumada T, Satomura S, Kasugai H, et al. Multicenter prospective analysis of newly diagnosed hepatocellular carcinoma with respect to the percentage of Lens culinaris agglutinin-reactive alpha-fetoprotein. *Journal of Gastroenterology and Hepatology* 2001;**16**(12):1378-83.

Okazaki 1984 *{published data only}*

Okazaki N, Yoshida T, Yoshino M, Matue H. Screening of patients with chronic liver disease for hepatocellular carcinoma by ultrasonography. *Clinical Oncology* 1984;**10**:241-6.

Omar 2017 *{published data only}*

Omar A, Omran D, Mahmoud M, El Baz T, Medhat E, Hosny A, et al. Hepatocellular cardinoma multidisciplinary clinic – Cairo University (HMC-CU) score; a new simple score for early diagnosis of HCC. *United European Gastroenterology* 2017;**5**(5S):A184.

Omran 2016 *{published data only}*

Omran MM, Emran TM, Farid K, Eltaweel FM, Omar MA, Bazeed FB. An easy and useful noninvasive score based on α-1-acid glycoprotein and C-reactive protein for diagnosis of patients with hepatocellular carcinoma associated with hepatitis C virus infection. *Journal of Immunoassay & Immunochemistry* 2016;**37**(3):273-88.

Omran 2020 *{published data only}*

Omran MM, Farid K, Omar MA, Emran TM, El-Taweel FM, Tabll AA. A combination of α-fetoprotein, midkine, thioredoxin and a metabolite for predicting hepatocellular carcinoma. *Annals of Hepatology* 2020;**19**(2):179-85.

Ozkan 2011 *{published data only}*

Özkan H, Erdal H, Koçak E, Tutkak H, Karaeren Z, Yakut M, et al. Diagnostic and prognostic role of serum glypican 3 in patients with hepatocellular carcinoma. *Journal of Clinical Laboratory Analysis* 2011;**25**:350-3.

Park 2017a *{published data only}*

Park SJ, Jang JY, Jeong SW, Cho YK, Lee SH, Kim SG, et al. Usefulness of AFP, AFP-L3, and PIVKA-II, and their combinations in diagnosing hepatocellular carcinoma. *Medicine* 2017;**96**(11):e5811.

Park 2017b *{published data only}*

Park YH, Hwang S. Diagnostic predictability of blood tumor markers for hepatocellular carcinoma in patients with liver cirrhosis undergoing liver transplantation. *18th Congress of European Society for Organ Transplantation Book of Abstracts* 2017;**30**(Suppl 2):286.

Park 2020 *{published data only}*

Park HJ, Jang HY, Kim SY, Lee SJ, Won HJ, Byun JH, et al. Non-enhanced magnetic resonance imaging as a surveillance tool for hepatocellular carcinoma: comparison with ultrasound. *Journal of Hepatology* 2020;**72**(4):718-24.

Passos-Castilho 2015 *{published data only}*

Passos-Castilho AM, Carvalho VM, Cardozo KH, Kikuchi L, Chagas AL, Gomes-Gouvêa MS, et al. Serum lipidomic profiling as a useful tool for screening potential biomarkers of hepatitis B-related hepatocellular carcinoma by ultraperformance liquid chromatography-mass spectrometry. *BMC Cancer* 2015;**15**:985.

Pateron 1994 *{published data only}*

Pateron D, Ganne N, Trinchet JC, Aurousseau MH, Mal F, Meicler C, et al. Prospective study of screening for hepatocellular carcinoma in Caucasian patients with cirrhosis. *Journal of Hepatology* 1994;**20**(1):65-71.

Paul 2007 *{published data only}*

Paul SB, Gulati MS, Sreenivas V, Madan K, Gupta AK, Mukhopadhyay S, et al. Evaluating patients with cirrhosis for hepatocellular carcinoma: value of clinical symptomatology, imaging and alpha-fetoprotein. *Oncology* 2007;**72**(Suppl 1):117-23.

Piciocchi 2013 *{published data only}*

Piciocchi M, Cardin R, Vitale A, Vanin V, Giacomin A, Pozzan C, et al. Circulating free DNA in the progression of liver damage to hepatocellular carcinoma. *Hepatology International* 2013;**7**(4):1050-7.

Pinero 2015 *{published data only}*

Pinero F, Marciano S, Anders M, Orozco F, Zerega A, Cabrera CR, et al. Screening for liver cancer during transplant waiting list: a multicenter study from South America. *European Journal of Gastroenterology & Hepatology* 2015;**27**(3):355-60.

Pompili 2003 *{published data only}*

Pompili M, Addolorato G, Pignataro G, Rossi C, Zuppi C, Covino M, et al. Evaluation of the albumin-gamma-glutamyltransferase isoenzyme as a diagnostic marker of hepatocellular carcinoma-complicating liver cirrhosis. *Journal of Gastroenterology and Hepatology* 2003;**18**(3):288-95.

Poon 2001 *{published data only}*

Poon TC, Chan AT, Zee B, Ho SK, Mok TS, Leung TW, et al. Application of classification tree and neural network algorithms to the identification of serological liver marker profiles for the diagnosis of hepatocellular carcinoma. *Oncology* 2001;**61**(4):275-83.

Porta 2008 *{published data only}*

Porta C, De Amici M, Quaglini S, Paglino C, Tagliani F, Boncimino A, et al. Circulating interleukin-6 as a tumor marker for hepatocellular carcinoma. *Annals of Oncology* 2008;**19**(2):353-8.

Pote 2015 *{published data only}*

Pote N, Cauchy F, Albuquerque M, Voitot H, Belghiti J, Castera L, et al. Performance of PIVKA-II for early hepatocellular carcinoma diagnosis and prediction of microvascular invasion. *Journal of Hepatology* 2015;**62**(4):848-54.

Powell-Jackson 1987 *{published data only}*

Powell-Jackson P, Forbes A, Michell M, Williams R. Effect of cirrhosis on accuracy of 99Tcm-tin colloid scintigraphy and ultrasound scanning in diagnosis of hepatocellular carcinoma. *British Journal of Radiology* 1987;**60**(720):1221-2.

Qi 2020 *{published data only}*

Qi F, Zhou A, Yan L, Yuan X, Wang D, Chang R, et al. The diagnostic value of PIVKA-II, AFP, AFP-L3, CEA, and their combinations in primary and metastatic hepatocellular carcinoma. *Journal of Clinical Laboratory Analysis* 2020;**34**(5):e23158.

Raedle 1995 *{published data only}*

Raedle J, Roth WK, Oremek G, Caspary WF, Zeuzem S. Alpha-fetoprotein and p53 autoantibodies in patients with chronic hepatitis C. *Digestive Diseases and Sciences* 1995;**40**(12):2587-94.

Raedle 1998 *{published data only}*

Raedle J, Oremek G, Truschnowitsch M, Lorenz M, Roth WK, Caspary WF, et al. Clinical evaluation of autoantibodies to p53 protein in patients with chronic liver disease and hepatocellular carcinoma. *European Journal of Cancer* 1998;**34**(8):1198-203.

Raff 2014 *{published data only}*

Raff E, Kakati DD, Shoreibah MG, Bloomer JR, Singal AK. CT and MRI scans are overused for hepatocellular carcinoma screening in patients with cirrhosis. *Gastroenterology* 2014;**146**(5 Suppl 1):S-606.

Reichl 2015 *{published data only}*

Reichl P, Fang M, Starlinger P, Staufer K, Nenutil R, Muller P, et al. Multicenter analysis of soluble Axl reveals diagnostic value for very early stage hepatocellular carcinoma. *International Journal of Cancer. Journal International du Cancer* 2015;**137**(2):385-94.

Ricco 2018 *{published data only}*

Ricco G, Cavallone D, Cosma C, Caviglia GP, Oliveri F, Biasiolo A, et al. Impact of etiology of chronic liver disease on hepatocellular carcinoma biomarkers. *Cancer Biomarkers* 2018;**21**(3):603-12.

Saada 1997 *{published data only}*

Saada J, Bhattacharya S, Dhillon AP, Dick R, Burroughs AK, Rolles K, et al. Detection of small hepatocellular carcinomas in cirrhotic livers using iodised oil computed tomography. *Gut* 1997;**41**:404-7.

Sadeghi 2015 *{published data only}*

Sadeghi M, Lahdou I, Oweira H, Daniel V, Terness P, Schmidt J, et al. Serum levels of chemokines CCL4 and CCL5 in cirrhotic patients indicate the presence of hepatocellular carcinoma. *British Journal of Cancer* 2015;**113**(5):756-62.

Sadik 2019 *{published data only}*

Sadik NA, Ahmed NR, Mohamed MF, Ashoush OA. Serum vascular endothelial growth factor in patients with hepatocellular carcinoma and its validity as a tumor biomarker. *Open Biomarkers Journal* 2019;**9**:84-94.

Saitta 2017 *{published data only}*

Saitta C, Raffa G, Alibrandi A, Brancatelli S, Lombardo D, Tripodi G, et al. PIVKA-II is a useful tool for diagnostic characterization of ultrasound-detected liver nodules in cirrhotic patients. *Medicine (Baltimore)* 2017;**96**(26):e7266.

Salem 2013 *{published data only}*

Salem M, Atti SA, Raziky ME, Darweesh SK, Sharkawy ME. Clinical significance of plasma osteopontin level asa biomarker of hepatocellular carcinoma. *Gastroenterology Research* 2013;**6**(5):191-9.

Sanai 2010 *{published data only}*

Sanai FM, Sobki S, Bzeizi KI, Shaikh SA, Alswat K, Al-Hamoudi W, et al. Assessment of alpha-fetoprotein in the diagnosis of hepatocellular carcinoma in Middle Eastern patients. *Digestive Diseases and Sciences* 2010;**55**(12):3568-75.

Sarwar 2014 *{published data only}*

Sarwar S, Khan AA, Tarique S. Validity of alpha fetoprotein for diagnosis of hepatocellular carcinoma in cirrhosis. *Journal of College of Physicians and Surgeons Pakistan* 2014;**24**(1):18-22.

Sassa 1999 *{published data only}*

Sassa T, Kumada T, Nakano S, Uematsu T. Clinical utility of simultaneous measurement of serum high sensitivity des-gamma-carboxy-prothrombin and Lens culinaris agglutinin A-reactive alpha-fetoptrotein in patient with small hepatocellular carcinoma. *European Journal of Gastroentrology & Hepatology* 1999;**11**:1387-92.

Sato 1993 *{published data only}*

Sato Y, Nakata K, Kato Y, Shima M, Ishii N, Koji T, et al. Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. *New England Journal of Medicine* 1993;**328**(25):1802-6.

Seo 2015 *{published data only}*

Seo SI, Kim HS, Kim WJ, Shin WG, Kim DJ, Kim KH, et al. Diagnostic value of PIVKA-II and alpha-fetoprotein in hepatitis B virus-associated hepatocellular carcinoma. *World Journal of Gastroenterology* 2015;**21**(13):3928-35.

Shaheen 2015 *{published data only}*

Shaheen KY, Abdel-Mageed AI, Safwat E, AlBreedy AM. The value of serum midkine level in diagnosis of hepatocellular carcinoma. *International Journal of Hepatology* 2015;**2015**:146389.

Shaheen 2018 *{published data only}*

Shaheen NM, Zayed N, Riad NM, Tamim HH, Shahin RM, Labib DA, et al. Role of circulating miR-182 and miR-150 as biomarkers for cirrhosis and hepatocellular carcinoma post HCV infection in Egyptian patients. *Virus Research* 2018;**255**:77-84.

Shang 2012a *{published data only}*

Shang S, Plymoth A, Ge S, Feng Z, Rosen HR, Sangrajrang S, et al. Identification of osteopontin as a novel marker for early hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2012;**55**(2):483-90.

Shang 2012b *{published data only}*

Shang S, Plymoth A, Ge S, Feng Z, Rosen HR, Sangrajrang S, et al. Identification of osteopontin as a novel marker for early hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2012;**55**(2):483-90.

Shariff 2010 *{published data only}*

Shariff MI, Ladep NG, Cox IJ, Williams HR, Okeke E, Malu A, et al. Characterization of urinary biomarkers of hepatocellular carcinoma using magnetic resonance spectroscopy in a Nigerian population. *Journal of Proteome Research* 2010;**9**(2):1096-103.

Shariff 2016 *{published data only}*

Shariff MI, Kim JU, Ladep NG, Crossey MM, Koomson LK, Zabron A, et al. Urinary metabotyping of hepatocellular carcinoma in a UK cohort using proton nuclear magnetic resonance spectroscopy. *Journal of Clinical and Experimental Hepatology* 2016;**6**(3):186-94.

Sharma 2010 *{published data only}*

Sharma B, Srinivasan R, Chawla YK, Kapil S, Saini N, Singla B, et al. Clinical utility of prothrombin induced by vitamin K absence in the detection of hepatocellular carcinoma in Indian population. *Hepatology International* 2010;**4**(3):569-76.

Shen 2012a *{published data only}*

Shen Q, Fan J, Yang XR, Tan Y, Zhao W, Xu Y, et al. Serum DKK1 as a protein biomarker for the diagnosis of hepatocellular carcinoma: a large-scale, multicentre study. *Lancet Oncology* 2012;**13**(8):817-26.

Shen 2012b *{published data only}*

Shen Q, Fan J, Yang XR, Tan Y, Zhao W, Xu Y, et al. Serum DKK1 as a protein biomarker for the diagnosis of hepatocellular carcinoma: a large-scale, multicentre study. *Lancet Oncology* 2012;**13**(8):817-26.

Sherman 1995 *{published data only}*

Sherman M, Peltekian KM, Lee C. Screening for hepatocellular carcinoma in chronic carriers of hepatitis B virus: incidence and prevalence of hepatocellular carcinoma in a North American urban population. *Hepatology (Baltimore, Md.)* 1995;**22**(2):432-8.

Shimizu 2002 *{published data only}*

Shimizu A, Shiraki K, Ito T, Sugimoto K, Sakai T, Ohmori S, et al. Sequential fluctuation pattern of serum des-gamma-carboxy prothrombin levels detected by high-sensitive electro

chemiluminescence system as an early predictive marker for hepatocellular carcinoma in patients with cirrhosis. *International Journal of Molecular Medicine* 2002;**9**(3):245-50.

Shu 2010 *{published data only}*

Shu H, Kang X, Guo K, Li S, Li M, Sun L, et al. Diagnostic value of serum haptoglobin protein as hepatocellular carcinoma candidate marker complementary to α fetoprotein. *Oncology Reports* 2010;**24**(5):1271-6.

Simão 2015 *{published data only}*

Simão A, Madaleno J, Silva N, Rodrigues F, Caseiro P, Costa JN, et al. Plasma osteopontin is a biomarker for the severity of alcoholic liver cirrhosis, not for hepatocellular carcinoma screening. *BMC Gastroenterology* 2015;**15**:73.

Singal 2012 *{published data only}*

Singal AG, Conjeevaram HS, Volk ML, Fu S, Fontana RJ, Askari F, et al. Effectiveness of hepatocellular carcinoma surveillance in patients with cirrhosis. *Cancer Epidemiology, Biomarkers & Prevention* 2012;**21**(5):793-9.

Snowberger 2007 *{published data only}*

Snowberger N, Chinnakotla S, Lepe RM, Peattie J, Goldstein R, Klintmalm GB, et al. Alpha fetoprotein, ultrasound, computerized tomography and magnetic resonance imaging for detection of hepatocellular carcinoma in patients with advanced cirrhosis. *Alimentary Pharmacology & Therapeutics* 2007;**26**(9):1187-94.

Son 2019 *{published data only}*

Son JH, Choi SH, Kim SY, Jang HY, Byun JH, Won HJ, et al. Validation of US liver imaging reporting and data system version 2017 in patients at high risk for hepatocellular carcinoma. *Radiology* 2019;**292**:390-7.

Song 2002 *{published data only}*

Song BC, Chung YH, Kim JA, Choi WB, Suh DD, Pyo SI, et al. Transforming growth factor-beta1 as a useful serologic marker of small hepatocellular carcinoma. *Cancer* 2002;**94**(1):175-80.

Song 2011 *{published data only}*

Song J, Kim J. Significance of alpha fetoprotein in the surveillance of HBV-associated hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2011;**54**(4):1419A.

Song 2014 *{published data only}*

Song P, Feng X, Inagaki Y, Song T, Zhang K, Wang Z, et al. Clinical utility of simultaneous measurement of alpha-fetoprotein and des-γ-carboxy prothrombin for diagnosis of patients with hepatocellular carcinoma in China: a multi-center case-controlled study of 1,153 subjects. *Bioscience Trends* 2014;**8**(5):266-73.

Song 2020a *{published data only}*

Song T, Wang L, Su B, Zeng W, Jiang T, Zhang T, at al. Diagnostic value of alpha-fetoprotein, Lensculinaris agglutinin-reactive alpha-fetoprotein, and des gamma-carboxy prothrombin in hepatitis B virus-related hepatocellular carcinoma. *Journal of International Medical Research* 2020;**48**(3):300060519889270.

Song 2020b *{published data only}*

Song X, Wu A, Ding Z, Liang S, Zhang C. Soluble AXL is a novel diagnostic biomarker of hepatocellular carcinoma in Chinese patients with chronic hepatitis B virus infection. *Cancer Research and Treatment* 2020;**52**(3):789-97.

Soroida 2012 *{published data only}*

Soroida Y, Ohkawa R, Nakagawa H, Satoh Y, Yoshida H, Kinoshita H, et al. Increased activity of serum mitochondrial isoenzyme of creatine kinase in hepatocellular carcinoma patients predominantly with recurrence. *Journal of Hepatology* 2012;**57**(2):330-6.

Sterling 2009 *{published data only}*

Sterling RK, Jeffers L, Gordon F, Venook AP, Reddy KR, Satomura S, et al. Utility of Lens culinaris agglutinin-reactive fraction of alpha-fetoprotein and des-gamma-carboxy prothrombin, alone or in combination, as biomarkers for hepatocellular carcinoma. *Clinical Gastroenterology and Hepatology* 2009;**7**(1):104-13.

Sterling 2012 *{published data only}*

Sterling RK, Wright EC, Morgan TR, Seeff LB, Hoefs JC, Di Bisceglie AM, et al. Frequency of elevated hepatocellular carcinoma (HCC) biomarkers in patients with advanced hepatitis C. *American Journal of Gastroenterology* 2012;**107**(1):64-74.

Sultanik 2017 *{published data only}*

Sultanik P, Ginguay A, Vandame J, Popovici T, Meritet JF, Cynober L, et al. Diagnostic accuracy of des-gamma-carboxy prothrombin for hepatocellular carcinoma in a French cohort using the Lumipulse® G600 analyzer. *Journal of Viral Hepatitis* 2017;**24**(1):80-5.

Sun 2010 *{published data only}*

Sun S, Poon RT, Lee NP, Yeung C, Chan KL, Ng IO, et al. Proteomics of hepatocellular carcinoma: serum vimentin as a surrogate marker for small tumors (<or=2 cm). *Journal of Proteome Research* 2010;**9**(4):1923-30.

Sun 2020 *{published data only}*

Sun Q, Li J, Jin B, Wang T, Gu J. Evaluation of miR-331-3p and miR-23b-3p as serum biomarkers for hepatitis C virus-related hepatocellular carcinoma at early stage. *Clinics and Research in Hepatology and Gastroenterology* 2020;**44**(1):21-8.

Sutherland 2017 *{published data only}*

Sutherland T, Watts J, Ryan M, Galvin A, Temple F, Vuong J, et al. Diffusion-weighted MRI for hepatocellular carcinoma screening in chronic liver disease: direct comparison with ultrasound screening. *Journal of Medical Imaging and Radiation Oncology* 2017;**61**(1):34-9.

Tahon 2019 *{published data only}*

Tahon AM, El-Ghanam MZ, Zaky S, Emran TM, Bersy, AM El-Raey F, et al. Significance of glypican-3 in early detection of hepatocellular carcinoma in cirrhotic patients. *Journal of Gastrointestinal Cancer* 2019;**50**(3):434-41.

Takaya 2019 *{published data only}*

Takaya H, Namisaki T, Kitade M, Kaji K, Nakanishi K, Tsuji Y, et al. VWF/ADAMTS13 ratio as a potential biomarker for early detection of hepatocellular carcinoma. *BMC Gastroenterology* 2019;**19**(1):167.

Takikawa 1992 *{published data only}*

Takikawa Y, Suzuki K, Yamazaki K, Goto T, Madarame T, Miura Y, et al. Plasma abnormal prothrombin (PIVKA-II): a new and reliable marker for the detection of hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology* 1992;**7**(1):1-6.

Talkahn 2018 *{published data only}*

Talkahn M, El Sayed EY, Fouad MH, Ghait RSA. Evaluation of serum endoglin level as a new diagnostic biomarker for hepatocellular carcinoma. *United European Gastroenterology Journal* 2018;**6**(8S):A157.

Tan 2012 *{published data only}*

Tan Y, Yin P, Tang L, Xing W, Huang Q, Cao D, et al. Metabolomics study of stepwise hepatocarcinogenesis from the model rats to patients: potential biomarkers effective for small hepatocellular carcinoma diagnosis. *Molecular & Cellular Proteomics* 2012;**11**(2):M111.010694.

Tan 2014 *{published data only}*

Tan Y, Ge G, Pan T, Wen D, Chen L, Yu X, et al. A serum microRNA panel as potential biomarkers for hepatocellular carcinoma related with hepatitis B virus. *PLOS One* 2014;**9**(9):e107986.

Tanaka 1986 *{published data only}*

Tanaka S, Kitamura T, Ohshima A, Umeda K, Okuda S, Ohtani T, et al. Diagnostic accuracy of ultrasonography for hepatocellular carcinoma. *Cancer* 1986;**588**:344-7.

Tang 2017a *{published data only}*

Tang XQ, Li H, Yan LB, Zhou LY, Chen EQ, Liu M, et al. Diagnostic value of PIVKA-II in detecting hepatocellular carcinoma in Chinese patients with hepatitis B virus infection. *Future Virology* 2017;**12**(5):259–67.

Tanglijvanich 2010 *{published data only}*

Tangkijvanich P, Chanmee T, Komtong S, Mahachai V, Wisedopas N, Pothacharoen P, et al. Diagnostic role of serum glypican-3 in differentiating hepatocellular carcinoma from non-malignant chronic liver disease and other liver cancers. *Journal of Gastroenterology and Hepatology* 2010;**25**(1):129-37.

Tayob 2016a *{published data only}*

Tayob N, Lok AS, Do KA, Feng Z. Improved detection of hepatocellular carcinoma by using a longitudinal alpha-fetoprotein screening algorithm. *Clinical Gastroenterology and Hepatology* 2016;**14**(3):469-75.

Tayob 2016b *{published data only}*

Tayob N, Lok AS, Do KA, Feng Z. Improved detection of hepatocellular carcinoma by using a longitudinal alpha-fetoprotein screening algorithm. *Clinical Gastroenterology and Hepatology* 2016;**14**(3):469-75.

Tayob 2019 *{published data only}*

Tayob N, Christie I, Richardson P, Feng Z, White DL, Davila J, et al. Validation of the hepatocellular carcinoma early detection screening (HES) algorithm in a cohort of veterans with cirrhosis. *Clinical Gastroenterology and Hepatology* 2019;**17**(9):1886-93.

Teefey 2003 *{published data only}*

Teefey SA, Hildeboldt CC, Dehdashti F, Siegel BA, Peters MG, Heiken JP, et al. Detection of primary hepatic malignancy in liver transplant candidates: prospective comparison of CT, MR imaging, US, and PET. *Radiology* 2003;**226**(2):533-42.

Teng 2016 *{published data only}*

Teng Y, Fan YC, Mu NN, Zhao J, Sun FK, Wang K. Serum SOX11 promoter methylation is a novel biomarker for the diagnosis of hepatitis B virus-related hepatocellular carcinoma. *Neoplasma* 2016;**63**(3):419-26.

Tian 2017 *{published data only}*

Tian MM, Fan YC, Zhao J, Gao S, Zhao ZH, Chen LY, et al. Hepatocellular carcinoma suppressor 1 promoter hypermethylation in serum. A diagnostic and prognostic study in hepatitis B. *Clinics and Research in Hepatology and Gastroenterology* 2017;**41**(2):171-80.

Tong 2001 *{published data only}*

Tong MJ, Blatt LM, Kao VWC. Surveillance for hepatocellular carcinoma in patients with chronic viral hepatitis in the United States of America. *Journal of Gastroenterology and Hepatology* 2001;**16**:553-9.

Toraih 2018 *{published data only}*

Toraih EA, Ellawindy A, Fala SY, Al Ageeli E, Gouda NS, Fawzy MS, et al. Oncogenic long noncoding RNA MALAT1 and HCV-related hepatocellular carcinoma. *Biomedecine & Pharmacotherapie [Biomedicine & Pharmacotherapy]* 2018;**102**:653-69.

Tremolada 1989 *{published data only}*

Tremolada F, Benvegnù L, Drago C, Casarin C, Cecchetto A, et al. Early detection of hepatocellular carcinoma in patients with cirrhosis by alpha-fetoprotein, ultrasound and fine-needle biopsy. *Hepato-gastroenterology* 1989;**36**(6):519-21.

Trevisani 2001 *{published data only}*

Trevisani F, D'Intino PE, Morselli-Labate AM, Mazzella G, Accogli E, Caraceni P, et al. Serum alpha-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: influence of HBsAg and anti-HCV status. *Journal of Hepatology* 2001;**34**(4):570-5.

Tsai 1995 *{published data only}*

Tsai JF, Jeng JE, Ho MS, Chang WY, Lin ZY, Tsai JH. Clinical evaluation of serum alpha-fetoprotein and circulating immune complexes as tumour markers of hepatocellular carcinoma. *British Journal of Cancer* 1995;**72**(2):442-6.

Tsai 1997 *{published data only}*

Tsai JF, Jeng JE, Chuang LY, Yang ML, Ho MS, Chang WY, et al. Clinical evaluation of urinary transforming growth factor-beta1 and serum alpha-fetoprotein as tumour markers

of hepatocellular carcinoma. *British Journal of Cancer* 1997;**75**(10):1460-6.

Tsai 2017 *{published data only}*

Tsai JF, Chen SC, Lin ZY, Dai CY, Yi ML, Chuang WL, et al. The optimal cut-off value of serum alpha-fetoprotein in the diagnosis of hepatocellular carcinoma. *Journal of Hepatology* 2017;**66**:S454.

Tsuda 2004 *{published data only}*

Tsuda Y, Fukuda A, Kobayashi H, Itou D, Yoshimoto S, Iwata K, et al. Serum neopterin as a marker for screening of hepatocellular carcinoma. *Pteridines* 2004;**15**:161-9.

Ungtrakul 2016 *{published data only}*

Ungtrakul T, Mahidol C, Chun-On P, Laohapand C, Siripongsakun S, Worakitsitisatorn A, et al. Hepatocellular carcinoma screening and surveillance in 2293 chronic hepatitis B patients in an endemic area. *World Journal of Gastroenterology* 2016;**22**(34):7806-12.

Unic 2013 *{published data only}*

Unic A, Grgurevic I, Derek L, Marijancevic D, Serdar T, Romic Z. AFP-L3 – screening marker for hepatocellular carcinoma in patients with alcoholic cirrhosis. *Biochimica Clinica* 2013;**37**:S679.

Van Thiel 2004 *{published data only}*

Van Thiel DH, Yong S, Li SD, Kennedy M, Brems J. The development of de novo hepatocellular carcinoma in patients on a liver transplant list: frequency, size, and assessment of current screening methods. *Liver Transplantation* 2004;**10**(5):631-7.

Villacastin Ruiz 2016 *{published data only}*

Villacastín Ruiz E, Caro-Patón Gómez A, Calero Aguilar H, Pérez Saborido B, García Pajares F, Sánchez Antolín G, et al. Review of imaging techniques in the diagnosis of hepatocellular carcinoma in patients who require a liver transplant. *European Journal of Gastroenterology & Hepatology* 2016;**28**:412-20.

Volk 2007 *{published data only}*

Volk ML, Hernandez JC, Su GL, Lok AS, Marrero JA. Risk factors for hepatocellular carcinoma may impair the performance of biomarkers: a comparison of AFP, DCP, and AFP-L3. *Cancer Biomarkers* 2007;**3**(2):79-87.

Vongsuvanh 2016 *{published data only}*

Vongsuvanh R, van der Poorten D, Iseli T, Strasser SI, McCaughan GW, George J. Midkine increases diagnostic yield in AFP negative and NASH-related hepatocellular carcinoma. *PLOS One* 2016;**11**(5):e0155800.

Wang 2005 *{published data only}*

Wang CS, Lin CL, Lee HC, Chen KY, Chiang MF, Chen HS, et al. Usefulness of serum des-gamma-carboxy prothrombin in detection of hepatocellular carcinoma. *World Journal of Gastroenterology* 2005;**11**(39):6115-9.

Wang 2009 *{published data only}*

Wang M, Long RE, Comunale MA, Junaidi O, Marrero J, Di Bisceglie AM, et al. Novel fucosylated biomarkers for the early detection of hepatocellular carcinoma. *Cancer Epidemiology, Biomarkers & Prevention* 2009;**18**(6):1914-21.

Wang 2013a *{published data only}*

Wang W, Li H, Zhou Y, Jie S. Peripheral blood microvesicles are potential biomarkers for hepatocellular carcinoma. *Cancer Biomarkers* 2013;**13**(5):351-7.

Wang 2013b *{published data only}*

Wang WW, Ang SF, Kumar R, Heah C, Utama A, Tania NP, et al. Identification of serum monocyte chemoattractant protein-1 and prolactin as potential tumor markers in hepatocellular carcinoma. *PLOS One* 2013;**8**(7):e68904.

Wang 2014a *{published data only}*

Wang N, Cao Y, Song W, He K, Li T, Wang J, et al. Serum peptide pattern that differentially diagnoses hepatitis B virus-related hepatocellular carcinoma from liver cirrhosis. *Journal of Gastroenterology and Hepatology* 2014;**29**(7):1544-50.

Wang 2014b *{published data only}*

Wang Y, Yang H, Xu H, Lu X, Sang X, Zhong S, et al. Golgi protein 73, not Glypican-3, may be a tumor marker complementary to α-Fetoprotein for hepatocellular carcinoma diagnosis. *Journal of Gastroenterology and Hepatology* 2014;**29**(3):597-602.

Wang 2016a *{published data only}*

Wang M, Devarajan K, Singal AG, Marrero JA, Dai J, Feng Z, et al. The Doylestown algorithm: a test to improve the performance of AFP in the detection of hepatocellular carcinoma. *Cancer Prevention Research (Philadelphia, Pa.)* 2016;**9**(2):172-9.

Wang 2016b *{published data only}*

Wang M, Devarajan K, Singal AG, Marrero JA, Dai J, Feng Z, et al. The Doylestown algorithm: a test to improve the performance of AFP in the detection of hepatocellular carcinoma. *Cancer Prevention Research (Philadelphia, Pa.)* 2016;**9**(2):172-9.

Wang 2016c *{published data only}*

Wang M, Devarajan K, Singal AG, Marrero JA, Dai J, Feng Z, et al. The Doylestown algorithm: a test to improve the performance of AFP in the detection of hepatocellular carcinoma. *Cancer Prevention Research (Philadelphia, Pa.)* 2016;**9**(2):172-9.

Wang 2016d *{published data only}*

Wang M, Devarajan K, Singal AG, Marrero JA, Dai J, Feng Z, et al. The Doylestown algorithm: a test to improve the performance of AFP in the detection of hepatocellular carcinoma. *Cancer Prevention Research (Philadelphia, Pa.)* 2016;**9**(2):172-9.

Wang 2016e *{published data only}*

Wang M, Devarajan K, Singal AG, Marrero JA, Dai J, Feng Z, et al. The Doylestown algorithm: a test to improve the performance of AFP in the detection of hepatocellular carcinoma. *Cancer Prevention Research (Philadelphia, Pa.)* 2016;**9**(2):172-9.

Wang 2017 *{published data only}*

Wang X, Zhang W, Liu Y, Gong W, Sun P, Kong X, et al. Diagnostic value of prothrombin induced by the absence of vitamin K or antagonist-II (PIVKA-II) for early stage HBV related hepatocellular carcinoma. *Infectious Agents and Cancer* 2017;**12**:47-55.

Wang 2019a *{published data only}*

Wang Q, Chen Q, Zhang X, Lu XL, Du Q, Zhu T, et al. Diagnostic value of gamma-glutamyltransferase/aspartate aminotransferase ratio, protein induced by vitamin K absence or antagonist II, and alpha-fetoprotein in hepatitis B virus-related hepatocellular carcinoma. *World Journal of Gastroenterology* 2019;**25**(36):5515-29.

Wang 2019b *{published data only}*

Wang X, Li MM, Niu Y, Zhang X, Yin J, Zhao C, et al. Serum zonulin in HBV-associated chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma. *Disease Markers* 2019;**2019**:5945721.

Weiss 2019 *{published data only}*

Weis A, Marquart L, Calvopina DA, Genz B, Ramm GA, Skoien R. Serum microRNAs as biomarkers in hepatitis C: preliminary evidence of a microRNApanel for the diagnosis of hepatocellular carcinoma. *International Journal of Molecular Sciences* 2019;**20**(4):864.

Wong 2008 *{published data only}*

Wong GL, Wong VW, Tan GM, Ip KI, Lai WK, Li YW, et al. Surveillance programme for hepatocellular carcinoma improves the survival of patients with chronic viral hepatitis. *Liver International* 2008;**28**(1):79-87.

Wong 2009 *{published data only}*

Wong VW, Yu J, Cheng AS, Wong GL, Chan HY, Chu ES, et al. High serum interleukin-6 level predicts future hepatocellular carcinoma development in patients with chronic hepatitis B. *International Journal of Cancer. Journal International du Cancer* 2009;**124**(12):2766-70.

Wong 2014a *{published data only}*

Wong GL, Chan HL, Tse YK, Chan HY, Tse CH, Lo AO, et al. On-treatment alpha-fetoprotein is a specific tumor marker for hepatocellular carcinoma in patients with chronic hepatitis B receiving entecavir. *Hepatology (Baltimore, Md.)* 2014;**59**(3):986-95.

Wong 2014b *{published data only}*

Wong GL, Chan HL, Tse YK, Chan HY, Tse CH, Lo AO, et al. On-treatment alpha-fetoprotein is a specific tumor marker for hepatocellular carcinoma in patients with chronic hepatitis B receiving entecavir. *Hepatology (Baltimore, Md.)* 2014;**59**(3):986-95.

Wu 2009 *{published data only}*

Wu C, Wang Z, Liu L, Zhao P, Wang W, Yao D, et al. Surface enhanced laser desorption/ionization profiling: new diagnostic method of HBV-related hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology* 2009;**24**(1):55-62.

Wu 2017 *{published data only}*

Wu XM, Xi ZF, Liao P, Huang HD, Huang XY, Wang C, et al. Diagnostic and prognostic potential of serum microRNA-4651 for patients with hepatocellular carcinoma related to aflatoxin B1. *Oncotarget* 2017;**8**(46):81235-49.

Wu 2018 *{published data only}*

Wu J, Xiang Z, Bai L, He L, Tan L, Hu M, et al. Diagnostic value of serum PIVKA-II levels for BCLC early hepatocellular carcinoma and correlation with HBV DNA. *Cancer Biomarkers* 2018;**23**(2):235-42.

Wu 2020 *{published data only}*

Wu M, Liu Z, Li X, Zhang A, Li N. Dynamic changes in serum markers and their utility in the early diagnosis of all stages of hepatitis B-associated hepatocellular carcinoma. *OncoTargets and Therapy* 2020;**13**:827-40.

Xing 2019 *{published data only}*

Xing H, Qiu H, Ding X, Han, Li Z, Wu H, et al. Clinical performance of α-L-fucosidase for early detection of hepatocellular carcinoma. *Biomarkers in Medicine* 2019;**13**(7):545-55.

Xu 2018 *{published data only}*

Xu H, Dong X, Chen Y, Wang X. Serum exosomal hnRNPH1 mRNA as a novel marker for hepatocellular carcinoma. *Clinical Chemistry and Laboratory Medicine* 2018;**56**(3):479-84.

Yan 2018 *{published data only}*

Yan L, Chen Y, Zhou J, Zhao H, Zhang H, Wang G. Diagnostic value of circulating cell-free DNA levels for hepatocellular carcinoma. *International Journal of Infectious Diseases* 2018;**67**:92-7.

Yang 2013a *{published data only}*

Yang L, Rong W, Xiao T, Zhang Y, Xu B, Liu Y, et al. Secretory/releasing proteome-based identification of plasma biomarkers in HBV-associated hepatocellular carcinoma. *Science China. Life Sciences* 2013;**56**:638-46.

Yang 2014 *{published data only}*

Yang Y, Fan YC, Gao S, Dou CY, Zhang JJ, Sun FK, et al. Methylated cysteine dioxygenase-1 gene promoter in the serum is a potential biomarker for hepatitis B virus-related hepatocellular carcinoma. *Tohoku Journal of Experimental Medicine* 2014;**232**(3):187-94.

Yang 2017 *{published data only}*

Yang JD, Dai J, Singal AG, Gopal P, Addissie BD, Nguyen MH, et al. Improved performance of serum alpha-fetoprotein for hepatocellular carcinoma diagnosis in HCV cirrhosis with normal alanine transaminase. *Cancer Epidemiology, Biomarkers & Prevention* 2017;**26**(7):1085-92.

Yang 2019 *{published data only}*

Yang JD, Addissie BD, Mara KC, Harmsen WS, Dai J, Zhang N, et al. GALAD score for hepatocellular carcinoma detection in comparison with liver ultrasound and proposal of GALADUS score. *Cancer Epidemiology, Biomarkers & Prevention* 2019;**28**(3):531-8.

Yao 2016 *{published data only}*

Yao M, Zhao J, Lu F. Alpha-fetoprotein still is a valuable diagnostic and prognosis predicting biomarker in hepatitis B virus infection-related hepatocellular carcinoma. *Oncotarget* 2016;**7**(4):3702-8.

Ye 2019a *{published data only}*

Ye X, Li C, Zu X, Lin M, Liu Q, Liu J, et al. A large-scale multicenter study validates aldo-keto reductase family 1 member B10 as a prevalent serum marker for detection of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2019;**69**(6):2489-501.

Ye 2019b *{published data only}*

Ye X, Li C, Zu X, Lin M, Liu Q, Liu J, et al. A large-scale multicenter study validates aldo-keto reductase family 1 member B10 as a prevalent serum marker for detection of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2019;**69**(6):2489-501.

Yoon 2009 *{published data only}*

Yoon YJ, Han KH, Kim DY. Role of serum prothrombin induced by vitamin K absence or antagonist-II in the early detection of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *Scandinavian Journal of Gastroenterology* 2009;**44**(7):861-6.

Youns 2013 *{published data only}*

Youns MM, Abdel Wahab AH, Hassan ZA, Attia MS. Serum talin-1 is a potential novel biomarker for diagnosis of hepatocellular carcinoma in Egyptian patients. *Asian Pacific Journal of Cancer Prevention* 2013;**14**(6):3819-23.

Yu 2011 *{published data only}*

Yu NC, Chaudhari V, Raman SS, Lassman C, Tong MJ, Busuttil RW, et al. CT and MRI improve detection of hepatocellular carcinoma, compared with ultrasound alone, in patients with cirrhosis. *Clinical Gastroenterology and Hepatology* 2011;**9**(2):161-7.

Yu 2016 *{published data only}*

Yu R, Xiang X, Tan Z, Zhou Y, Wang H, Deng G. Efficacy of PIVKA-II in prediction and early detection of hepatocellular carcinoma: a nested case-control study in Chinese patients. *Scientific Reports* 2016;**6**:35050.

Yu 2020a *{published data only}*

Yu J, Ding WB, Wang MC, Guo X, Xu J, Xu Q, et al. Plasma circular RNA panel to diagnose hepatitis B virus-related hepatocellular carcinoma: a large-scale, multicenter study. *International Journal of Cancer. Journal International du Cancer* 2020;**146**(6):1754-63.

Yu 2020b *{published data only}*

Yu J, Ding WB, Wang MC, Guo X, Xu J, Xu Q, et al. Plasma circular RNA panel to diagnose hepatitis B virus-related hepatocellular carcinoma: a large-scale, multicenter study. *International Journal of Cancer. Journal International du Cancer* 2020;**146**(6):1754-63.

Yu 2020c *{published data only}*

Yu J, Ding WB, Wang MC, Guo X, Xu J, Xu Q, et al. Plasma circular RNA panel to diagnose hepatitis B virus-related hepatocellular carcinoma: a large-scale, multicenter study. *International Journal of Cancer. Journal International du Cancer* 2020;**146**(6):1754-63.

Zekri 2013 *{published data only}*

Zekri AR, El-Kassas M, Saad Y, Bahnassy A, El-Din HK, Darweesh SK, et al. Caspase recruitment domains. New potential markers for diagnosis of hepatocellular carcinoma associated with HCV in Egyptian patients. *Annals of Hepatology* 2013;**12**(5):774-81.

Zhan 2020 *{published data only}*

Zhan Z, Guan Y, Mew K, Zeng W, Peng M, Hu P, et al. Urine α-fetoprotein and orosomucoid 1 as biomarkers of hepatitis B virus-associated hepatocellular carcinoma. *American Journal of Physiology. Gastrointestinal and Liver Physiology* 2020;**318**(2):G305-12.

Zhang 2020 *{published data only}*

Zhang W, Fu T, Guo Z, Zhang, Y, Zhang L, Su H, et al. Serum miR-375 levels are closely related to disease progression from HBV infection to HBV-related hepatocellular carcinoma. *BioMed Research International* 2020;**2020**:5819385.

Zheng 2017 *{published data only}*

Zheng Y, Yang T, Soh A, Beshiri A. Application of PIVKA-II in hepatocellular carcinoma diagnosis in Chinese population: preliminary data from Asian multicenter study. *Tumor Biology* 2017;**2017**:4-5.

Zhou 2012 *{published data only}*

Zhou Z, Xia D, Wang C, Lin C, Zhao W, Dong C. Clinical evaluation of single or joint of Golgi protein 73 and alpha-fetoprotein in hepatocellular carcinoma diagnosing. *Chinese-German Journal of Clinical Oncology* 2012;**11**:650.

Zhou 2019 *{published data only}*

Zhou J, Yang W, Zhang S, He X, Lin J, Zhou T, et al. Diagnostic value of angiopoietin-like protein 2 for CHB-related hepatocellular carcinoma. *Cancer Management and Research* 2019;**11**:7159-69.

Zhu 2013 *{published data only}*

Zhu WW, Guo JJ, Guo L, Jia HL, Zhu M, Zhang JB, et al. Evaluation of midkine as a diagnostic serum biomarker in hepatocellular carcinoma. *Clinical Cancer Research* 2013;**19**(14):3994-54.

Zhu 2020 *{published data only}*

Zhu M, Zheng J, Wu F, Kang B, Liang J, Heskia F, et al. OPN is a promising serological biomarker for hepatocellular carcinoma diagnosis. Journal of Medical Virology 2020 Feb 11 [Epub ahead of print]. [DOI: 10.1002/jmv.25704]

Ziada 2016 *{published data only}*

Ziada DH, El Sadany S, Soliman H, Abd-Elsalam S, Salama M, Hawash N, et al. Prevalence of hepatocellular carcinoma in chronic hepatitis C patients in Mid Delta, Egypt: a single center study. *Journal of the Egyptian National Cancer Institute* 2016;**28**(4):257-62.

Zinkin 2008 *{published data only}*

Zinkin NT, Grall F, Bhaskar K, Otu HH, Spentzos D, Kalmowitz B, et al. Serum proteomics and biomarkers in hepatocellular carcinoma and chronic liver disease. *Clinical Cancer Research* 2008;**14**(2):470-7.

Zuo 2016 *{published data only}*

Zuo D, Chen L, Liu X, Wang X, Xi Q, Luo Y, et al. Combination of miR-125b and miR-27a enhances sensitivity and specificity of AFP-based diagnosis of hepatocellular carcinoma. *Tumour Biology* 2016;**37**(5):6539-49.

References to studies excluded from this review

Abdelgawad 2013 *{published data only}*

Abdelgawad IA, Mossallam GI, Radwan NH, Elzawahry HM, Elhifnawy NM. Can Glypican3 be diagnostic for early hepatocellular carcinoma among Egyptian patients? *Asian Pacific Journal of Cancer Prevention* 2013;**14**(12):7345-9.

Abd El Gawad 2014 *{published data only}*

Abd El Gawad IA, Mossallam GI, Radwan NH, Elzawahry HM, Elhifnawy NM. Comparing prothrombin induced by vitamin K absence-II (PIVKA-II) with the oncofetal proteins glypican-3, Alpha feto protein and carcinoembryonic antigen in diagnosing hepatocellular carcinoma among Egyptian patients. *Journal of Egyptian National Cancer Institute* 2014;**26**(2):79-85.

Abdel-Hafiz 2018 *{published data only}*

Abdel-Hafiz SM, Hamdy HE, Khorshed FM, Aboushousha TS, Safwat G, Saber MA, et al. Evaluation of osteopontin as a biomarker in hepatocellular carcinomas in Egyptian patients with chronic HCV cirrhosis. *Asian Pacific Journal of Cancer Prevention* 2018;**19**(4):10212-1027.

Abelev 1971 *{published data only}*

Abelev GI, Tsvetkov VS, Biryulina TI, Él'gort DA, Olovnikov AM, Gusev AI, et al. Value of highly sensitive methods of α-fetoprotein determination for the diagnosis of hepatocellular carcinoma and teratoblastoma. *Bulletin of Experimental Biology and Medicine* 1971;**71**(4):75-81.

Abouzied 2017 *{published data only}*

Abouzied MM, Nazmy MH, Mohamed RM, Fawzy M, Eltahir H. Diagnostic utility of leptin and insulin-like growth factor binding protein-2 in hepatocellular carcinoma of diabetic and non-diabetic Egyptian patients. *Tropical Journal of Pharmaceutical Research* 2017;**16**(1):211-8.

Aburano 1979 *{published data only}*

Aburano T, Kuwajima A, Tada A, Ichiyanagi K, Tonami N, Hisada K. Combined use of nuclear medicine and ultrasound in the diagnosis of hepatic tumor (author's transl). *Rinsho Hoshasen* 1979;**24**(7):729-37.

Asim 2010 *{published data only}*

Asim M, Malik A, Sarma MP, Polipalli SK, Begum N, Ahmad I, et al. Hepatitis B virus BCP, Precore/core, X gene mutations/genotypes and the risk of hepatocellular carcinoma in India. *Journal of Medical Virology* 2010;**82**:1115-25.

Åström 2017 *{published data only}*

Åström E, Stål P, Zenlander R, Edenvik P, Alexandersson C, Haglund M, et al. Reverse lectin ELISA for detecting fucosylated forms of α1-acid glycoprotein associated with hepatocellular carcinoma. *PLOS One* 2017;**12**(3):e0173897.

Bago 1993 *{published data only}*

Bago J, Bilic A, Babic Z, Babic D, Jagic V, Andrijevic B, et al. The role of alpha-fetoprotein in differentiation between primary and secondary liver tumors. *Croatian Journal of Gastroenterology and Hepatology* 1993;**2**(2):45-8.

Baig 2009 *{published data only}*

Baig JA, Alam JM, Mahmood SR, Baig M, Shaheen R, Sultana I, et al. Hepatocellular carcinoma (HCC) and diagnostic significance of A-fetoprotein (AFP). *Journal of Ayub Medical College, Abbottabad* 2009;**21**(1):72-5.

Banales 2019 *{published data only}*

Banales JM, Iñarrairaegui M, Arbelaiz A, Milkiewicz P, Muntané J, Muñoz-Bellviset L, et al. Serum metabolites as diagnostic biomarkers for cholangiocarcinoma, hepatocellular carcinoma, and primary sclerosing cholangitis. *Hepatology (Baltimore, Md.)* 2019;**70**(2):547-62.

Bao 2013 *{published data only}*

Bao YX, Yang Y, Zhao HR, Mao R, Xiao L, Zhang YF, et al. Clinical significance and diagnostic value of Golgi-protein 73 in patients with early-stage primary hepatocellular carcinoma. *Zhonghua Zhong Liu Za Zhi* 2013;**35**(7):505-8.

Baumgarten 2001 *{published data only}*

Baumgarten R. Early diagnosis of HCC: how effective are screening examinations [Fruhdiagnose HCC - wie effektiv sin screeninguntersuchungen?]. *Die Medizinische Welt* 2001;**52**:137-40.

Beaugrand 2000 *{published data only}*

Beaugrand M. Screening for hepatocellular carcinoma in cirrhotic patients [Depistage du carcinome hepatocellulaire sur cirrhose]. *Gastroenterologie Clinique et Biologique* 2000;**24**:892-4.

Ben Hassine 2007 *{published data only}*

Ben Hassine L, Daghfous MH, Mami A, Chammakhi-Jemli C, Zouaoui W, Saddoud W, et al. Imaging in the screening and diagnosis of hepatocellular carcinoma in cirrhosis liver in Tunisia. A series of 30 cases [L'imagerie dans le dépistage et le diagnostic du carcinome hépatocellulaire sur foie de cirrhose en Tunisie. A propos de 30 cas]. *La Tunisie Medicale* 2007;**85**(5):421-6.

Bialecki 2006 *{published data only}*

Bialecki ES, Ezenekwe AM, Brunt EM, Collins BT, Ponder TB, Bieneman BK, et al. Comparison of liver biopsy and noninvasive methods for diagnosis of hepatocellular carcinoma. *Clinical Gastroenterology and Hepatology* 2006;**4**(3):361-8.

Bird 2016 *{published data only}*

Bird TG, Dimitropoulou P, Turner RM, Jenks SJ, Cusack P, Hey S, et al. Alpha-Fetoprotein detection of hepatocellular carcinoma

leads to a standardized analysis of dynamic AFP to Improve screening based detection. *PLOS One* 2016;**11**(6):e0156801.

Biwole Sida 1992 *{published data only}*

Biwole-Sida M, Amvene N, Oyono E, Mbakop A, Manfo C, Tapko JB, et al. Screening of hepatocellular carcinoma in the middle of a high risk population in Cameroon [Dépistage du carcinome hépatocellulaire au sein d'une population de sujets à haut risque au Cameroun]. *Annales de Gastroenterologie et d'Hepatologie* 1992;**28**(5):213-6.

Bolondi 1990 *{published data only}*

Bolondi L, Benzi G, Santi V, Gaiani S, Li Bassi SL, Zironi G, et al. Relationship between alpha-fetoprotein serum levels, tumour volume and growth rate of hepatocellular carcinoma in a western population. *Italian Journal of Gastroenterology* 1990;**22**:190-4.

Bottelli 1998 *{published data only}*

Bottelli R, Tibballs J, Hochhauser D, Watkinson A, Dick R, Burroughs AK. Ultrasound screening for hepatocellular carcinoma (HCC) in cirrhosis: the evidence for an established clinical practice. *Clinical Radiology* 1998;**53**(10):713-6.

Bowry 1980 *{published data only}*

Bowry TR, Shaha MV. A study of hepatitis Bs antigen titres and alpha foetoprotein levels in primary hepatocellular carcinoma in Kenya. *East African Medical Journal* 1980;**57**(6):382-9.

Bröker 2014 *{published data only}*

Bröker ME, Ijzermans JN, Witjes CD, van Vuuren HJ, de Man RA. The predictive value of Golgi protein 73 in differentiating benign from malignant liver tumors. *PLOS One* 2014;**9**(7):e100187.

Buell 2001 *{published data only}*

Buell JF, Layman R, Yoshida A, Cronin DC, Kim G, Hart J, et al. Efficacy of computed tomography in detecting hepatocellular cancer in adult liver transplant candidates. *Transplantation Proceedings* 2001;**33**(1-2):1854-5.

Cai 2019 *{published data only}*

Cai J, Chen L, Zhang Z, Zhang X, Lu X, Liu W, et al. Genome-widemapping of 5-hydroxymethylcytosinesin circulating cell-freeDNA as a non-invasive approach for early detection of hepatocellular carcinoma. *Gut* 2019;**68**:2195–205.

Carriere 1993 *{published data only}*

Carriere, F, Jammet P, Legmann P, Hazebroucq V, Pariente D, Abecassis JP, et al. Monitoring of idiopathic liver hemochromatosis. Comparison of ultrasound and MRI in 46 patients [Surveillance hepatique des sujects porteur d'une hemochromatose idiopatique. Compairaison echographie et IRM chez 46 patients]. *Revue d'Imagerie Medicale* 1993;**5**(4):239-45.

Chen 1995 *{published data only}*

Chen CJ, Lu SN, You SL, Wu MH, Wang LY, Lee LT, et al. Community-based hepatocellular carcinoma screening in seven townships in Taiwan. *Journal of the Formosan Medical Association* 1995;**94**(Suppl 2):94-102.

Chen 2002 *{published data only}*

Chen TH, Chen CJ, Yen MF, Lu SN, Sun CA, Huang GT, et al. Ultrasound screening and risk factors for death from hepatocellular carcinoma in a high risk group in Taiwan. *International Journal of Cancer. Journal International du Cancer* 2002;**98**:257-61.

Chen 2010 *{published data only}*

Chen Y, Zhoub Y, Qiuc S, Wang K, Liua S, Peng XX, et al. Autoantibodies to tumor-associated antigens combined with abnormal alpha-fetoprotein enhance immunodiagnosis of hepatocellular carcinoma. *Cancer Letters* 2010;**289**(1):32-9.

Chen 2011 *{published data only}*

Chen F, Xue J, Zhou L, Wu S, Chen Z. Identification of serum biomarkers of hepatocarcinoma through liquid chromatography/mass spectrometry-based metabonomic method. *Analytical and Bioanalytical Chemistry* 2011;**401**:1899-904.

Chen 2013 *{published data only}*

Chen M, Li G, Yan J, Lu X, Cui J, Ni Z, et al. Reevaluation of glypican-3 as a serological marker for hepatocellular carcinoma. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2013;**423**:105-11.

Cheng 2009 *{published data only}*

Cheng KS, Tang HL, Chou FT, Chou JW, Hsu CH, Yu CJ, et al. Cytokine evaluation in liver cirrhosis and hepatocellular carcinoma. *Hepato-gastroenterology* 2009;**56**(93):1105-10.

Choi 2013 *{published data only}*

Choi JY, Jung SW, Kim HY, Kim M, Kim Y, Kim DG, et al. Diagnostic value of AFP-L3 and PIVKA-Ⅱ in hepatocellular carcinoma according to total-AFP. *World Journal of Gastroenterology* 2013;**19**(3):339-46.

Chun 2015 *{published data only}*

Chun S, Rhie SY, Ki CS, Kim JE, Park HD. Evaluation of alpha-fetoprotein as a screening marker for hepatocellular carcinoma in hepatitis prevalent areas. *Annals of Hepatology* 2015;**14**(6):882-8.

Colombo 1991 *{published data only}*

Colombo M, De Franchis R, Del Ninno E. Hepatocellular carcinoma in Italian patients with cirrhosis. *New England Journal of Medicine* 1991;**325**(10):675.

Cui 2013 *{published data only}*

Cui Z, Yu X, Guo L, Wei Y, Zheng S, Li W, et al. Combined analysis of serum alpha-fetoprotein and mage-a3-specific cytotoxicT lymphocytes in peripheral blood for diagnosis of hepatocellular carcinoma. *Disease Markers* 2013;**35**(6):915-23.

Del Vecchio-Blanco 1977 *{published data only}*

Del Vecchio-Blanco C, Caporaso N, Balzano A, Grande R. Comparative study of the plasma levels of alpha-1 fetoprotein and carcinoembryonic antigen in chronic liver diseases and malignant neoplasias. *Quaderni Sclavo di Diagnostica Clinica e di Laboratorio* 1977;**13**(3):288-97.

Dengler 2017 *{published data only}*

Dengler M, Staufer K, Huber H, Stauber R, Bantel H, Weiss KH, et al. Soluble Axl is an accurate biomarker of cirrhosis and hepatocellular carcinoma development: results from a large scale multicenter analysis. *Oncotarget* 2017;**8**(28):46234-48.

Deshpande 1981 *{published data only}*

Deshpande AD, Patil SD, Talib SH. Comparative study of needle biopsy and alpha-foetoprotein detection in the diagnosis of primary hepatocellular carcinoma. *Journal of the Association of Physicians of India* 1981;**29**(2):129-32.

Di Martino 2013 *{published data only}*

Di Martino M, De Filippis G, De Santis A, Geiger D, Del Monte M, Lombardo CV, et al. Hepatocellular carcinoma in cirrhotic patients: prospective comparison of US, CT and MR imaging. *European Radiology* 2013;**23**(4):887-96.

Ding 1995 *{published data only}*

Ding S, Li H, Shun S. Evaluation of the combined assay of alpha-fetoprotein and gamma-glutamyl transpeptidase isoenzyme II in the diagnosis of primary hepatic carcinoma. *Zhonghua Nei Ke Za Zhi* 1995;**34**(10):690-2.

Di Poto 2017 *{published data only}*

Di Poto C, Ferrarini A, Zhao Y, Varghese RS, Tu C, Zuo Y, et al. Metabolomic characterization of hepatocellular carcinoma in patients with liver cirrhosis for biomarker discovery. *Cancer Epidemiology, Biomarkers & Prevention* 2017;**26**(5):675-83.

Di Poto 2018 *{published data only}*

Di Poto C, He S, Varghese RS, Zhao Y, Ferrarini A, Su S, et al. Identification of race-associated metabolite biomarkers for hepatocellular carcinoma in patients with liver cirrhosis and hepatitis C virus infection. *PLOS One* 2018;**13**(3):e0192748.

Divella 2012 *{published data only}*

Divella R, Daniele A, Gadaleta C, Tufaro A, Venneri MT, Paradiso A, et al. Circulating transforming growth factor-β and epidermal growth factor receptor as related to virus infection in liver carcinogenesis. *Anticancer Research* 2012;**32**:141-6.

Donato 1995 *{published data only}*

Donato F, Rodella S, Chiesa R, Picoco C, Donati LF, Nardi G. Diagnostic accuracy for primary liver cancer. implications for cancer registration. *Tumori* 1995;**81**:86-90.

Dong 2008 *{published data only}*

Dong ZZ, Yao DF, Yao M, Qiu LW, Zong L, Wu W, et al. Clinical impact of plasma TGF-beta1 and circulating TGF-beta1 mRNA in diagnosis of hepatocellular carcinoma. *Hepatobiliary & Pancreatic Diseases International* 2008;**7**(3):288-95.

Dou 2016 *{published data only}*

Dou CY, Fan YC, Cao CJ, Yang Y, Wang K. Sera DNA methylation of CDH1, DNMT3b and ESR1 promoters as biomarker for the early diagnosis of hepatitis B virus-related hepatocellular carcinoma. *Digestive Diseases and Sciences* 2016;**61**(4):1130-8.

El Abd 2016 *{published data only}*

El Abd N, Fawzy A, Elbaz T, Hamdy S. Evaluation of annexin A2 and as potential biomarkers for hepatocellular carcinoma. *Tumour Biology* 2016;**37**:211-6.

El-Ahwany 2019 *{published data only}*

El-Ahwany EG, Mourad L, Zoheiry MM, Abu-Taleb H, Hassan M, Atta R, et al. MicroRNA-122a as a non-invasive biomarker for HCV genotype 4-related hepatocellular carcinoma in Egyptian patients. *Archives of Medical Science* 2019;**15**(6):1454–61.

El-Attar 2010 *{published data only}*

El-Attar HA, Kandil MH, El-Kerm YM, El-Ghandou MK. Comparison of serum survivin and alpha fetoprotein in Egyptian patients with hepatocellular carcinoma associated with hepatitis C viral infection. *Asian Pacific Journal of Cancer Prevention* 2010;**11**:897-903.

El Azm 2013 *{published data only}*

El Azm AR, Yousef M, Salah R, Mayah W, Tawfeek S, Ghorabah H, et al. Serum anti-P53 antibodies and alpha-fetoprotein in patients with non-B non-C hepatocellular carcinoma. *Springer Plus* 2013;**2**:69.

El-Emshaty 2014 *{published data only}*

EI-Emshaty HM, Gadelhak SA, Abdelaziz MM, Abbas AT, Gadelhak NA. Serum P53 Abs in HCC patients with viral hepatitis - type C. *Hepato-gastroenterology* 2014;**134**:1688-95.

El-Emshaty 2015 *{published data only}*

El-Emshaty HM, Nasif WA, Mohamed IE. Serum cytokine of il-10 and il-12 in chronic liver disease: the immune and inflammatory response. *Disease Markers* 2015;**2015**:707254.

El Gedawy 2017 *{published data only}*

El Gedawy G, Obada M, Kelani A, El-Said H, Ghanayem NM. Circulating MiRNA-21 and programed cell death (PDCD) 4 gene expression in hepatocellular carcinoma (HCC) in Egyptian patients. *Egyptian Journal of Medical Human Genetics* 2017;**18**:137-45.

Elghoroury 2017 *{published data only}*

Elghoroury EA, Abdel Maksoud SA, Farouk H, Shalabi AA, Hassan EM, Hassan M, et al. Circulating MicroRNAs: new diagnostic approach for hepatocellular carcinoma in viral hepatitis in Egyptians. *Research Journal of Pharmaceutical, Biological and Chemical Sciences* 2017;**8**(2):1307-16.

El-Mazny 2014 *{published data only}*

El-Mazny A, Sayed M, Sharaf S. Human telomerase reverse transcriptase messenger RNA (TERT mRNA) as a tumour marker for early detection of hepatocellular carcinoma. *Arab Journal of Gastroenterology* 2014;**15**:68-71.

El-Saadany 2018 *{published data only}*

El-Saadany S, El-Demerdash T, Helmy A, Mayah WW, El-Sayed Hussein B, Hassanien M, et al. Diagnostic value of Glypican-3 for hepatocellular carcinomas. *Asian Pacific Journal of Cancer Prevention* 2018;**19**(3):811-7.

El-Serag 2005 *{published data only}*

El-Serag HB, Mallat DB, Rabeneck L. Management of the single liver nodule in a cirrhotic patient: a decision analysis model. *Journal of Clinical Gastroenterology* 2005;**39**:152-9.

Elsewify 2020 *{published data only}*

Elsewify WA, Hassan EH, Mekky MA, El-Din AS, El-Rehim A, El-Abdeen Z, et al. Usefulness of circulating methylated p16 as a noninvasive molecular biomarker for hepatitis C-related hepatocellular carcinoma with normal serum alpha-fetoprotein levels. *International Journal of General Medicine* 2020;**13**:147–55.

Elshimi 2018 *{published data only}*

Elshimi E, Sakr MA, Morad WS, Mohammad L. Optimizing the diagnostic role of alpha-fetoprotein and abdominal ultrasound by adding overexpressed blood mRNA matrix metalloproteinase-12 for diagnosis of HCV-related hepatocellular carcinoma. *Gastrointestinal Tumors* 2018;**5**:100-8.

Eltabbakh 2015 *{published data only}*

Eltabbakh M, Zaghla H, Abdel-Razek W, Elshinnawy H, Ezzat S, Gomaa A, et al. Utility and cost-effectiveness of screening for hepatocellular carcinoma in a resource-limited setting. *Medical Oncology (Northwood, London, England)* 2015;**32**(1):432.

El-Zefzafy 2015 *{published data only}*

El-Zefzafy W, Eltokhy H, Mohamed NA, Abu-Zahab Z. Significance of serum cytokeratin-18 in prediction of hepatocellular carcinoma in chronic hepatitis C infected Egyptian patients. *Macedonian Journal of Medical Sciences* 2015;**3**(1):117-23.

Esfeh 2020 *{published data only}*

EsfehJM, Hajifathalian K, Ansari-Gilani K. Sensitivity of ultrasound in detecting hepatocellular carcinoma in obese patients compared to explant pathology as the gold standard. *Clinical and Molecular Hepatology* 2020;**26**:54-9.

Fan 2017 *{published data only}*

Fan SS, Liao CS, Cao YD, Xiao PL, Deng T, Luo RC, et al. A low serum Tat-interacting protein 30 level is a diagnostic and prognostic biomarker for hepatocellular carcinoma. *Oncology Letters* 2017;**13**:4208-14.

Farag 2018 *{published data only}*

Farag RM, AlAyobi D, Alsaleh KA, Kwon HJ, EL-Ansary A, Dawoud EA. Influence of glypican-3 as a newly diagnostic biomarker. *Biomedical & Pharmacology Journal* 2018;**11**(4):1789-96.

Fouad 2014 *{published data only}*

Fouad SA, Elsaaid NH, Mohamed NA, Abutaleb OM. Diagnostic value of serum level of soluble tumor necrosis factor receptor IIα in Egyptian patients with chronic hepatitis C virus infection and hepatocellular carcinoma. *Hepatitis Monthly* 2014;**14**(9):e19346.

Fouad 2015 *{published data only}*

Fouad SA, Mohamed NA, Fawzy MW, Moustafa DA. Plasma osteopontin level in chronic liver disease and hepatocellular carcinoma. *Hepatitis Monthly* 2015;**15**(9):e30753.

Frey 2015 *{published data only}*

Frey RS, Boldanova T, Heim MH. Ultrasound surveillance for hepatocellular carcinoma: real-life performance in a hepatology outpatient clinic. *Swiss Medical Weekly* 2015;**145**:w14200.

Fujiyama 1991 *{published data only}*

Fujiyama S, Izuno K, Gohshi K, Shibata J, Sato T. Clinical usefulness of des-gamma-carboxy prothrombin assay in early diagnosis of hepatocellular carcinoma. *Digestive Diseases and Sciences* 1991;**36**(12):1787-92.

Gandolfi 1987 *{published data only}*

Gandolfi L, Solmi L, Bertoni F, Muratori R, Stasi G. Small hepatocellular carcinoma: an Italian experience. *Hepato-gastroenterology* 1987;**34**(3):100-2.

Ganne-Carrié 1996 *{published data only}*

Ganne-Carrié N, Chastang C, Chapel F, Munz C, Pateron D, Sibony M, et al. Predictive score for the development of hepatocellular carcinoma and additional value of liver large cell dysplasia in Western patients with cirrhosis. *Hepatology (Baltimore, Md.)* 1996;**23**(5):1112-8.

Gao 2012 *{published data only}*

Gao DM, Sun L, Guo K, Li Y, Liu YK, Kang XN. Applicability of the multiplex quantitative antibody array system for early diagnosis of hepatocellular carcinoma. *Zhonghua Gan Zang Bing Za Zhi* 2012;**20**(10):785-8.

Geramizadeh 2013 *{published data only}*

Geramizadeh B, Nikeghbalian S, Kazemi K, Shamsaifar A, Bahador A, Salahi H, et al. Hepatocellular carcinoma in explanted livers of patients with genotype D HBV cirrhosis: report of the first experience from Iran. *Archives of Iranian Medicine* 2013;**16**(6):348-50.

Geramizadeh 2017 *{published data only}*

Geramizadeh B, Baghernezhad M, Salehi H, Nikeghbalian S, Shamsaeefar A, Kazemi K, et al. Clinicopathological discrepancies in the diagnosis of hepatocellular carcinoma in explanted livers, a single center study on more than 1500 transplanted livers. *Hepatitis Monthly* 2017;**17**(10):e11836.

Gheorghe 1986 *{published data only}*

Gheorghe N, Boerescu I, Ciuntu L, Coman M, Boros I. Serum alphaphetoprotein in chronic hepatitis scintigraphic correlations. *Physiologie (Bucarest)* 1986;**23**(2):131-8.

Gheorghe 2017 *{published data only}*

Gheorghe L, Iacob S, Prelipcean CC, Mihai C, Grasu M, Dumitru R, et al. Alpha fetoprotein and a useful tool for surveillance of interferon-free treated cirrhotic patients for hepatocellular carcinoma after SVR. *Gastroenterology* 2017;**152**(5 Suppl 1):S298.

Giangregorio 2009 *{published data only}*

Giangregorio F, Comparato G, Marinone MG, Di Stasi M, Sbolli G, Aragona G, et al. Imaging detection of new HCCs in cirrhotic patients treated with different techniques: comparison of conventional US, spiral CT, and 3-dimensional contrast-enhanced US with the navigator technique (Nav 3D CEUS). *Journal of Ultrasound* 2009;**12**(1):12-21.

Giannini 2012 *{published data only}*

Giannini EG, Marenco S, Borgonovo G, Savarino V, Farinati F, Del Poggio P, et al, Italian Liver Cancer (ITALICA) group. Alpha-fetoprotein has no prognostic role in small hepatocellular carcinoma identified during surveillance in compensated cirrhosis. *Hepatology (Baltimore, Md.)* 2012;**56**(4):1371-9.

Giannini 2013 *{published data only}*

Giannini EG, Cucchetti A, Erroi V, Garuti F, Odaldi F, Trevisani F. Surveillance for early diagnosis of hepatocellular carcinoma: how best to do it? *World journal of Gastroenterology* 2013;**19**(47):8808-21.

Giannini 2014 *{published data only}*

Giannini EG, Trevisani F. Accuracy of α-fetoprotein measurement in detection of hepatocellular carcinoma--1 more nail in the coffin. *Clinical Gastroenterology and Hepatology* 2014;**12**(12):2138-9.

Giusti 2005 *{published data only}*

Giusti S, Donati F, Paolicchi A, Bartolozzi C. Hepatocellular adenoma: imaging findings and pathological correlation. *Digestive and Liver Disease* 2005;**37**(3):200-5.

Goldaracena 2019 *{published data only}*

Goldaracena N, Mehta N, Scalera I, Sposito C, Atenafu EG, Yao FY, et al. Multicenter validation of a score to predict prognosis after the development of HCC recurrence following liver transplantation. *HPB : the Official Journal of the International Hepato Pancreato Biliary Association* 2019;**21**(6):731-8.

Gomez Rodriguez 2012 *{published data only}*

Gómez-Rodríguez R, Romero-Gutiérrez M, Artaza-Varasa T, González-Frutos C, Ciampi-Dopazo JJ, de-la-Cruz-Pérez G, et al. The value of the Barcelona Clinic Liver Cancer and alpha-fetoprotein in the prognosis of hepatocellular carcinoma. *Revista Espanola de Enfermedades Digestivas* 2012;**104**(6):298-304.

Gomez Rubio 2005 *{published data only}*

Gómez-Rubio M, Moya-Valdés M, García J. Diagnostic laparoscopy and laparoscopic ultrasonography with local anesthesia in hepatocellular carcinoma. *World Journal of Gastroenterology* 2005;**11**(26):4120-3.

Gorbatenko 1974 *{published data only}*

Gorbatenko VP, Zolotarenskiĭ VB, Mikhaĭlova VS. Diagnostic significance of the study of alpha-fetoprotein in liver cirrhosis and primary cancer of the liver. *Sovetskaia Meditsina* 1974;**37**(6):128-30.

Gounder 2016 *{published data only}*

Gounder PP, Bulkow LR, Meltzer MI, Bruce MG, Hennessy TW, Snowball M, et al. Cost-effectiveness analysis of hepatocellular carcinoma screening by combinations of ultrasound and alphafetoprotein among Alaska Native people, 1983-2012. *International Journal of Circumpolar Health* 2016;**75**:31115.

Grąt 2016 *{published data only}*

Grąt M, Krasnodębski M, Patkowski W, Wronka KM, Masior Ł, Stypułkowski J, et al. Relevance of pre-transplant a-fetoprotein dynamics in liver transplantation for hepatocellular cancer. *Annals of Transplantation* 2016;**18**(21):115-224.

Ha 2012 *{published data only}*

Ha NB, Ha NB, Ahmed A, Ayoub W, Daugherty TJ, Chang ET, et al. Risk factors for hepatocellular carcinoma in patients with chronic liver disease: a case-control study. *Cancer Causes & Control* 2012;**23**(3):455-62.

Hacque 2016 *{published data only}*

Haque SS, Kumari R, Muzaffar A, Kumar U, Sharan A, Kumari B. Estimation of serum alpha feto-protein (AFP), interleukin-6 and des-γ-carboxyprothrombin (DCP) incase of hepatocellular carcinoma. *Bangladesh Journal of Medical Science* 2016;**15**(2):230-3.

Hagag 2020 *{published data only}*

Hagag NA, Ali YB, Elsharawy AA, Talaat RM. Clinical impact of circulated miR-1291 in plasma of patients with liver cirrhosis (LC) and hepatocellular carcinoma (HCC): implication on glypican-3 expression. *Gastrointestinal Cancer* 2020;**51**(1):234-41.

Hajiani 2005 *{published data only}*

Hajiani E, Masjedizadeh R, Hashemi J, Azmi M, Rajabi T. Risk factors for hepatocellular carcinoma in Southern Iran. *Saudi Medical Journal* 2005;**26**(6):974-7.

Han 2018b *{published data only}*

Han C, Gao L, Bai H, Dou X. Identification of a role for serum aldo-keto reductase family 1 member B10 in early detection of hepatocellular carcinoma. *Oncology Letters* 2018;**16**:7123-30.

Hanaoka 2011 *{published data only}*

Hanaoka T, Sato S, Tobita H, Miyake T, Ishihara S, Akagi S, et al. Clinical significance of the highly sensitive fucosylated fraction of α-fetoprotein in patients with chronic liver disease. *Journal of Gastroenterology and Hepatology* 2011;**26**(4):739-44.

Hashemi 2008 *{published data only}*

Hashemi J, Esmaeilzadeh A, Dabagh Kakhaki VR, Hosseini MR. Accuracy of gray-scale and colordoppler sonography in diagnosis of hepatic hemangioma, hepatocellular carcinoma and liver metastasis. *Iranian Journal of Radiology* 2008;**5**(3):129-34.

Hass 2017 *{published data only}*

Hass HG, Markmann H-U, Schäffer M, Wellhäußer U, Smith U, Denzlinger C. Diagnostic and prognostic aspects of hepatocellular carcinoma - a retrospective analysis in 145

patients. *Journal of Gastroenterology and Hepatology Research* 2017;**6**(3):2358-64.

Hemken 2019 *{published data only}*

Hemken PM, Sokoll LJ, Yang X, Dai J, Elliott D, Gawel SH, et al. Validation of a novel model for the early detection of hepatocellular carcinoma. *Clinical Proteomics* 2019;**16**:2-9.

Hernandez 2011 *{published data only}*

Hernandez JC, Samada M, Roque A, Cruz Y, Howland I, Fernandez I. Diagnostic value of alpha-fetoprotein for hepatocellular carcinoma. *Biotecnologia Aplicada* 2011;**28**:34-9.

Heyward 1985 *{published data only}*

Heyward WL, Lanier AP, McMahon BJ, Fitzgerald MA, Kilkenny S, Paprocki TR. Early detection of primary carcinoma. Screening of primary hepatocellular carcinoma among persons infected with hepatitis B virus. *JAMA* 1985;**254**(21):3052-4.

Hiraoka 2016 *{published data only}*

Hiraoka A, Ochi M, Matsuda R, Aibiki T, Okudaira T, Kawamura T, et al. Ultrasonography screening for hepatocellular carcinoma in Japanese patients with diabetes mellitus. *Journal of Diabetes* 2016;**8**(5):640-6.

Hu 2010 *{published data only}*

Hu JS, Wu DW, Liang S, Miao XY. GP73, a resident Golgi glycoprotein, is sensibility and specificity for hepatocellular carcinoma of diagnosis in a hepatitis B-endemic Asian population. *Medical Oncology (Northwood, London, England)* 2010;**27**(2):339-45.

Hussein 2008 *{published data only}*

Hussein SM, Ibrahim AA, Abdella HM, Montasser IF. Evaluation of serum squamous cell carcinoma antigen as a novel biomarker for diagnosis of hepatocellular carcinoma in Egyptian patients. *Indian Journal of Cancer* 2008;**45**(4):167-72.

Hwang A 2018 *{published data only}***10.1371/journal.pone.0198937**

Hwang A, Shi C, Zhu E, Naaz F, Zhou P, Rasheed Z, et al. Supervised learning reveals circulating biomarker levels diagnostic of hepatocellular carcinoma in a clinically relevant model of non-alcoholic steatohepatitis; an OAD to NASH. *PLOS One* 2018;**13**(6):e0198937.

Hwang H 2018 *{published data only}*

Hwang HM, Heo CK, Lee HJ, Kwak SS, Lim WH, Yoo JS, et al. Identification of `anti-SF3B1` autoantibody as a diagnostic marker in patients with hepatocellular carcinoma. *Journal of Translational Medicine* 2018;**16**:177.

Imberti 1993 *{published data only}*

Imberti D, Fornari F, Sbolli G, Buscarini E, Squassante L, Buscarini L. Hepatocellular carcinoma in liver cirrhosis. A prospective study. *Scandinavian Journal of Gastroenterology* 1993;**28**(6):540-4.

Izuno 1995 *{published data only}*

Izuno K, Fujiyama S, Yamasaki K, Sato M, Sato T. Early detection of hepatocellular carcinoma associated with cirrhosis by combined assay of des-gamma-carboxy prothrombin and alpha-fetoprotein: a prospective study. *Hepato-gastroenterology* 1995;**42**(4):387-93.

Izzo 1998 *{published data only}*

Izzo F, Cremona F, Ruffolo F, Palaia R, Parisi V, Curley SA. Outcome of 67 patients with hepatocellular cancer detected during screening of 1125 patients with chronic hepatitis. *Annals of Surgery* 1998;**27**(4):513-8.

Jirun 2011 *{published data only}*

Jirun P, Zhang G, Kim HK, Ha SA, Zhongtian J, Shishi Q, et al. Clinical utility of alpha fetoprotein and HCCR-1, alone or in combination, in patients with chronic hepatitis, liver cirrhosis and hepatocellular carcinoma. *Disease Markers* 2011;**30**(6):307-15.

Johnson 1997 *{published data only}*

Johnson PJ, Leung N, Cheng P, Welby C, Leung WT, Lau WY. Hepatoma-specific alpha fetoprotein may permit preclinical diagnosis of malignant change in patients with chronic liver disease. *British Journal of Cancer* 1997;**75**(2):236-40.

Jun 2019 *{published data only}*

Jun T, Hsu UC, Ogawa S, Huang YT, Yeh ML, Tseng CH, et al. Mac-2 binding protein glycosylation isomer as a hepatocellular carcinoma marker in patients with chronic hepatitis B or C infection. *Hepatology Communications* 2019;**3**(4):493-503.

Junna 2017 *{published data only}*

Junna Z, Gongde C, Jinying X, Xiu Z. Serum AFU, 5'-NT and AFP as biomarkers for primary hepatocellular carcinoma diagnosis. *Open Medicine* 2017;**12**:254-358.

Kim 2011a *{published data only}*

Kim SE, Lee HC, Shim JH, Park HJ, Kim KM, Kim PN, et al. Noninvasive diagnostic criteria for hepatocellular carcinoma in hepatic masses 42 cm in a hepatitis B virus-endemic area. *Liver International* 2011;**31**(10):1468-76.

Kim 2011b *{published data only}*

Kim SE, Lee HC, Shim JH, Park HJ, Kim KM, Kim PN, et al. Noninvasive diagnostic criteria for hepatocellular carcinoma in hepatic masses 42 cm in a hepatitis B virus-endemic area. *Liver International* 2011;**31**(10):1468-76.

Kim 2013 *{published data only}*

Kim JM, Kwon CH, Joh JW, Park JB, Lee JH, Kim SJ, et al. PIVKA-II is a useful marker in patients with modified UICC T3 stage hepatocellular carcinoma. *Hepato-gastroenterology* 2013;**60**(126):1456-62.

Kim 2015 *{published data only}*

Kim SU, Park JH, Kim HS, Lee JM, Lee HG, Kim H, et al. Serum dickkopf-1 as a biomarker for the diagnosis of hepatocellular carcinoma. *Yonsei Medical Journal* 2015;**56**(5):1296-306.

Kim 2017 *{published data only}*

Kim H, Park YJ, Kim Y, Sohn A, Yeo I, Yu SJ, et al. Serum fibronectin distinguishes the early stages of hepatocellular carcinoma. *Scientific Reports* 2017;**7**(1):9449.

King 1989 *{published data only}*

King MA, Kew MC, Kuyl JM, Atkinson PM. A comparison between des-gamma-carboxy prothrombin and alpha-fetoprotein as markers of hepatocellular carcinoma in southern African blacks. *Journal of Gastroenterology and Hepatology* 1989;**4**(1):17-24.

Kiyokawa 2017 *{published data only}*

Kiyokawa H, Yasuda H, Oikawa R, Okuse C, Matsumoto N, Ikeda H, et al. Serum monomeric laminin-c2 as a novel biomarker for hepatocellular carcinoma. *Cancer Science* 2017;**108**:1432-9.

Ko 2017 *{published data only}*

Ko YS, Bae JH, Sinn DH, Gwak GY, Kang W, Paik YH, et al. The clinical significance of serum alpha-fetoprotein in diagnosing hepatocellular carcinoma in a health screening population. *Korean Journal of Gastroenterology* 2017;**69**(4):232-8.

Kobayashi 1985 *{published data only}*

Kobayashi K, Sugimoto T, Makino H, Kumagai M, Unoura M, Tanaka N, et al. Screening methods for early detection of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 1985;**5**(6):1100-5.

Kobeisy 2012 *{published data only}*

Kobeisy MA, Morsy KH, Galal M, Sayed SK, Ashmawy MM, Mohammad FM. Clinical significance of elevated alpha-foetoprotein (AFP) in patients with chronic hepatitis C without hepatocellular carcinoma in upper Egypt. *Arab Journal of Gastroenterology* 2012;**13**(2):49-53.

Kuromatzo 1993 *{published data only}*

Kuromatsu R, Tanaka M, Tanikawa K. Serum alpha-fetoprotein and lens culinaris agglutinin-reactive fraction of alpha-fetoprotein in patients with hepatocellular carcinoma. *Liver* 1993;**13**:177-82.

Larcos 1998 *{published data only}*

Larcos G, Sorokopud H, Berry G, Farrel GC. Sonographic screening for hepatocellular carcinoma in patients with chronic hepatitis or cirrhosis: an evaluation. *AJR. American Journal of Roentgenology* 1998;**171**:433-5.

Lee 2016 *{published data only}*

Lee CS, Kim BH, Chung JW, Lee JH, Sanghyuk I, Jang ES, et al. Concomitant measurement of alpha-fetoprotein improves effectiveness of ultrasound-based surveillance of hepatocellular carcinoma in chronic hepatitis B patients. *Hepatology (Baltimore, Md.)* 2016;**64**(1):219A.

Lee 2017 *{published data only}*

Lee H, Yoon JH, Kim H, Yi NJ, Hong SK, Yoon KC, et al. False positive diagnosis of hepatocellular carcinoma in liver resection patients. *Journal of Korean Medical Science* 2017;**32**(2):315-20.

Leerapun 2007 *{published data only}*

Leerapun A, Suravarapu SV, Bida JP, Clark RJ, Sanders EL, Mettler TA, et al. The utility of lens culinaris agglutinin-reactive alpha-fetoprotein in the diagnosis of hepatocellular carcinoma: evaluation in a United States referral population. *Clinical Gastroenterology and Hepatology* 2007;**5**(3):394-402.

Li 2016 *{published data only}*

Li S, Wang H, Fan J, Wang C, Dong L, Zhong Z, et al. Cytoskeleton-linking membrane protein 63 as a serum biomarker for the early detection of hepatocellular carcinoma. *International Journal of Clinical and Experimental Pathology* 2016;**9**(2):1720-6.

Li 2018 *{published data only}*

Li S, Li X, Xu A, Zhang B, He X, Chen H, et al. Screening and clinical evaluation of dominant peptides of centromere protein F antigen for early diagnosis of hepatocellular carcinoma. *Molecular Medicine Reports* 2018;**17**(3):4720-8.

Li 2019b *{published data only}*

Li J, Qiyu S, Wang T, Jin B, Li N. Improving the detection of hepatocellular carcinoma using serum AFP expression in combination with GPC3 and micro-RNA miR-122 expression. *Open Life Science* 2019;**14**:53-61.

Li 2019c *{published data only}*

Li Y, Chen J. Serum des-gamma-carboxy prothrombin for diagnosis of adult primary cancer in liver. *Journal of the College of Physicians and Surgeons--Pakistan : JCPSP* 2019;**29**(10):972-6.

Liaw 1986 *{published data only}*

Liaw YF, Tai DI, Chu CM, Lin DY, Sheen IS, Chen TJ, et al. Early detection of hepatocellular carcinoma in patients with chronic type B hepatitis. A prospective study. *Gastroenterology* 1986;**90**(2):263-7.

Lim 2006 *{published data only}*

Lim JH, Kim SH, Lee WJ, Choi D, Kim SH, Lim HK. Ultrasonographic detection of hepatocellular carcinoma: correlation of preoperative ultrasonography and resected liver pathology. *Clinical Radiology* 2006;**61**:191-7.

Liu 2003 *{published data only}*

Liu WC, Lim JH, Park CK, Kim MJ, Kim SH, Lee SJ, et al. Poor sensitivity of sonography in detection of hepatocellular carcinoma in advanced liver cirrhosis: accuracy of pretransplantation sonography in 118 patients. *European Radiology* 2003;**13**:1693-8.

Liu 2008 *{published data only}*

Liu YH, Li YX, Shen HH. Diagnostic value of simultaneous measurement of AFP, GGT II, MM9 in primary hepatic carcinoma. *Journal of Jilin University (Medical Edition)* 2008;**34**(5):875-8.

Liu 2010b *{published data only}*

Liu Y, He J, Li C, Benitez R, Fu S, Marrero J, et al. Identification and confirmation of biomarkers using an integrated platform for quantitative analysis of glycoproteins and their glycosylations. *Journal of Proteome Research* 2010;**9**(2):798-805.

Liu 2013 *{published data only}*

Liu B, Yang Y, Han S, Zhang L, Bai Y, Fang X. Plasma CD24 as a novel useful biomarker in differentiating hepatocellular carcinoma patients from normal individuals. *Hepato-gastroenterology* 2013;**60**:1245-50.

Liu 2014 *{published data only}*

Liu J, Zhou G, Lu W. Plasma interleukin 17 in the diagnosis of hepatocellular carcinoma: a retrospective study of 39 cases. *Journal of Cancer Research* 2014;**10**(Suppl):304-6.

Liu 2017a *{published data only}*

Liu J, Hu HH, Lee MH, Korenaga M, Jen CL, Batrla-Utermann R, et al. Serum levels of M2BPGi as short-term predictors of hepatocellular carcinoma in untreated chronic hepatitis B patients. *Scientific Reports* 2017;**7**(1):14352.

Liu 2017b *{published data only}*

Liu T, Liu D, Liu R, Jiang H, Yan G, Li W, et al. Discovering potential serological biomarker for chronic HepatitisB Virus-related hepatocellular carcinoma in Chinese population by MAL-associated serum glycoproteomics analysis. *Scientific Reports* 2017;**7**:38918.

Lu 2008 *{published data only}*

Lu SN, Wang JH, Chen PF, Tung HD, Tseng PL, Hung CH, et al. Community-based mass ultrasonographic screening of hepatocellular carcinoma among thrombocytopenic adults. *Cancer Epidemiology, Biomarkers & Prevention* 2008;**17**(7):1813-21.

Luning 1991 *{published data only}*

Luning M, Koch M, Abet L, Wolff II, Wenig B, Buchalt K, et al. The accuracy of the imaging procedures (sonography, MRT, CT, angio-CT, nuclear medicine) in characterizing liver tumors [Treffsicherheit bildgebender Verfahren (Sonographie, MRT, CT, Angio-CT, Nuklearmedizin) bei der Charakterisierung von Lebertumoren]. *Fortschritte Auf Dem Gebiete der Rontgenstrahlen* 1991;**154**(4):398-406.

Lv 2013 *{published data only}*

Lv Y, Wang W, Jia WD, Sun QK, Huang M, Zhou HC, et al. High preoperative levels of serum periostin are associated with poor prognosis in patients with hepatocellular carcinoma after hepatectomy. *European Journal of Surgical Oncology* 2013;**39**(10):1129-35.

Mac Kinnon 1985 *{published data only}*

Mac Kinnon JD, Llana PO. Neoplastic liver: scintigraphy or ecotomography. Experience with 58 patients with the clinical suspicion of liver malignancy. *Revista Medica Chile* 1985;**113**:661-4.

Maeda 2019 *{published data only}*

Maeda T, Kanzaki H, Chiba T, Ao J, Kanayama K, Maruta S, et al. Serum fibroblast growth factor 19 serves asa potential novel biomarker for hepatocellular carcinoma. *BMC Cancer* 2019;**19**:1088.

Mao 2010 *{published data only}*

Mao Y, Yang H, Xu H, Lu X, Sang X, Du S, et al. Golgi protein 73 (GOLPH2) is a valuable serum marker for hepatocellular carcinoma. *Gut* 2010;**59**(12):1687-93.

Matboli 2018 *{published data only}*

Matbol M, E Shafei A, Ali MA, Ashry AM, Kamal KM, Agag MA, et al. circRNAs (hsa_circ_00156, hsa_circ _000224, and hsa_circ_000520) are novel potential biomarkers in hepatocellular carcinoma. *Journal of Cellular Biochemistry* 2018;**120**:7711-24.

Matboli 2020 *{published data only}*

Matboli M, Labib ME, Nasser HE, El-Tawdi AH, Habib EK, Ali-Labib R. Exosomal miR-1298 and lncRNA-RP11-583F2.2 expression in hepatocellular carcinoma. *Current Genomics* 2020;**21**:46-55.

Matsubara 2013 *{published data only}*

Matsubara T, Kanto T, Kuroda S, Yoshio S, Higashitani K, Kakita N, et al. TIE2-expressing monocytes as a diagnostic marker for hepatocellular carcinoma correlates with angiogenesis. *Hepatology (Baltimore, Md.)* 2013;**57**(4):1416-25.

Maussier 1990 *{published data only}*

Maussier ML, Valenza V, Schinco G, Galli G. AFP, CEA, CA 19-9 and TPA in hepatocellular carcinoma. *International Journal of Biological Markers* 1990;**5**(3):121-6.

Mclntire 1972 *{published data only}*

Mclntire KR, Vogel CL, Princler GL, Patel IA. Serum a-Fetoprotein as a Biochemical Marker for Hepatocellular carcinoma. *Cancer Research* 1972;**32**:1941-6.

Mebazaa 1985 *{published data only}*

Mebazaa A, Ben Naceur M, Rachdi MT, Garoui H, Gargouri M. Diagnostic significance of alpha-fetoproteins in hepatic pathology [Valeur diagnostique du dosage de l'alpha-foeto-protéine en pathologie hépatique]. *La Tunisie Medical* 1985;**63**(3):237-40.

Mehta 2018 *{published data only}*

Mehta AS, Lau DT, Wang M, Aslam A, Nasir B, Javaid A, et al. Application of the Doylestown algorithm for the early detection of hepatocellular carcinoma. *PLOS One* 2018;**13**(8):e0203149.

Melia 1983 *{published data only}*

Melia WM, Bullock S Johnson PJ, Williams R. Serum ferritin in hepatocellular carcinoma: a comparison with alphafetoprotein. *Cancer* 1983;**51**:2112-5.

Merchante 2019 *{published data only}*

Merchante N, Figueruela B, Rodríguez-Fernández M, Rodríguez-Arrondo F, Revollo B, Ibarra S, et al. Low performance of ultrasound surveillance for the diagnosis of hepatocellular carcinoma in HIV-infected patients. *AIDS* 2019;**33**(2):269-78.

Mita 1998 *{published data only}*

Mita Y, Aoyagi Y, Yanagi M, Suda T, Suzuki Y, Asakura H. The usefulness of determining des-gamma-carboxy prothrombin by sensitive enzyme immunoassay in the early diagnosis of patients with hepatocellular carcinoma. *Cancer* 1998;**82**(9):1643-8.

Morimoto 2002 *{published data only}*

Morimoto Y, Kubo S, Shuto T, Tanaka H, Hirohashi K, Yamamoto T, et al. Power Doppler ultrasonographic diagnosis of small hepatocellular carcinomas. *Digestive Surgery* 2002;**19**(5):379-87; discussion 387-8.

Morimoto 2012 *{published data only}*

Morimoto M, Numata K, Nozaki A, Kondo M, Moriya S, Taguri M, et al. Novel Lens culinaris agglutinin-reactive fraction of α-fetoprotein: a biomarker of hepatocellular carcinoma recurrence in patients with low α-fetoprotein concentrations. *International Journal of Clinical Oncology* 2012;**17**(4):373-9.

Morota 2011 *{published data only}*

Morota K, Nakagawa M, Sekiya R, Hemken PM, Sokoll LJ, Elliott D, et al. A comparative evaluation of Golgi protein-73, fucosylated hemopexin, α-fetoprotein, and PIVKA-II in the serum of patients with chronic hepatitis, cirrhosis, and hepatocellular carcinoma. *Clinical Chemistry and Laboratory Medicine* 2011;**49**(4):711-8.

Nayak 1988 *{published data only}*

Nayak SS, Kamath SS, Kundaje GN, Aroor AR. Diagnostic significance of estimation of serum apolipoprotein A along with a-fetoproteinin alcoholic cirrhosis and hepatocellular carcinoma patients. *Clinica Chimica Acta* 1988;**173**:157-64.

Oka 1990 *{published data only}*

Oka H, Kurioka N, Kim K, Kanno T, Kuroki T, Mizoguchi Y, et al. Prospective study of early detection of hepatocellular carcinoma in patients with cirrhosis. *Hepatology (Baltimore, Md.)* 1990;**12**(4):680-7.

Ola 2006 *{published data only}*

Ola OS, Ogunbiyi OJ, Olaleye OD, Ayoola EA. Tumour markers and HCV infection in Nigerian patients with liver disease. *Nigerian Journal of Medicine* 2006;**15**:417-20.

Pan 2019 *{published data only}*

Pan Y, Dai J, Hua X, Chen J, Chen Y, Liao Y. The value of combined detection of PIVKA-II, AFP and AFP-L3 in the diagnosis of hepatocellular carcinoma. *Acta Medica Mediterranea* 2019;**35**:2793-7.

Peterson 2000 *{published data only}*

Peterson MS, Baron RL, Marsh JW, Oliver JH III, Confer SR, Hunt LE. Pretransplantation surveillance for possible hepatocellular carcinoma inpatients with cirrhosis:epidemiology and CT-based tumor detection rate in 430 cases with surgical pathologic correlation. *Radiology* 2000;**217**:743-9.

Pocha 2013 *{published data only}*

Pocha C, Dieperink E, McMaken KA, Knott A, Thuras P, Ho SB. Surveillance for hepatocellular cancer with ultrasonography vs. computed tomography -- a randomised study. *Alimentary Pharmacology & Therapeutics* 2013;**38**(3):303-12.

Pratedrat 2020 *{published data only}*

Pratedrat P, Chuaypen N, Nimsamer P, Payungporn S, Pinjaroen N, Sirichindakul B, et al. Diagnostic and prognostic roles of circulating miRNA-223-3p in hepatitis B virus–related hepatocellular carcinoma. *PLOS One* 2020;**154**(4):e0232211.

Qiao 2011 *{published data only}*

Qiao SS, Cui ZQ, Gong L, Han H, Chen PC, Guo LM, et al. Simultaneous measurements of serum AFP, GPC-3 and HCCR for diagnosing hepatocellular carcinoma. *Hepato-gastroenterology* 2011;**58**(110-111):1718-24.

Qu 2011 *{published data only}*

Qu KZ, Zhang K, Ma W, Li H, Wang X, Zhang X, et al. Ubiquitin-proteasome profiling for enhanced detection of hepatocellular carcinoma in patients with chronic liver disease. *Journal of Gastroenterology and Hepatology* 2011;**26**(4):751-8.

Quaia 2004 *{published data only}*

Quaia E, Calliada F, Bertolotto M, Rossi S, Garioni L, Rosa L, et al. Characterization of focal liver lesions with contrast specific US modes and a sulfur hexafluoride–filled microbubble contrast agent:diagnostic performance and confidence. *Radiology* 2004;**232**:420-30.

Rao 2003 *{published data only}*

Rao AR, Chui AK, Shi LW, Waugh R, Pillay P, Sheil AG. Sensitivity of radiological investigations in diagnosing hepatocellular carcinoma in cirrhotic livers. *Transplantation Proceedings* 2003;**35**(1):348-9.

Rickes 2003 *{published data only}*

Rickes S, Schulze S, Neye H, Ocran KW, Wermke W. Improved diagnosing of small hepatocellular carcinomas by echo-enhanced power Doppler sonography in patients with cirrhosis. *European Journal of Gastroenterology & Hepatology* 2003;**15**:893-900.

Rizzi 1994 *{published data only}*

Rizzi PM, Kane PA, Ryder SD, Ramage JK, Gane E, Tan KC, et al. Accuracy of radiology in detection of hepatocellular carcinoma before liver transplantation. *Gastroenterology* 1994;**107**(5):1425-9.

Rode 2001 *{published data only}*

Rode A, Bancel B, Douek P, Chevallier M, Vilgrain V, Picaud G, et al. Small nodule detection in cirrhotic livers: evaluation with US, spiral CT, and MRI and correlation with pathologic examination of explanted liver. *Journal of Computer Assisted Tomography* 2001;**25**(3):327-36.

Saber 2017 *{published data only}*

Saber MA, AbdelHafiz SM, Khorshed FE, Aboushousha TS, Hamdy HE, Seleem MI, et al. differential expression of glypican-3 and insulin–like growth factor-ii mRNAs and alpha-fetoprotein and KI-67 markers in HCV related hepatocellular carcinomas in Egyptian patients. *Asian Pacific Journal of Cancer Prevention* 2017;**18**(1):121-7.

Sakai 1991 *{published data only}*

Sakai T, Kishikawa Y. Follow-up studies on hepatoma developing from liver cirrhosis. *Nippon Naika Gakkai Zasshi* 1991;**80**(10):1637-44.

Salmi 1988 *{published data only}*

Salmi A, Rangoni G, Martinazzoli F, Salmi R. A clinical study on primary liver cancer in cirrhosis with ascites [Studio clinico del carcinoma epatico primitivo su cirrosi con ascite]. *Minerva Dietologica e Gastroenterologica* 1988;**34**:163-6.

Sangiovanni 2004 *{published data only}*

Sangiovanni A, Del Ninno E, Fasani P, De Fazio C, Ronchi G, Romeo R, et al. Increased survival of cirrhotic patients with a hepatocellular carcinoma detected during surveillance. *Gastroenterology* 2004;**126**(4):1005-14.

Santagostino 2003 *{published data only}*

Santagostino E, Colombo M, Rivi M, Rumi MG, Rocino A, Linari S, et al, Study Group of the Association of Italian Hemophilia Centers. A 6-month versus a 12-month surveillance for hepatocellular carcinoma in 559 hemophiliacs infected with the hepatitis C virus. *Blood* 2003;**102**(1):78-82.

Sato 2009 *{published data only}*

Sato T, Tateishi R, Yoshida H, Ohki T, Masuzaki R, Imamura J, et al. Ultrasound surveillance for early detection of hepatocellular carcinoma among patients with chronic hepatitis C. *Hepatology International* 2009;**3**:544-50.

Seif 2019 *{published data only}*

Seif AA, Aly HH, Elzoghby DM, Elbreedy AM, Lotfy M. Aberrant p16 methylation as an early diagnostic marker in blood of hepatocellular carcinoma patients. *Egyptian Journal of Medical Human Genetics* 2019;**20**:27.

Sekoguchi 1994 *{published data only}*

Sekoguchi B, Horiguchi Y, Imai H, Suzuki T, Miyoshi A, Itoh H, et al. Integrated diagnosis of small hepatocellular carcinoma with imaging diagnosis. *Japanese Journal of Clinical Pathology* 1994;**42**:1036-42.

Shalably 2019 *{published data only}*

Shalably NM, Badawi R, Hawash N, Abd-Elsalam S, Elkhalawany W, EIHameed AA. Evaluation of fucosylated haptoglobin as a diagnostic biomarker for hepatocellular carcinoma in Egypt. *Open Biomarkers Journal* 2019;**9**:31-7.

Shao 2015 *{published data only}*

Shao H, Cheng W, Yu L, Li Y, Hui J. User of F-FDG PET scan and ultrasound guided biopsy in the diagnosis of hepatic carcinoma. *Hepato-Gastroenterology* 2015;**62**:978-81.

Shapiro 1996 *{published data only}*

Shapiro RS, Katz R, Mendelson DS, Halton KP, Schwartz ME, Miller CM. Detection of hepatocellular carcinoma in cirrhotic patients: sensitivity of CT and ultrasonography. *Journal of Ultrasound in Medicine* 1996;**15**(7):497-502; quiz 503-4.

Shehab-Eldeen 2019 *{published data only}*

Shehab-Eldeen S, Nada A, Abou-Elela D, El-Naidany S, Arafat E, Omar T. Diagnostic performance of microRNA-122 and microRNA-224 in hepatitis C virus-Induced hepatocellular carcinoma (HCC). *Asian Pacific Journal of Cancer Prevention* 2019;**20**(8):2515-22.

Sherman 2017 *{published data only}*

Sherman M, Feld J, Yamada H, Mori Y, Janssen H. A randomized controlled trial of US vs US + biomarkers for the diagnosis of hepatocellular carcinoma: an interim report. *Journal of Hepatology* 2017;**66**:S1-32.

Sheu 1985 *{published data only}*

Sheu JC, Sung JL, Chen DS, Lai MY, Wang TH, Yu JY, et al. Early detection of hepatocellular carcinoma by real-time ultrasonography. A prospective study. *Cancer* 1985;**56**:660-6.

Shinagawa 1984 *{published data only}*

Shinagawa T, Ohto M, Kimura K, Tsunetomi S, Morita M, Saisho H, et al. Diagnosis and clinical features of small hepatocellular carcinoma with emphasis on the utility of real-time ultrasonography. A study in 51 patients. *Gastroenterology* 1984;**86**(3):495-502.

Singal 2017 *{published data only}*

Singal AG, Mittal S, Yerokun OA, Ahn C, Marrero JA, Yopp AC, et al. Hepatocellular carcinoma screening associated with early tumor detection and improved survival among patients with cirrhosis in the US. *American Journal of Medicine* 2017;**130**(9):1099-106.

Tameda 2013 *{published data only}*

Tameda M, Shiraki K, Sugimoto K, Ogura S, Inagaki Y, Yamamoto N, et al. Des-γ-carboxy prothrombin ratio measured by P-11 and P-16 antibodies is a novel biomarker for hepatocellular carcinoma. *Cancer Science* 2013;**104**(6):725-31.

Thakur 2014 *{published data only}*

Thakur S, Jhobta A, Dhiman DS, Sood RG, Chauhan A, Thakur CS. Role of contrast enhanced ultrasound in characterization of focal liver lesions. *Egyptian Journal of Radiology and Nuclear Medicine* 2014;**45**:7-17.

Toyoda 2011 *{published data only}*

Toyoda H, Kumada T, Tada T, Kaneoka Y, Maeda A, Kanke F, et al. Clinical utility of highly sensitive Lens culinaris agglutinin-reactive alpha-fetoprotein in hepatocellular carcinoma patients with alpha-fetoprotein <20 ng/mL. *Cancer Science* 2011;**102**(5):1025-31.

Tradati 1998 *{published data only}*

Tradati F, Colombo M, Mannucci PM, Rumi MG, De Fazio C, Gamba G, et al. A prospective multicenter study of hepatocellular carcinoma in Italian hemophiliacs with chronic hepatitis C. The Study Group of the Association of Italian Hemophilia Centers. *Blood* 1998;**91**(4):1173-7.

Tung 2012 *{published data only}*

Tung EK, Ng IO. Significance of serum DKK1 as a diagnostic biomarker in hepatocellular carcinoma. *Future Oncology (London, England)* 2012;**8**(12):1525-8.

Ueda 1995 *{published data only}*

Ueda K, Kitagawa K, Kadoya M, Matsui O, Takashima T, Yamahana T. Detection of hypervascular hepatocellular carcinoma by using spiral volumetric CT: comparison of US and MR imaging. *Abdominal Imaging* 1995;**20**(6):547-53.

Uenishi 2006 *{published data only}*

Uenishi T, Yamazaki O, Yamamoto T, Hirohashi K, Tanaka H, Tanaka S, et al. Clinical significance of serum cytokeratin-19 fragment (CYFRA 21-1) in hepatocellular carcinoma. *Journal of Hepato-biliary-pancreatic Surgery* 2006;**13**(3):239-44.

Wang 2020 *{published data only}*

Wang S, Yang Y, Sun L, Qiao G, Song Y, Liu B. Exosomal microRNAs as liquid biopsy biomarkers in hepatocellular carcinoma. *OncoTargets and Therapy* 2020;**13**:2021-30.

Wei 2012 *{published data only}*

Wei S, Suryani Y, Gowda GA, Skill N, Maluccio M, Raftery D. Differentiating hepatocellular carcinoma from hepatitis C using metabolite profiling. *Metabolites* 2012;**2**(4):701-16.

Welberry 2020 *{published data only}*

Welberry C, Macdonald I, McElveen J, Parsy-Kowalska C, Allen J, Healey G. Tumor-associated autoantibodies in combination with alpha-fetoprotein for detection of early stage hepatocellular carcinoma. *PLOS One* 2020;**15**(5):e0232247.

Worland 2018 *{published data only}*

Worland T, Harrison B, Delmenico L, Dowling D. Hepatocellular carcinoma screening utilising serum alpha-fetoprotein measurement and abdominal ultrasound is more effective than ultrasound alone in patients with non-viral cirrhosis. *Journal of Gastrointestinal Cancer* 2018;**49**(4):476-80.

Xiao 2019 *{published data only}*

Xiao J, Long F, Peng T, Hu LB, Cai H, Chen R, et al. Development and potential application of a simultaneous multiplex assay of Golgi protein 73 and alpha-fetoprotein for hepatocellular carcinoma diagnosis. *European Review for Medical and Pharmacological Sciences* 2019;**238**(8):3302-10.

Xu 1990 *{published data only}*

Xu KC, Shi YC, Meng XY, Wei Q. Reappraisal of diagnostic significance of a hepatoma-specific band of serum gamma-glutamyl transferase. *Chinese Medical Journal* 1990;**103**(3):228-32.

Xu 2012 *{published data only}*

Xu HX, Lu MD, Liu LN, Zhang YF, Guo LH, Xu JM, et al. Discrimination between neoplastic and non-neoplastic lesions in cirrhotic liver using contrast-enhanced ultrasound. *British Journal of Radiology* 2012;**85**(1018):1376-84.

Xu 2014 *{published data only}*

Xu JB, Qi FZ, Xu G, Chen GF, Qin LX, Zhang JH. Value of alpha-fetoprotein and clinical characteristics in patients with liver neoplasm. *Neoplasma* 2014;**61**(2):218-24.

Xu 2019 *{published data only}*

Xu R, Wang J, Huang X, Zhang Q, Xie Y, Pang L, et al. Clinical value of spectral CT imaging combined with AFP in identifying liver cancer and hepatic focal nodular hyperplasia. *Official journal of the Balkan Union of Oncology* 2019;**24**(4):1429-34.

Yamamoto 2009 *{published data only}*

Yamamoto K, Imamura H, Matsuyama Y, Hasegawa K, Beck Y, Sugawara Y, et al. Significance of alpha-fetoprotein and des-gamma-carboxy prothrombin in patients with hepatocellular carcinoma undergoing hepatectomy. *Annals of Surgical Oncology* 2009;**16**(10):2795-804.

Yamamoto 2010 *{published data only}*

Yamamoto K, Imamura H, Matsuyama Y, Kume Y, Ikeda H, Norman GL, et al. AFP, AFP-L3, DCP, and GP73 as markers for monitoring treatment response and recurrence and as surrogate markers of clinicopathological variables of HCC. *Journal of Gastroenterology* 2010;**45**(12):1272-82.

Yamashiki 2011 *{published data only}*

Yamashiki N, Sugawara Y, Tamura S, Kaneko J, Yoshida H, Aoki T, et al. Diagnostic accuracy of α-fetoprotein and des-γ-carboxy prothrombin in screening for hepatocellular carcinoma in liver transplant candidates. *Hepatology Research* 2011;**41**(12):1199-207.

Yamashita 2020 *{published data only}*

Yamashita S, Kato A, Akatsuka T, Sawada T, Asai T, Koyama N, et al. Clinical relevance of increased serum preneoplastic antigen in hepatitis C-related hepatocellular carcinoma. *World Journal of Gastroenterology* 2020;**26**(13):1463-73.

Yang 2013b *{published data only}*

Yang H, Chen GD, Fang F, Liu Z, Lau SH, Zhang JF, et al. Dickkopf-1: as a diagnostic and prognostic serum marker for early hepatocellular carcinoma. *International Journal of Biological Markers* 2013;**28**(3):286-97.

Yang 2019 *{published data only}*

Yang T, Xing H, Wang G, Wang N, Liu M, Yan C, et al. A novel online calculator based on serum biomarkers to detect hepatocellular carcinoma among patients with hepatitis B. *Clinical Chemistry* 2019;**65**(12):1543–53.

Yao 2013 *{published data only}*

Yao M, Yao DF, Bian YZ, Wu W, Yan XD, Yu DD, et al. Values of circulating GPC-3 mRNA and alpha-fetoprotein in detecting patients with hepatocellular carcinoma. *Hepatobiliary & Pancreatic Diseases International* 2013;**12**(2):171-9.

Yasmin Anum 2009 *{published data only}*

Yasmin Anum MY, Looi ML, Nor Aini AH, Merican I, Wahidah A, Mohd Radzi AH, et al. Combined assessment of TGF-beta-1 and alpha-fetoprotein values improves specificity in the diagnosis of hepatocellular carcinoma and other chronic liver diseases in Malaysia. *Medical Journal of Malaysia* 2009;**64**(3):223-7.

Yasuda 2010 *{published data only}*

Yasuda E, Kumada T, Toyoda H, Kaneoka Y, Maeda A, Okuda S, et al. Evaluation for clinical utility of GPC3, measured by a commercially available ELISA kit with Glypican-3 (GPC3) antibody, as a serological and histological marker for hepatocellular carcinoma. *Hepatology Research* 2010;**40**(5):477-85.

Younis 2019 *{published data only}*

Younis YS, Alegaily HS, Elagawy W, Semeya AA, Esam-Eldin Abo-Amer Y, El-Abgeegy M, et al. Serum Dickopff 1 as a novel biomarker in hepatocellular carcinoma diagnosis and follow up after ablative therapy. *Cancer Management and Research* 2019;**11**:10555-62.

Yvamoto 2015 *{published data only}*

Yvamoto EY, Ferreira RF, Nogueira V, Pinhe MA, Tenani GD, Andrade JG, et al. Influence of vascular endothelial growth factor and alpha-fetoprotein on hepatocellular carcinoma. *Genetics and Molecular Research* 2015;**14**(4):17453-62.

Zhang 1999 *{published data only}*

Zhang B, Yang B. Combined alpha fetoprotein testing and ultrasonography as a screening test for primary liver cancer. *Journal of Medical Screening* 1999;**6**(2):108-10.

Zhang 2010 *{published data only}*

Zhang Q, Xiao Q, Lin Z, Ying X, Li Z, Lin JM. Development of a competitive radioimmunoassay for glypican-3 and the clinical application in diagnosis of hepatocellular carcinoma. *Clinical Biochemistry* 2010;**43**(12):1003-8.

Zhao 2011 *{published data only}*

Zhao YM, Wang L, Dai Z, Wang DD, Hei ZY, Zhang N, et al. Validity of plasma macrophage migration inhibitory factor for diagnosis and prognosis of hepatocellular carcinoma. *International Journal of Cancer. Journal International du Cancer* 2011;**129**(10):2463-72.

Zheng 2014 *{published data only}*

Zheng T, Chen M, Han S, Zhang L, Bai Y, Fang X, et al. Plasma minichromosome maintenance complex component 6 is a novel biomarker for hepatocellular carcinoma patients. *Hepatology Research* 2014;**44**(13):1347-56.

Zheng 2018 *{published data only}*

Zheng W, Yang J, Dong Z, Wang L, Fang M, Wu W, et al. High mobility group box 3 as an emerging biomarker in diagnosis and prognosis of hepatocellular carcinoma. *Cancer Management and Research* 2018;**10**:5979-89.

Zheng 2019 *{published data only}*

Zheng W, Yao M, Fang M, Pan L, Wang L, Yang J, et al. Oncogenic Wnt3a: a candidate specific marker and novel molecular target for hepatocellular carcinoma. *Journal of Cancer* 2019;**10**(23):5862-73.

Additional references

Arif-Tiwari 2014

Arif-Tiwari H, Kalb B, Chundru S, Sharma P, Costello J, Guessner RW, et al. MRI of hepatocellular carcinoma: an update of current practices. *Diagnostic and Interventional Radiology* 2014;**20**:209-21.

Balshem 2011

Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 2011;**64**(4):401-6.

Bourliere 2015

Bourliere M, Bronowicki JP, de Ledinghen V, Hezode C, Zoulim F, Mathurin P, et al. Ledipasvir-sofosbuvir with or without ribavirin to treat patients with HCV genotype 1 infection and cirrhosis non-responsive to previous protease-inhibitor therapy: a randomised, double-blind, phase 2 trial (SIRIUS). *Lancet Infectious Diseases* 2015;**15**:397–404.

Bralet 2000

Bralet MP, Regimbeau JM, Pineau P, Dubois S, Loas G, Degos F, et al. Hepatocellular carcinoma occurring in nonfibrotic liver: epidemiologic and histopathologic analysis of 80 French cases. *Hepatology (Baltimore, Md.)* 2000;**32**(2):200-4.

Bray 2018

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians* 2018;**68**(6):394-424.

Bruix 2011

Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology (Baltimore, Md.)* 2011;**53**(3):1020.

Butt 2018

Butt AS, Sharif F, Abid S. Impact of direct acting antivirals on occurrence and recurrence of hepatocellular carcinoma: biologically plausible or an epiphenomenon? *World Journal of Hepatology* 2018;**10**(2):267-76.

Calvaruso 2018

Calvaruso V, Cabibbo G, Cacciola I, Petta S, Madonia S, Bellia A, et al. Incidence of hepatocellular carcinoma in patients with HCV-associated cirrhosis treated with direct-acting antiviral agents. *Gastroenterology* 2018;**155**(2):411-21.e4.

Charlton 2015

Charlton M, Everson GT, Flamm SL, Kumar P, Landis C, Brown RS Jr, et al. Ledipasvir and sofosbuvir plus ribavirin for treatment of HCV infection in patients with advanced liver disease. *Gastroenterology* 2015;**149**:649–59.

Choi 2014

Choi JY, Lee JM, Sirlin CB. CT and MR imaging diagnosis and staging of hepatocellular carcinoma: part I. Development, growth, and spread: key pathologic and imaging aspects. *Radiology* 2014;**272**:635-54.

Chou 2015

Chou R, Cuevas C, Fu R, Devine B, Wasson N, Ginsburg A, et al. Imaging techniques for the diagnosis of hepatocellular carcinoma: a systematic review and meta-analysis. *Annals of Internal Medicine* 2015;**162**:697-711.

Chung 2015

Chung YE, Kim KW. Contrast-enhanced ultrasonography: advance and current status in abdominal imaging. *Ultrasonography* 2015;**34**(1):3-18.

Colli 2006

Colli A, Fraquelli M, Casazza G, Massironi S, Colucci A, Conte D, et al. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. *American Journal of Gastroenterology* 2006;**101**:513-23.

Colli 2014

Colli A, Fraquelli M, Casazza G, Conte D, Nikolova D, Duca P, et al. The architecture of diagnostic research: from bench to bedside--research guidelines using liver stiffness as an example. *Hepatology* 2014;**60**(1):408-18.

Davila 2012

Davila JA, Duan Z, McGlynn KA, El-Serag HB. Utilization and outcomes of palliative therapy for hepatocellular carcinoma: a population-based study in the United States. *Journal of Clinical Gastroenterology* 2012;**46**:71-7.

de Franchis 2015

de Franchis R. Expanding consensus in portal hypertension. Report of the Baveno VI Consensus Workshop: stratifying risk and individualizing care for portal hypertension. *Journal of Hepatology* 2015;**63**:743–52.

Di Bisceglie 2005

Di Bisceglie AM, Sterling RK, Chung RT, Everhart JE, Dienstag JL, Bonkovsky HL, et al. Serum alpha-fetoprotein levels inpatients with advanced hepatitis C: results from the HALT-C Trial. *Journal of Hepatology* 2005;**43**:434–41.

DTA Handbook 2013

Deeks JJ, Bossuyt PM, Gatsonis C, editor(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2013. Available from srdta.cochrane.org.

EASL 2018

European Association for the Study of the Liver. EASL Clinical Practice Guidelines: managementof hepatocellular carcinoma. *Journal of Hepatology* 2018;**69**:182-236.

EASL-EORTC 2012

European Association For The Study Of The Liver, European Organisation For Research And Treatment Of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *Journal of Hepatology* 2012;**56**(4):908-43.

Forner 2012

Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet* 2012;**379**(9822):1245-55.

Forner 2018

Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* 2018;**391**:1301-14.

Fraquelli 2019

Fraquelli M, Nadarevic T, Giljaca V, Colli A, Miletic D, Štimac D, et al. Contrast-enhanced ultrasound for the diagnosis of hepatocellular carcinoma in advanced chronic liver disease. *Cochrane Database of Systematic Reviews* 2019, Issue 11. Art. No: CD013483. [DOI: 10.1002/14651858.CD013483]

GBD 2017

Global Burden of Disease Liver Cancer Collaboration. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015. *JAMA Oncology* 2017;**3**(12):1683-91.

GRADEpro GDT [Computer program]

McMaster University (developed by Evidence Prime) GRADEpro GDT. Version accessed 5 November 2020. Hamilton (ON): McMaster University (developed by Evidence Prime), 2015. Available at gradepro.org.

Heimbach 2018

Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 2018;**67**:358-80.

Hennedige 2012

Hennedige T, Venkatesh SK. Imaging of hepatocellular carcinoma: diagnosis, staging and treatment monitoring. *Cancer Imaging* 2012;**12**(3):530-47.

Hussain 2002

Hussain SM, Zondervan PE, IJzermans JN, Schalm SW, de Man RA, Krestin GP. Benign versus malignant hepatic nodules: MR imaging findings with pathologic correlation. *Radiographics* 2002;**22**:1023–39.

Imamura 1999

Imamura H, Matsuyama Y, Miyagawa Y, Ishida K, Shimada R, Miyagawa S, et al. Prognostic significance of anatomical resection and des-gamma-carboxy prothrombin in patients with hepatocellular carcinoma. *British Journal of Surgery* 1999;**89**:1032–8.

Jakobsen 2017

Jakobsen JC, Nielsen EE, Feinberg J, Katakam KK, Fobian K, Hauser G, et al. Direct-acting antivirals for chronic hepatitis C. *Cochrane Database of Systematic Reviews* 2017, Issue 9. Art. No: CD012143. [DOI: 10.1002/14651858.CD012143.pub3]

Kansagara 2014

Kansagara D, Papak J, Pasha AS, O'Neil M, Freeman M, Relevo R, et al. Screening for hepatocellular carcinoma in chronic liver disease. *Annals of Internal Medicine* 2014;**161**:261-9.

Kew 1975

Kew MC. Alpha-fetoprotein. In: Read AE, editors(s). Modern Trends in Gastroenterology. Vol. **5**. London: Butterworths, 1975:91.

Kinoshita 2015

Kinoshita A, Onoda H, Fushiya N, Koike K, Nishino H, Tajiri H. Staging systems for hepatocellular carcinoma: current status and future perspectives. *World Journal of Hepatology* 2015;**7**(3):406–24.

Koike 2001

Koike Y, Shiratori Y, Sato S, Obi S, Teratani T, Imamura M, et al. Des-gamma-carboxy prothrombin as a useful predisposing factor for the development of portal venous invasion in patients with hepatocellular carcinoma: a prospective analysis of 227 patients. *Cancer* 2001;**91**:561–9.

Lee 2012

Lee JM, Yoon JH, Kim KW. Diagnosis of hepatocellular carcinoma: newer radiological tools. *Seminars in Oncology* 2012;**39**:399-409.

Leroy 2016

Leroy V, Angus P, Bronowicki JP, Dore GJ, Hezode C, Pianko S, et al. Daclatasvir, sofosbuvir, and ribavirin for hepatitis C virus genotype 3 and advanced liver disease: a randomized phase III study (ALLY-3 +). *Hepatology (Baltimore, Md.)* 2016;**63**:1430–41.

Li 2014

Li C, Zhang Z, Zhang P, Liu J. Diagnostic accuracy of desgamma-carboxy prothrombin versus alpha-fetoprotein for hepatocellular carcinoma: a systematic review. *Hepatology Research* 2014;**44**:E11–25.

LI-RADS 2018

The American College of Radiology. CT/MRI LI-RADS v2018. www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS/CT-MRI-LI-RADS-v2018 2018 (accessed 1 June 2020).

Llovet 1999

Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Seminars in Liver Disease* 1999;**19**:329-38.

Llovet 2003

Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. *Lancet* 2003;**362**:1907-17.

Llovet 2008

Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. *Journal of the National Cancer Institute* 2008;**100**:698–711.

Lok 2009

Lok AS, Seeff LB, Morgan TR, di Bisceglie AM, Sterling RK, Curto TM, et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease. *Gatroenterology* 2009;**136**(1):138-48.

Mazzaferro 1996

Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *New England Journal of Medicine* 1996;**334**:693-9.

Mazzaferro 2011

Mazzaferro V, Bhoori S, Sposito C, Bongini M, Langer M, Miceli R, et al. Milan criteria in liver transplantation for hepatocellular carcinoma: an evidence based analysis of 15 years of experience. *Liver Transplantation* 2011;**17**:S44-57.

Nadarevic 2019

Nadarevic T, Giljaca V, Colli A, Fraquelli M, Casazza G, Miletic D, et al. Computed tomography for the diagnosis of hepatocellular carcinoma in chronic advanced liver disease. *Cochrane Database of Systematic Reviews* 2019, Issue 6. Art. No: CD013362. [DOI: 10.1002/14651858.CD013362]

Nadarevic 2020

Nadarevic T, Colli A, Giljaca V, Fraquelli M, Casazza G, Manzotti C, et al. Magnetic resonance imaging for the diagnosis of hepatocellular carcinoma in people with chronic liver disease [Protocol]. Cochrane Database of Systematic Reviews 2020 (to be published).

O'Neill 2015

O'Neill EK, Cogley JR, Miller FH. The ins and outs of liver imaging. *Clinics in Liver Disease* 2015;**19**:99-121.

Omata 2017

Omata M, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, et al. Asia–Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatology International* 2017;**11**:317–70.

Park 2017

Park HJ, Choi BI, Lee ES, Park SB, Lee JB. How to differentiate borderline hepatic nodules in hepatocarcinogenesis: emphasis on imaging diagnosis. *Liver Cancer* 2017;**6**:189–203.

Poustchi 2011

Poustchi H, Farrell GC, Strasser SI, Lee AU, McCaughan GW, George J. Feasibility of conducting a randomized control trial for liver cancer screening: is a randomized controlled trial for liver cancer screening feasible or still needed? *Hepatology (Baltimore, Md.)* 2011;**54**:1998-2004.

Pucci 1991

Pucci P, Siciliano R, Malorni A, Marino G, Tecce MF, Ceccarini C, et al. Human alpha-fetoprotein primary structure: a mass spectrometric study. *Biochemistry* 1991;**30**(20):5061-6.

Review Manager 2020 [Computer program]

The Cochrane Collaboration Review Manager (RevMan). Version 5.4. The Cochrane Collaboration, 2020.

Roberts 2018

Roberts LR, Sirlin CB, Zaiem F, Almasri J, Prokop LJ, Heimbach JK, et al. Imaging for the diagnosis of hepatocellular carcinoma: a systematic review and meta-analysis. *Hepatology (Baltimore, Md.)* 2018;**67**(1):401-21.

Rodgers 2019

Rodgers SK, Fetzer DT, Gabriel H, Seow JH, Choi HH, Maturen KE, et al. Role of US LI-RADS in the LI-RADS algorithm. *Radiographics* 2019;**39**:690-708.

Royle 2003

Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *International Journal of Technology Assessment in Health Care* 2003;**19**(4):591–603.

Rutter 2001

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2001;**20**(19):2865-84.

Ryerson 2016

Ryerson AB, Eheman CR, Altekruse SF, Ward JW, Jemal A, Sherman RL, et al. Annual Report to the Nation on the Status of Cancer, 1975-2012, featuring the increasing incidence of liver cancer. *Cancer* 2016;**122**(9):1312-37.

Schuetz 2012

Schuetz GM, Schlattmann P, Dewey M. Use of 3x2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies. *BMJ (Clinical Research Ed.)* 2012;**345**:e6717.

Schünemann 2008

Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ (Clinical Research Ed.)* 2008;**336**(7653):1106-11.

Schünemann 2016

Schünemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G, et al. GRADE Working Group. GRADEGuidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *Journal of Clinical Epidemiology* 2016;**76**:89-98.

Schünemann 2020a

Schünemann HJ, Mustafa R, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 1. Study design, risk of bias and indirectness in rating the certainty across a body of evidence for test accuracy. *Journal of Clinical Epidemiology* 2020;**122**:129-41.

Schünemann 2020b

Schünemann HJ, Mustafa R, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 2. Inconsistency, Imprecision, publication bias and other domains for rating the certainty of evidence for test accuracy and presenting it in evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2020;**122**:142-52.

Shah 2014

Shah S, Shukla A, Paunipagar B. Radiological features of hepatocellular carcinoma. *Journal of Clinical and Experimental Hepatology* 2014;**4**:63-6.

Simmons 2017

Simmons O, Fetzer DT, Yokoo T, Marrero JA, Yopp A, Kono Y, et al. Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis. *Alimentary Pharmacology & Therapeutics* 2017;**45**(1):169-77.

Singal 2009

Singal A, Volk ML, Waljee A, Salgia R, Higgins P, Rogers MA, et al. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Alimentary Pharmacology & Therapeutics* 2009;**30**:37-47.

Singal 2014

Singal AG, Pillai A, Tiro J. Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. *PLOS Medicine* 2014;**11**:e1001624.

Tateishi 2008

Tateishi R, Yoshida H, Matsuyama Y, Mine N, Kondo Y, Omata M. Diagnostic accuracy of tumor markers for hepatocellular carcinoma: a systematic review. *Hepatology International* 2008;**2**:17-30.

Tzartzeva 2018

Tzartzeva K, Obi J, Rich NE, Parikh ND, Marrero JA, Yopp A, et al. Surveillance Imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: a meta-analysis. *Gastroenterology* 2018;**154**:1706-18.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Yang 2011

Yang JD, Harmsen WS, Slettedahl SW, Chaiteerakij R, Enders FT, Therneau TM, et al. Factors that affect the risk for hepatocellular carcinoma and effects of surveillance. *Clinical Gastroenterology and Hepatology* 2011;**9**(7):617-23.

Yokoyama 1990

Yokoyama I, Todo S, Iwatsuki S, Starzl TE. Liver transplantation in the treatment of primary liver cancer. *Hepato-gastroenterology* 1990;**37**(2):188-93.

Young 2012

Young AL, Adair R, Prasad KR, Toogood GJ, Lodge JP. Hepatocellular carcinoma within a noncirrhotic, nonfibrotic, seronegative liver: surgical approaches and outcomes. *Journal of the American College of Surgeons* 2012;**214**(2):174-83.

Zhou 2018

Zhou F, Shang W, Yu X, Tian J. Glypican-3: a promising biomarker for hepatocellular carcinoma diagnosis and treatment. *Medicinal Research Reviews* 2018;**38**(2):741-67.

References to other published versions of this review

Colli 2019

Colli A, Nadarević T, Miletić D, Giljaca V, Fraquelli M, Štimac D, et al. Abdominal ultrasound and alpha-fetoprotein for the diagnosis of hepatocellular carcinoma. *Cochrane Database of Systematic Reviews* 2019, Issue 6. Art. No: CD013346. [DOI: 10.1002/14651858.CD013346]

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Abdel-Aziz 2016

Study characteristics			
Patient Sampling	A total of 86 participants with chronic liver disease were included; 68 with HCC and 18 without HCC (+ 20 healthy adults). Age range: 38-54. Males 85%		
Patient characteristics and setting	Patients with chronic liver disease at tertiary referral centre in Egypt, selected on the presence of HCC		
Index tests	Serum alpha-foetoprotein was measured using Electrochemilus-cence Immunoassay (Roche).		
Target condition and reference standard(s)	HCC occurrences were detected by US, CT, AFP and confirmed by histology; controls: liver cirrhosis without evidence of HCC		
Flow and timing	No information on interval between index test and reference stan-dard		
Comparative			
Notes	Authors declared no conflicts of interest		
Methodological quality			
Item	**Authors' judge-ment**	**Risk of bias**	**Applicability con-cerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or inter-pretation differ from the review question?			Low concern

Abdel-Aziz 2016 *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Abdelghany 2018

Study characteristics	
Patient Sampling	30 participants with HCC on HCV liver cirrhosis and 10 participants with chronic liver disease as control. Consecutuvely enrolled. Quote: "Subjects with malignancies other than HCC, autoimmune diseases, chronic liver diseases other than viral hepatitis, benign liver tumours or secondary (metastatic) liver tumours and BCLC stage C or D disease were excluded from the study." Age range: 30-70. Males 85%
Patient characteristics and setting	Patients with chronic liver disease at tertiary referral centre in Egypt selected on the presence absence of HCC
Index tests	AFP was assayed by electro-chemiluminescence on a Cobas e411 immunoassay autoanalyzer.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was based on non-invasive imaging techniques; either triphasic multidetector CT scan or dynamic contrast-enhanced magnetic resonance imaging, according to AASLD guidelines. For patients with hepatic nodules beyond 1 cm in diameter, one imaging technique was required, while in those patients with smaller lesions, both techniques were performed for confirmation. Pathological diagnosis was performed for the typical HCC criteria.

Abdelghany 2018 *(Continued)*

	Chronic liver disease: clinical evaluation, laboratory tests, and abdominal ultrasound
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Authors declared no conflicts of interest, no funding.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Abdelghany 2018 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Abdel-Hamid 2014

Study characteristics	
Patient Sampling	Group 1: 20 healthy volunteers, sex matched to other groups with the following clinical conditions: apparently healthy, normal clinical examination, abdominal ultrasonography, liver function tests, and seronegative for HCV markers. Group II: 20 patients with HCC regardless the etiology with the following clinical diagnosis: deterioration of health, right hypochondrial pain and hepatomegaly with nodular surface, the abdominal ultrasonography showing hepatic focal lesions (single or multiple) or heterogeneous areas in the liver. Positive histopathological examination of liver biopsy or aspirate for malignancy (only when the patient's clinical condition and prothrombin time and concentration allowed the performance) and/or raised AFP above 400 ng/mL. Group III: 20 patients with chronic HCV, matched in sex to group I with the following clinical conditions: fatigue, anorexia, with high aminotransferase values and hyperbilirubinaemia. They were not under interferon-α2+ribavirin (IF-α2 +RV) treatment. Group IV: 20 patients with chronic HCV, under IF-α2 (weekly subcutaneous single dose, 160 ug/ampoule) plus RV (1200 mg/day, per os doses after meals divided into 3 doses) for one year: this group of patients matches to group I in gender, with the same diagnosis as group III Age range: not reported. Males 82%
Patient characteristics and setting	Patients with chronic liver disease at tertiary referral centre in Egypt selected on the presence absence of HCC
Index tests	The rest of plasma and serum were separated in aliquots and frozen at -70°C for measurement of DCP and AFP. Sandwich principle was employed to determine the AFP concentration via ELISA technique according to manufacturers' instructions (Anogen, Canada)
Target condition and reference standard(s)	
Flow and timing	No information on interval between index test and reference standard
Comparative	

Abdel-Hamid 2014 *(Continued)*

Notes	No information on conflicts of interest or funding		

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Abdel-Hamid 2014 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Abdel-Razik 2016

Study characteristics	
Patient Sampling	309 participants with chronic hepatitis C were prospectively and consecutively enrolled in a tertiary university centre in Egypt.; 47 excluded according to exclusion criteria (hepatitis B virus (HBV) and HIV, liver transplant, patients who developed HCC in addition to HBV infection, patients with nonalcoholic steatohepatitis and patients receiving certain medications that may increase serotonin levels such as antidepressants or migraine headache medications, patients with hyperlipidaemia, peripheral vascular disease, hypertension, heart failure, and autoimmune diseases). Age range: 29-70. Males 71%
Patient characteristics and setting	Patients with chronic liver disease at tertiary referral centre in Egypt
Index tests	Serum total AFP was assayed using the chemiluminescent immunometric technique on an Immulite 2000 system (Siemens Medical Solutions Diagnostics, Los Angeles, California, USA). The cut-off value of 11.8 ng/mL was derived as the optimal cut-off.
Target condition and reference standard(s)	Hepatocellular carcinoma. Quote: "All studied patients were subjected to a full assessment of history, clinical examination, abdominal ultrasonography, and computed tomography scan to confirm and/or exclude the presence of small HCC. Liver cirrhosis was diagnosed by abnormal biochemical changes, histological examination of liver biopsy, ultrasonography, or endoscopic results suggesting advanced liver disease with portal hypertension. The diagnosis of HCC was made on the basis of a clinical algorithm, triphasic spiral computed tomography of the abdomen, dynamic contrast enhanced MRI of the abdomen, and measurement of AFP."
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Authors declared no conflicts of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Abdel-Razik 2016 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Aboelfotoh 2018

Study characteristics			
Patient Sampling	Quote: "In this prospective study, we recruited 96 patients from Ain Shams University hospitals' clinics and inpatient department, then classified them into three groups; 1) Cirrhosis group: 40 patients with liver cirrhosis without HCC, 2) HCC group: 40 patients with liver cirrhosis and HCC as diagnosed by triphasic CT, 3) Control group: 16 healthy volunteers, with matched age and gender ..." Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum AFP: no specification		
Target condition and reference standard(s)	HCC as diagnosed by triphasic CT: control group: unclear		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest or funding		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Aboelfotoh 2018 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Abu El Makarem 2011

Study characteristics	
Patient Sampling	This prospective case controlled hospital-based study recruiting three groups of individuals attending the Internal Medicine Department of Minia University Hospital between February 2009 and January 2010. A series of patients with HCC with chronic hepatitis C (CHC) was compared with two different groups: one consisted of patients with liver cirrhosis (LC) and the other one, the controls, included individuals who were treated in our hospital for a wide spectrum of acute conditions, other than liver diseases as the primary diagnosis of hospital admission. HCC group consisted of 113 (97 [85.8%] patients were males and 16 [14.1%] patients were females) consecutive patients with HCC, 98 patient (86.7%) diagnosed by means of cytological or histological examination of hepatic focal lesions, while the remaining 15 patients (13.2%) were diagnosed by the appropriate imaging characteristics as defined by accepted guidelines. Liver Cirrhosis group comprised 120 patients (84 (70%) patients were males, 36 (30%) patients were females) with HCV-related LC, by selecting from 250 cirrhotic patients of our department, matched according to age (± 5 years), gender, Patients with LC were admitted to our hospital for diagnosis, staging or therapy of LC. The presence of cirrhosis was defined by histology or non-histologically by evidence of portal hypertension in the presence of chronic liver disease.Controls must have an ultrasound, CT or MRI showing no evidence of hepatic mass within 6 months prior to enrolment. Patients with an elevated AFP (> 20 ng/mL) at enrolment were excluded. Age range: 28-77. Males 78%
Patient characteristics and setting	

Abu El Makarem 2011 *(Continued)*

Index tests	Plasma AFP levels were measured in a plasma sample by the chemiluminescence method using Elecsys AFP kits (Roche Diagnostic GmbH, Mannheim, Germany) according to the manufacturer's instructions.
Target condition and reference standard(s)	HCC diagnosed by means of cytological or histological examination of hepatic focal lesions, while the remaining 15 patients (13.2%) were diagnosed by the appropriate imaging characteristics as defined by accepted guidelines. Controls must have an ultrasound, CT or MRI showing no evidence of hepatic mass within 6 months prior to enrolment.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest or funding

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Abu El Makarem 2011 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Ahmed Mohamed 2016

Study characteristics	
Patient Sampling	Serum samples were obtained from sixty patients with chronic liver disease, divided into two groups: Group (I) included 40 patients with HCC. Patients with cancers other than HCC or metastatic liver cancer were excluded. Group (II) included 20 patients with liver cirrhosis and without any evidence of HCC, and Group (III) included 20 healthy adults recruited as controls. Age range: 48-89. Males 77.5%
Patient characteristics and setting	
Index tests	Serum AFP and osteopontin levels were determined using an enzyme-linked binding protein assay kit. AFP was assayed by an enzyme immunoassay (EIA) Kit (Roche Mannheim, Germany).
Target condition and reference standard(s)	HCC was diagnosed by abdominal US and confirmed by triphasic CT scan. AFP was assayed by an enzyme immunoassay (EIA) Kit (Roche Mannheim, Germany).
Flow and timing	No information on interval between index test and reference standard
Comparative	

Ahmed Mohamed 2016 *(Continued)*

Notes	Authors declared no conflicts of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Ahmed Mohamed 2016 *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ahn 2016

Study characteristics			
Patient Sampling	Between January 2005 and September 2012, we identified consecutive cases of newly diagnosed HCC at three university-affiliated hospitals (the Samsung Medical Center, Seoul St. Mary's Hospital, and Chung-Ang University Hospital) in Seoul, Republic of Korea. For each HCC patient, we selected a cirrhosis control patient matched for age, sex, aetiology, and Child-Pugh classification. Those with end-stage or significant medical comorbidities, in which survival was predicted to be less than 1 year, were excluded. Age range not reported. Males 72%		
Patient characteristics and setting			
Index tests	Serum alpha-foetoprotein		
Target condition and reference standard(s)	We established the diagnosis of HCC by histological examination; the presence of cirrhosis was defined by histology or by evidence of unequivocal clinical and laboratory evidence of cirrhosis, such as ultrasound (US) and/or computed tomography (CT) findings indicating cirrhosis (an irregular liver surface, splenomegaly, etc.) and the detection of signs/symptoms consistent with decompensated cirrhosis (jaundice, varices due to portal hypertension, ascites, or hepatic coma.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Authors declared no conflicts of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Ahn 2016 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Alexander 1978

Study characteristics	
Patient Sampling	Patients who attended the Liver Clinic at Groote Schuur Hospital (South Africa) over the 2-year period 1975 - 1976: 35 HCC; 8 chronic hepatitis B; 12 chronic active hepatitis; 43 alcoholic cirrhosis Age range and % of males not reported
Patient characteristics and setting	

Alexander 1978 *(Continued)*

Index tests	AFP was measured by radio-immunoassay as described by Purves and Purves; the assay is sensitive in the nanogram range and the upper limit of normal is 30 ng/mL.
Target condition and reference standard(s)	The clinical diagnosis of hepatoma was always supported by arteriography, liver scanning and histological examination. Alcoholic cirrhosis was diagnosed if the history indicated prolonged alcohol abuse, and if clinical evidence of cirrhosis and portal hypertension was found on examination. Histological confirmation was obtained when the prothrombin index and platelet count permitted biopsy. The 'diagnosis of chronic active hepatitis was confirmed histologically in every case.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest or funding

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Alexander 1978 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ali 2019

Study characteristics	
Patient Sampling	This study is a case–control, hospital-based study. 120 patients with chronic HCV related liver diseases were included and 60 apparently healthy participants. Group II composed of 60 patients with post HCV liver cirrhosis (LC) diagnosed by clinical, biochemical, and abdominal ultrasonographic findings. Group III composed of 60 patients with HCV-associated HCC on top of LC. Age range: 45-81. Males 68%
Patient characteristics and setting	
Index tests	Serum AFP without other specification; no predefined cut-off value
Target condition and reference standard(s)	HCC was defined on the basis of ultrasound, computed tomography (CT) or magnetic resonance imaging characteristics, serology (AFP), and liver function tests. No definition for controls
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors report no conflicts of interest in this work.
Methodological quality	

Ali 2019 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			

Ali 2019 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Almani 2004

Study characteristics			
Patient Sampling	This case-control study was carried out from 02/2000 till 12/2002 at the department of Internal Medicine, Liquat Univeristy Hospital Jamshoro, Sindh. Among 200 persons studies, 100 were diagnosed with HCC. Age range: 20-65. Males % not reported		
Patient characteristics and setting			
Index tests	AFP was analysed by enzyme immunoassay-based kit. Cut-off value was prespecified at 8.6 ng/mL.		
Target condition and reference standard(s)	HCC: patients presented with liver mass or other symptoms were directed to liver pathology and later diagnosed/confirmed histopathologically as HCC.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Almani 2004 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Alpert 1971

Study characteristics			
Patient Sampling	Sera were obtained from 124 patients (from USA, Uganda and Taiwan), with histologically proven HCC. Control sera were obtained from 337 other patients with various liver diseases. Age range and % of males not reported.		
Patient characteristics and setting			
Index tests	Sera were tested by Ouchterlony double immunoelectrophoresis in agar gel and by quantitative radial immunodiffusion modified to increase sensitivity. Counterimmunoelctrophoresis was adopted from a previously published method.		
Target condition and reference standard(s)	HCC histologically proven; reference standard for control, with other liver diseases, unspecified		
Flow and timing	No information on interval between index test and reference standard: 124 patients with HCC included, analysed 117		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			

Alpert 1971 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Alsebaey 2016

Study characteristics	
Patient Sampling	Eighty-seven people were enrolled into the study. Twenty-two healthy people as a control group (n = 22), 22 patients in the cirrhosis group and finally 43 patients in the HCC group. The diagnosis of cirrhosis was based on clinical, laboratory, and ultrasonography findings (Schuppan and Afdhal, 2008). HCC was diagnosed according to the EASL guideline (European Association for the Study of the et al.). Exclusion criteria were sepsis, GIT bleeding, concurrent medical disease such as long standing diabetes mellitus, chest or cardiac disease. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP no specification
Target condition and reference standard(s)	HCC CT and EASL criteria; controls ultrasound
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest or funding

Alsebaey 2016 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	No		

Alsebaey 2016 *(Continued)*

Could the patient flow have introduced bias?	High risk

Al-Zoubi 2017

Study characteristics	
Patient Sampling	This study included three groups: the hepatocellular carcinoma group (HCC group) contained 26 patients who were diagnosed with HCC for the first time. The diagnosis of HCC was based on typical imaging patterns and/or histological examinations conducted according to EASL–EORTC Clinical Practice Guidelines (12). The chronic liver disease group (CLD group) contained 27 patients who were diagnosed in the same hospitals during the same period as the HCC group. Age range not reported. Males 77%
Patient characteristics and setting	
Index tests	All samples were collected between March 2014 and February 2015 at Al Assad University Hospital and Al Mouwasat University Hospital. Serum AFP levels were routinely evaluated in all patients. This finding was in agreement with the results of Kim in 2006, who found that some HCC patients had AFP levels under 400 ng/mL.
Target condition and reference standard(s)	The diagnosis of HCC was based on typical imaging patterns and/or histological examinations conducted according to EASL–EORTC Clinical Practice Guidelines (12). The chronic liver disease group (CLD group) contained 27 patients who were diagnosed in the same hospitals during the same period as the HCC group.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Authors declared no conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Al-Zoubi 2017 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Amuro 1988

Study characteristics	
Patient Sampling	52 patients (43 males and 9 females) with hepatocellular carcinoma and 42 (30 males and 12 females) with liver cirrhosis were investigated in this study. Age range not reported
Patient characteristics and setting	

Amuro 1988 *(Continued)*

Index tests	Alpha-fetoprotein in serum was determined by a commercially available radioimmunoassay kit (alpha-Feto RIA BEAD, Dinabott, Tokyo, Japan).
Target condition and reference standard(s)	Diagnosis of the diseases was made on the basis of the usual clinical, laboratory, and radiological findings and was confirmed by histological examination of the specimens obtained by liver biopsy, liver resection, or autopsy.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Amuro 1988 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Arrieta 2007

Study characteristics			
Patient Sampling	We obtained 212 files of patients with the diagnosis of HCC and 202 of patients with LC; from which 193 and 74 patients were included, respectively. The main causes of exclusion were: incomplete files, lack of AFP determinations, and an ambiguous diagnosis. Age range not reported. Males 66%		
Patient characteristics and setting			
Index tests	AFP without any specification		
Target condition and reference standard(s)	HCC histology Controls: US, or CT, or RM		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Authors declared no conflicts of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Arrieta 2007 *(Continued)*

Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Arrigoni 1988

Study characteristics	
Patient Sampling	The study population included 164 people with cirrhosis, referred to the Department of Turin from January 1981 to July 1986. The

Arrigoni 1988 *(Continued)*

	patients were prospectively followed as out patients for at least 12 months. Age range: 36-79. Males 66%
Patient characteristics and setting	
Index tests	AFP with commercial Kit (alpha fetoprotein Riabead, Dainabot Co. Ltd Tokyo Japan)
Target condition and reference standard(s)	HCC histology; chronic liver disease: US, AFP, CT
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest or funding

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Arrigoni 1988 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No	
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?	High risk	

Atiq 2017

Study characteristics	
Patient Sampling	We manually abstracted information on patient demographics, clinical history, laboratory data, and imaging results from the EMR. Dates of all HCC surveillance tests between July 2010 and July 2013 were abstracted. HCC surveillance at Parkland is typically performed using ultrasound, with or without AFP, per the AASLD guidelines with low use of surveillance CT or MRI. A total of 680 patients with cirrhosis met inclusion criteria. Age range not reported. Males 64%
Patient characteristics and setting	
Index tests	AFP with a cut-off value of 20 ng/mL; US: We recorded whether ultrasounds were normal (no suspicious masses), positive (suspicious liver mass 1 cm), or indeterminate (mass < 1 cm or unclear if mass is present, e.g. coarse echo texture). Abdominal US: no information on the test and positivity criteria
Target condition and reference standard(s)	HCC: US, CT, RM histology in patients with AFP > 20
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Conflicts of interest: Dr. Singal consults, advises, and is on the speakers' bureau for Bayer. He is on the speakers' bureau and received grants from Gilead. He advises Wako Diagnostics. Dr. Kono advises Wako Diagnostics. Dr. Yopp is on the speakers' bureau for Bayer. He received grants

Atiq 2017 *(Continued)*

	from Peregrine, Merck, and Novartis

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Atiq 2017 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Attallah 2011

Study characteristics	
Patient Sampling	All patients diagnosed with HCC (mean age ± SD, 58 ±10.5 years; male:female ratio, 3.4:1) at TropicalMedicine Unit,Mansoura University hospitals, Mansoura, Egypt between March 2008 to December 2010 were considered eligible for this study.The second group included 100 patients with cirrhosis (mean age ± SD, 50 ± 11.6 years; male:female ratio, 2.8:1). During a 3-year period (2008–2010), 150 consecutive HCC patients and 100 LC patients and 50 healthy individuals were enrolled in the study. Patients with rheumatoid arthritis, hepatitis B viral infection, alcohol abuse, autoimmune liver diseases and metabolic disorders, or other malignancies were not included. Age range not reported. Males 75%
Patient characteristics and setting	
Index tests	AFP level was performed by chemiluminescence, with Immulite AFP (1000) kit (Diagnostic Products Corporation; Los Angeles, CA, USA).
Target condition and reference standard(s)	The diagnosis of HCC was based on AFP levels N200 ng/mL, the presence of hepatic focal lesion (s) detected by liver ultrasound and confirmed by computed tomography and/or magnetic resonance as imaging techniques. The final diagnosis was confirmed by histopathological analysis on ultrasound assisted fine-needle biopsy, when indicated. All the studied participants underwent thorough clinical examination and ultrasonography of the abdomen.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest or funding
Methodological quality	

Attallah 2011 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Attallah 2013

Study characteristics	
Patient Sampling	The estimation population consisted of patients from the Tropical Medicine Unit (Mansoura University Hospitals, Mansoura, Egypt). In this retrospective study, all patients had chronic hepatitis C. Participants were divided into two main groups: group I – HCC which included 227 cirrhotic patients with proved HCC. The non-malignant chronic liver disease (CLD) group included 1124 patients with chronic hepatitis (836 males, 288 females). Patients with the following conditions were excluded from the study: presence of other causes of liver diseases, hepatitis B virus (HBV) infection, or other suspected malignancies. Age range not reported. Males 77%
Patient characteristics and setting	
Index tests	AFP level was performed by chemiluminescence, with IMMULITE AFP (1000) kit (Diagnostic Products Corporation, Los Angeles, CA, USA).
Target condition and reference standard(s)	The diagnosis of HCC in those patients was carried out according to the American Association for the Study of Liver Diseases (AASLD) Practice Guidelines (Bruix and Sherman, 2005). The diagnosis of HCC was based on AFP levels X400Ul1, presence of hepatic focal lesion (s) detected by liver ultrasound (US), and confirmed by computed tomography (CT) and/or magnetic resonance imaging (MRI) techniques. The final diagnosis was confirmed by histopathologic analysis on US-assisted fine-needle biopsy, when indicated. Diagnosis of CLD in this group was based on the standard clinical, biochemical, and ultrasonographic criteria, as well as the pathological data.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Authors declared no conflicts of interest.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Attallah 2013 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Attallah 2017

Study characteristics	
Patient Sampling	659 consecutive patients (318 patients with HCC and 341 with liver cirrhosis), admitted to the Tropical Medicine Unit (Mansoura University Hospitals, Mansoura, Egypt), were enrolled in this study. Age range and % of males not reported
Patient characteristics and setting	

Attallah 2017 *(Continued)*

Index tests	The AFP level was determined by chemiluminescence with an IMMULITE AFP (1000) kit (Diagnostic Products Corporation, Los Angeles, CA, USA
Target condition and reference standard(s)	All of the HCC patients had chronic hepatitis C or liver cirrhosis as the underlying liver disease. The diagnosis of HCC in those patients was made according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines. The tumours were detected by abdominal ultrasound (US) studies and/or AFP assays (> 400 U/L). Each focal lesion detected was further evaluated by multiphase spiral computed tomography (CT) or contrast-enhanced magnetic resonance imaging (MRI).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Attallah 2017 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Attallah 2018

Study characteristics			
Patient Sampling	A total of 182 patients, 110 patients with HCC and 72 patients with liver cirrhosis were included. Age range not reported. Males 60%		
Patient characteristics and setting	82 patients classified into 72 patients with liver cirrhosis, 44 male and 28 female had mean age ± SD (standard division); 52.5±7.1 years and 110 patients with HCC, 82 males and 28 females with age 54.6 ± 10.5 years.		
Index tests	Serum AFP was measured using Immulite AFP-1000 ELISA kit (Diagnostic Products Corporation, Los Angeles, CA, USA). GPC3 was determined by human GPC3 ELISA kit (Wuhan EIAab Science Co., Ltd., Hubei, China).		
Target condition and reference standard(s)	Diagnosis of HCC patients were initially diagnosed by image studies were included US, CT, or magnetic resonance (MRI).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest disclosure		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Attallah 2018 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Attallah 2020

Study characteristics			
Patient Sampling	121 patients with liver fibrosis (fibrosis stages F1-F3) 133 patients with liver cirrhosis (F4) and 148 patients with HCC Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Alpha-foetoprotein (AFP) was measured by chemiluminescence (Immulite 1000, Diagnostic Products Corporation. Cut-off value 400 ng/mL		
Target condition and reference standard(s)	HCC was diagnosed on the basis of liver histological findings or typical imaging characteristics by ultrasound and computed tomography. No definition for controls		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors declared that there was no conflict of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Attallah 2020 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Bachtiar 2009

Study characteristics	
Patient Sampling	Serum samples from 119 patients were collected from the Hepatology Division at the Department Internal Medicine, Cipto Man-

Bachtiar 2009 *(Continued)*

	gunkusumo Hospital, Indonesia. Sera were frozen immediately and stored at −80 °C before use. The group of patients with HCC included 65 patients. The control group of CLD patients comprised 54 patients. Patients displayed CLD related to either hepatitis B virus infection or hepatitis C virus infection (55.6%). Age range: 23-81. Males 76%
Patient characteristics and setting	
Index tests	The qualitative measurement of serum AFP was performed using enzyme immunoassay method (Diagnostic System Laboratories, Webster, TX).
Target condition and reference standard(s)	Diagnosis of HCC relied on the presence of a malignant liver nodule, as established on imaging techniques and by pathological analysis of liver biopsies. In total, there were 65 serum samples from patients with primary HCC at different clinical stages and with various AFP concentration (AFP≤ 200, n = 37 and AFPN200, n = 28). The control group consisted of 54 serum samples from patients with CLD only. CLD patients were defined as persons positive for hepatitis B surface antigen (HBsAg) or positive anti-HCV test for more than 6 months.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest. This work was supported by MRIN Funding (Budget no. cc042/2007).

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		

Bachtiar 2009 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Badr 2014

Study characteristics	
Patient Sampling	During the period from June 2012 to June 2013, we selected 60 patients from the tropical, and internal medicine departments, as well as the oncology centre, of the University hospital and Faculty of medicine, Menoufiya University, Egypt. Thirty of these patients were diagnosed with HCC. The remaining 30 patients had HCV liver cirrhosis.
Patient characteristics and setting	
Index tests	Serum AFP levels were measured by ELISA (MONBIND, Inc. Costa Mesa, CA92627 USA).
Target condition and reference standard(s)	HCC was diagnosed according to history, clinical examination, classic radiological investigations [abdominal ultrasonography (US) and/or triphasic computed tomography], serum AFP levels above 200 ng/mL, and/or histopathological examination of tis-

Badr 2014 *(Continued)*

	sue biopsy when available. All HCC patients were newly diagnosed cases and did not receive prior chemotherapy. Liver cirrhosis was diagnosed by history, clinical features of cirrhosis, abdominal US features, laboratory investigations and/or liver biopsy.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Declared no conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Badr 2014 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Baek 2009

Study characteristics			
Patient Sampling	Study conducted in an University hospital in Korea, enrolled 327 participants 237 HCC, 100 with liver cirrhosis. Age range not reported. Males 69%		
Patient characteristics and setting			
Index tests	AFP radioimmunoassay method, cut-off value 20 ng/mL		
Target condition and reference standard(s)	HCC: US, CT, MR, angiography, histology; controls: follow-up 12 months		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest and funding		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Baek 2009 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Bell 1982

Study characteristics	
Patient Sampling	14 participants with HCC, 110 patients with alcoholic liver disease. Age range 34-93. Males 74%
Patient characteristics and setting	
Index tests	Serum AFP measurement by radioimmunoassay; cut-off value predefined 20 ng/mL

Bell 1982 *(Continued)*

Target condition and reference standard(s)	HCC histology (surgical specimen or autopsy); alcoholic liver disease clinical follow-up
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Funded by Norvegian Cancer Society and National Institute for Alcohol Research. No conflicts of interest reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Bell 1982 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Beneduce 2004

Study characteristics			
Patient Sampling	Serum samples from 160 patients with different liver diseases and from 50 healthy blood donors were analysed. The first subgroup included 60 patients with HCC. The second subgroup included 50 patients with cirrhosis. Age range not reported. Males 64%		
Patient characteristics and setting			
Index tests	Serum AFP levels were determined in parallel in each sample using Beckman Coulter Access reagents for AFP on an Access® I analyzer (Beckman Coulter, CA,®USA).		
Target condition and reference standard(s)	The diagnosis of HCC was based on the presence of a focal liver lesion > 2 cm detected by ultrasonography and confirmed by computed tomography or magnetic resonance imaging. All patients of the control group underwent regular liver ultrasound screening to exclude the occurrence of liver nodules.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Funded by 1XEPTAGEN S.p.A., Pozzuoli (Naples Italy). Four authors are employers of this company.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Beneduce 2004 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Beneduce 2008

Study characteristics	
Patient Sampling	Serum samples from 31 patients with cirrhosis, 33 untreated HCC and 30 healthy controls were studied.

Beneduce 2008 *(Continued)*

	Age range not reported. Males 75%
Patient characteristics and setting	
Index tests	Serum AFP levels were determined using AFP ELISA kit (DRG Diagnostics)
Target condition and reference standard(s)	HCC was diagnosed by ultrasound, computed tomography and/or magnetic resonance and confirmed by histopathology, when indicated.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Some authors are employed by XEPTAGEN SpA, Marghera Venezia, Italy (L. Beneduce, G. Pesce, A. Gallotta, F. Zampieri, G. Fassina).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Beneduce 2008 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Bennett 2002

Study characteristics	
Patient Sampling	Quote: "Between December 1991 and December 2000, 455 patients underwent liver transplantation for cirrhosis at our hospital. We retrospectively reviewed the radiology database. A total of 200 patients were included in our study population. Patients with a lesion that had been detected on previous imaging or those treated with chemoembolisation for a known tumour before undergoing sonography were excluded." Age range: 23-70. Males 67%
Patient characteristics and setting	
Index tests	All sonograms were obtained on one of three types of sonography units. XP128 or Aspen (Acuson, Mountain View, CA) or AI 5200S (Acoustic Imaging, Tempe, AZ) scanners using 2.5-, 3.5-, or 4-MHz transducers. Examinations were performed by experienced technologists. All focal solid lesions were interpreted as potential hepatocellular carcinomas and were described with respect to size, location, and echotexture. Focal areas of heterogeneity were not considered positive findings. Lesions described, as simple cysts were not included in the analysis.
Target condition and reference standard(s)	Explanted livers were serially sectioned into 5 mm to 8 mm sections. Hepatocellular carcinomas and dysplastic nodules were identified grossly as distinct from surrounding nodules in terms of size, texture, colour, and degree of bulging beyond the cut surface of the liver. Nodules were classified using the International Working Party's terminology of nodular hepatocellular lesions.

Bennett 2002 *(Continued)*

Flow and timing	Quote: "We retrospectively reviewed the radiology database to determine which patients had been evaluated with sonography within 90 days before transplantation."
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Bennett 2002 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Bessa 2010

Study characteristics	
Patient Sampling	60 Egyptian patients with HCV-related liver cirrhosis (LC) were selected from those admitted to the Internal Medicine and Tropical Medicine Departments in Tanta University Hospital; among them, 30 patients with HCC and 30 patients without HCC. Age range not reported. Males 70%
Patient characteristics and setting	
Index tests	The second part of the blood sample was drawn into ethylenediaminetetraacetic acid (EDTA) tubes and plasma was obtained by centrifuging the blood sample for 15 minutes at room temperature at 1000 g within 30 minutes after collection, aliquoted, and stored at 80° C until measurements of osteopontin (OPN) and AFP levels. Plasma AFP levels were measured using a commercially available enzyme immunometric assay kit (CanAg AFP EIA kit, Fujirebio Diagnostics AB, Majnabbeterminalen,Goteborg, Sweden), according to the manufacturer's instructions.
Target condition and reference standard(s)	The diagnosis of HCC was based on typical imaging studies and/or histopathology according to American Association for the Study of Liver Diseases (AASLD) practice guidelines. The diagnosis of Liver Cirrhosis was established on the basis of clinical, laboratory, imaging (ultrasonography and computed tomography), and histological examinations.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Bessa 2010 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Best 2016

Study characteristics			
Patient Sampling	In this monocentric study, 285 HCC patients and 402 controls were enrolled from February 2007 to November 2008, and from July 2010 to February 2012 at the University Hospital Essen in Germany. Age range not reported. Males 60%		
Patient characteristics and setting			
Index tests	AFP, AFP-L3, and DCP were measured in the same serum sample using the µTASWakoTM i30 fully automated immunoanalyser (Wako Chemicals GmbH, Neuss, Germany.		
Target condition and reference standard(s)	HCC was diagnosed according to the EASL guidelines via histology or by 2 different imaging modalities. The Barcelona Clinic Liver Cancer (BCLC) staging system was used for determination of disease stage. Patients with viral hepatitis, nonalcoholic steatohepatitis (NASH), autoimmune hepatitis (AIH), liver cirrhosis, and other chronic liver diseases served as the control group. Liver cirrhosis was diagnosed by histology or typical findings such as portal hypertension in known chronic liver diseases.		
Flow and timing	Of 697 patients enrolled, 10 were excluded from analysis because of warfarin medication. No information on interval between index test and reference standard		
Comparative			
Notes	JB received travel grant from WAKO Chemicals GmbH, Neuss, Germany.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		

Best 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Best 2020

Study characteristics	
Patient Sampling	Three hundred fifty-six patients with NASH were enrolled in the German multicentre case-control study, including 125 with HCC and 231 without HCC. Age range: 44-75. Males 57%
Patient characteristics and setting	
Index tests	Serum AFP, no specification; cut-off 10 ng/mL
Target condition and reference standard(s)	HCC was diagnosed according to the European Association for the Study of the Liver (EASL) guidelines via histology or by 2 different imaging modalities (dynamic contrast computed tomography or magnetic resonance imaging of the liver). No specification for controls

Best 2020 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors disclosed no conflicts of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Best 2020 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Biselli 2015

Study characteristics	
Patient Sampling	"Between January 2000 and February 2009, we recruited 80 patients newly diagnosed with HCC in the outpatients' clinic of our centre during a regular semiannual surveillance program based on US and AFP measurement. Between January 2000 and February 2009, we recruited 80 patients newly diagnosed with HCC in the outpatients' clinic of our centres during a regular semiannual surveillance program based on US and AFP measurement. HCC patients (HCC cases) were matched at a 1:2 ratio for the training group and 1:3 for the validation group to simultaneously surveyed patients who remained cancer-free for at least 18 months after enrolment. Matching variables were gender, age (within a 5-year interval), aetiology of cirrhosis.To avoid interference with AFP levels, patients who began or stopped antiviral therapy during the 18 months preceding the HCC occurrence or the enrolment (controls) were excluded." Age range: 33-90. Males 67%
Patient characteristics and setting	
Index tests	AFP serum levels were measured using a commercially available Immunoassay (COBAS ROCHE Diagnostics GmbH, Milan, Italy).

Biselli 2015 *(Continued)*

Target condition and reference standard(s)	Patients with a negative US, but an AFP with a value > 10 ng mL and doubled compared with the previous one, underwent computed tomography. The HCC diagnosis was based on histology in 12 out of 80 (15%) patients in the training group and in 7 out of 36 (19.4%) patients in the validation group. In the remaining patients, it was based on recommended non-invasive criteria. HCC was staged by CT or MRI.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Biselli 2015 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Bolondi 2001

Study characteristics	
Patient Sampling	Between March 1989 and November 1991, a cohort of patients with liver cirrhosis and without HCC. Exclusion criteria were: (1) Child-Pugh C class 16 in patients older than 60 years; (2) a previous diagnosis of focal liver lesion at US; and (3) a serum AFP level > 200 ng/dL. Patients were withdrawn from further surveillance when they were > 60 years old and belonged to Child-Pugh C class, developed other neoplasms, or underwent orthotopic liver transplantation. Age range not reported. Males 62%
Patient characteristics and setting	
Index tests	Cut-off value 20 ng/mL
Target condition and reference standard(s)	Serum AFP determinations and abdominal US, together with physical examination and routine biochemical tests, were repeated every six months. The diagnostic protocol for detection of a nodular liver lesion at US was based on contrast enhanced computed tomography (CT) and echo guided biopsy (when feasible, according to location of the nodule and bleeding risk). When a negative result was obtained after CT and echo guided biopsy, a strict follow up procedure was followed (three month intervals) and the nodule was rebiopsied when an increase in size was detected at US.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	This research was supported by grants of MURST (Italian Ministry for Technological and Scientifi Research). No information on conflicts of interest

Bolondi 2001 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Bolondi 2001 *(Continued)*

Could the patient flow have introduced bias?	High risk

Bon 1998

Study characteristics			
Patient Sampling	37 participants with HCC on liver cirrhosis; controls: 23 participants with liver cirrhosis without HCC. Age range: 33-81. Males 70%		
Patient characteristics and setting			
Index tests	AFP method: chemiluminescence on automatic device ACS180 (Chiron Diagnostics)		
Target condition and reference standard(s)	HCC histology; controls: unspecified		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest and funding		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Bon 1998 *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Brunello 1993

Study characteristics	
Patient Sampling	39 participants with HCC and 16 controls (15 with cirrhosis 1 chronic active hepatitis) Age range: 43-80. Males 78%
Patient characteristics and setting	
Index tests	Serum AFP measurement by immunoturbidimetric method. Cut-off value 20 ng/mL
Target condition and reference standard(s)	HCC histology
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Funding and conflicts of interest not reported
Methodological quality	

Brunello 1993 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Buffet 1988

Study characteristics			
Patient Sampling	A retrospective study of 217 patients admitted to hospital (Paris France) with liver cirrhosis Age range: 16-85. Males 63%		
Patient characteristics and setting			
Index tests	AFP measurement with a cut-off level 250 mcg/mL; US: any focal lesion or diffuse heterogeneity with venous distortion; AFP > 250 + US		
Target condition and reference standard(s)	HCC: histology, CT arteriography or AFP > 250 ng/mL, or clinical follow-up (undefined interval)		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Buffet 1988 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Cabrera 2012

Study characteristics	
Patient Sampling	The study included 143 patients with HCC in the setting of cirrhosis, 61 liver disease controls.

Cabrera 2012 *(Continued)*

	Age range not reported. Males 76%
Patient characteristics and setting	
Index tests	AFP measurement in serum. No prespecified cut-off value
Target condition and reference standard(s)	HCC was diagnosed according to the non-invasive radiological criteria per the American Association for the Study of the Liver Diseases (AASLD) practice guidelines (2008). All controls were evaluated for stage of fibrosis with a liver biopsy and serum collection was performed on the same day as the biopsy. The patients with confirmed HCV-related cirrhosis were enrolled into the surveillance program, received serial cross-sectional imaging every six months, and had no liver masses on enrolment and 12 months after enrolment.
Flow and timing	AFP was measured in 46 of the 61 disease control patients. No information on interval between index test and reference standard
Comparative	
Notes	This study was supported by the NIH KL2 University of Florida Clinical Translational Science Scholar Award, NIH/NCRR award UL1RR029890 and NIH/NCI award K24CA139570.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Cabrera 2012 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Capurro 2003

Study characteristics	
Patient Sampling	Blood samples were obtained from 34 patients with HCC, 20 patients with hepatitis plus liver cirrhosis. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP measurement in serum. No prespecified cut-off value
Target condition and reference standard(s)	HCC was diagnosed histologically when a liver biopsy specimen was available or from clinical information following the guidelines of the European Association for the Study of the Liver (EASL). Sera from patients who were diagnosed with nonmalignant liver disease (hepatitis with liver cirrhosis) at the time of serum collection were only included in this study if there was no indication of malignant disease 6 months after such collection.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The author, M.C, was supported by a fellowship from the Cancer Research Society of Canada.

Capurro 2003 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Capurro 2003 *(Continued)*

Could the patient flow have introduced bias?	High risk

Caviglia 2016

Study characteristics	
Patient Sampling	This single-centre cross-sectional study included 98 prospectively enrolled outpatients (68 men, 30 women; mean age, 62.2 ± 14.1years) with chronic liver disease (CLD) or cirrhosis that underwent US screening for hepatic nodular lesions. All the patients were screened for HCC every 6 or 12 months with abdominal US according to presence or absence of cirrhosis, respectively. Age range not reported. Males 69%
Patient characteristics and setting	
Index tests	Sera were stored at –80°C and subsequently analysed for the concentration of AFP, AFP-L3, and des-γ-carboxy prothrombin (DCP) using an automated immunoassay system assay on the μTASWako i30 immuno-analyser (Wako Chemicals, Neuss, Germany).
Target condition and reference standard(s)	Final diagnosis of HCC was established by four-phase multidetector CT or dynamic contrast-enhanced MRI showing arterial hypervascularity and washout in the venous/late phase.3 The degree of liver disease was classified according to clinical, serological and histological criteria where appropriate. Liver cirrhosis was diagnosed by liver biopsy or by laboratory data and imaging findings (abdominal US and transient elastography).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Caviglia 2016 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Caviglia 2017

Study characteristics	
Patient Sampling	From a cohort of patients HBsAg positive with cirrhosis33 patients with HCC and 30 patients with cirrhosis HbSaG pos were enrolled between December 2012 and June 2015 exclusion criteria: anti HCV positivity, anti HIV positivity, alcohol intake > 40 g/day, concomitant other liver disease; unavailability of at least two serum samples. Age range: 50-64. Males 76%
Patient characteristics and setting	

Caviglia 2017 *(Continued)*

Index tests	Serum AFP measurement by CLEIA system. No predefined cut-off value
Target condition and reference standard(s)	CT for HCC; US for cirrhosis
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Authors declared no conflicts of interest; funded by a grant from the local university

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Caviglia 2017 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Cedrone 2000

Study characteristics			
Patient Sampling	A cohort of 350 consecutive participants with viral chronic liver disease undergoing liver biopsy in an university in Italy. Patients with other aetiologies than viral were excluded. Age range not reported. Males 58.5%		
Patient characteristics and setting			
Index tests	AFP measurement: radioimmunoassay Abbot USA; cut-off value > 20 ng/mL		
Target condition and reference standard(s)	US if positive for focal lesion, biopsy, and histology		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Cedrone 2000 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Chalasani 1999

Study characteristics	
Patient Sampling	Patients with cirrhosis who were evaluated for liver transplantation from January l 1994 – December 31, 1997. All patients evaluated for liver transplantation underwent initial screening consisting of AFP, liver ultrasound, and abdominal CT. Any focal lesions detected on ultrasound or abnormal AFP values (> 20 ng/mL) dur-

Chalasani 1999 *(Continued)*

	ing the extended screening were followed up with an abdominal CT scan. Age range: 39-59. Males 56%
Patient characteristics and setting	
Index tests	AFP with a cut-off value of 20ng/mL; ultrasound: any focal liver lesion
Target condition and reference standard(s)	HCC; biopsy, CT, follow-up
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Data on conflicts of interest not provided

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Chalasani 1999 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Chan 2013

Study characteristics	
Patient Sampling	This was a retrospective-prospective cohort study of consecutive entecavir-treated patients in the out-patient clinic. All patients received entecavir 0,5 mg daily for at least 12 months. Regular HCC surveillance was performed with AFP and US. All HCC cases diagnosed after at least 12 months of entecavir therapy were included. AFP at -12, -9, -6, -3 and 0 (time of HCC diagnosis) from HCC cases and at corresponding time points from non-HCC cases were analysed. Age range not reported. Males 78%
Patient characteristics and setting	
Index tests	AFP with a cut-off value of 6 ng/mL
Target condition and reference standard(s)	No definition or explanation of reference standard
Flow and timing	No information about reference standard nor about the interval between index test and reference standard
Comparative	

Chan 2013 *(Continued)*

Notes	Conflicts of interest: Henry Lik-Yuen Chan – Advisory Committees or Review Panels: Gilead, Vertex, Bristol-Myers Squibb, Abbott, Novartis Pharmaceuticals, Roche, MSD Grace LH Wong – Advisory Committees or Review Panels: Otsuka: Roche Pharmaceuticals, Gilead, Abbott; Speaking and Teaching: Bristol-Myers Squibb, Novartis Pharmaceuticals, Echosens

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Chan 2013 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		High
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Chan 2014

Study characteristics	
Patient Sampling	The study population was composed of consecutive patients who presented at the study centre with liver lesions from May 1997 to March 2003. All patients were managed in either the hepatobiliary surgical unit or the joint hepatoma clinic in the hospital Prince of Wales Hospital, Chinese University of Hong Kong; Inclusion criteria were: (i) presence of one or more focal liver lesions depicted on ultrasonography or computed tomography of the abdomen; (ii) availability of a histological diagnosis of the corresponding liver lesion(s) obtained by resection or percutaneous needle biopsy; (iii) availability of data on the serum AFP concentration within 1 month of the histological diagnosis and before the commencement of any treatment for cancer, and (iv) a patient age of 18 years. Age range not reported. Males 79%
Patient characteristics and setting	
Index tests	Serum AFP concentration was measured by electrochemiluminescence immunoassay (E170 Analytics; Roche Diagnostics Corp., Indianapolis, IN, USA
Target condition and reference standard(s)	Histology of surgical specimen or obtained by US-guided biopsy. Ultrasound-guided percutaneous biopsy was performed in single lesions or in the most suspicious of multiple lesions. If a lesion was diagnosed histologically as non-tumourous or as representing a benign condition, an additional 2-year period of clinical and radiological follow-up was initiated. A diagnosis of a benign condition was considered definitive if there was no change in clinical and radiological outcomes in the 2-year follow-up. Any results of repeated biopsies during the 2-year period were reviewed to ensure the hepatic lesion was not malignant.
Flow and timing	Serum AFP measured within 1 month prior to the histological diagnosis was reviewed
Comparative	
Notes	Conflicts of interest: none declared
Methodological quality	

Chan 2014 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		

Chan 2014 *(Continued)*

Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Chang 1988

Study characteristics			
Patient Sampling	From March 1986 to November 1986, all LC patients with or without HCC who were admitted to the Medical Ward of Kaohsiung Medical College Hospital were recruited for a prospective study. Abdominal sonography, AFP, and complements were examined for HCC screening. Age range: 26-79. Males 87%		
Patient characteristics and setting			
Index tests	AFF levels were examined with a commercial kit (Abbott, North Chicago, IL) The cut-off value of serum alpha-fetoprotein (AFP) level was chosen at 400 ng/mL, as suggested by Chen and Sung.		
Target condition and reference standard(s)	Aspiration cytology, needle biopsy, hepatic angiography, and/or abdominal computed tomography (CT) were performed if HCC was indicated. HCC was verified when the tumour was pathologically proven or the tumour was proven by cyt010gy'~ accompanied with positive findings on angiography or dynamic abdominal CT.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			

Chang 1988 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Chang 2015

Study characteristics	
Patient Sampling	This is a retrospective study that analysed all patients with cirrhosis aged ≥20 years, who were subjected to HCC surveillance between January 2002 and July 2010. The medical histories of these patients were reviewed and recorded until the time of HCC emergence, death, loss to follow-up, or 30 June 2013. The patients were classified according to the aetiologies of liver disease, namely hepatitis B (HBV, the presence of the hepatitis B surface antigen in serum), hepatitis C (HCV, the presence of the hepatitis C antibody in serum), dual HBV and HCV (BC, the presence of both HBV and HCV), and non-B, non-C (NBNC, negative for both HBV and HCV). The exclusion criteria were as follows: (1) the development of a focal liver lesion within the first 18 months, detected using US; (2)

Chang 2015 *(Continued)*

	concurrent extrahepatic neoplasms; (3) a previous liver tumour; and (4) a follow-up duration of < 18 months. Age range: 45-69. Males 65%
Patient characteristics and setting	
Index tests	AFP with cut-off value 20 ng/mL, US: no information on positivity criteria, AFP + US (AFP cut-off value 20 ng/mL)
Target condition and reference standard(s)	The diagnostic workup for HCC was initiated when AFP levels were elevated or a mass lesion was observed on US images. The diagnosis of HCC was based on triple-phase contrast enhanced computed tomography, magnetic resonance imaging, or histopathology. In general, a strict follow-up procedure was followed (1- to 3-month intervals) and computed tomography, magnetic resonance imaging, or biopsy of the liver mass was repeated as indicated, when the initial diagnosis of the mass was inconclusive.
Flow and timing	The diagnosis of HCC was based on triple-phase contrast enhanced computed tomography, magnetic resonance imaging, or histopathology. No information on interval between index test and reference standard.
Comparative	
Notes	Potential competing interests: none

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Chang 2015 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Chayvialle 1977

Study characteristics			
Patient Sampling	200 participants admitted to hospital in Rennes or Lyon from December 1974 to June 1976 underwent AFP measurement and a follow-up (3-18 months). Age range: 25-81. Males 78%		
Patient characteristics and setting	137 participants with cirrhosis (115 alcoholic cirrhosis) and 63 with haemochromatosis (30 with cirrhosis)		
Index tests	Serum AFP radioimmunoassay, cut-off value 7.7 ng/mL		
Target condition and reference standard(s)	Follow up 3-18 months; pathology on surgical specimen or autopsy		
Flow and timing	Tthe interval between index test and reference standard was at least 270 days.		
Comparative			
Notes	The study was funded by the Institut National de la Santé et de la Recherche Médicale (INSERM); no conflicts of interest reported		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Chayvialle 1977 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Chen 1977

Study characteristics	
Patient Sampling	Sera were obtained from 125 patients with hepatocellular carcinoma, and from 74 with cirrhosis of the liver. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP: double immunodiffusion (Micro-Ouchterlony) was carried out on a slide covered with 1% agarose in veronal buffer (pH 8.6, ionic strength 0.075). The antiserum was obtained by immuno-rabbits using a method specified by Gitlin. 5-gR IA was performed with Dainabot Alpha-Feto-125 Kit (Dainabot Radioisotope Lab.,Ltd., Tokyo).
Target condition and reference standard(s)	HCC histology; liver cirrhosis histology
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding and conflicts of interest
Methodological quality	

Item	Authors' judgement	Risk of bias	Applicability concerns

Chen 1977 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Chen 2003

Study characteristics	
Patient Sampling	The target population comprised 60,000 men aged 30–59 living in 15 townships. A blood specimen was obtained from 36,381 of these men, and tested for AFP and HBsAg. In April 1992, a further 1681 high-risk people aged 30–69 were identified using the same enrolment criteria in an additional eight townships of Qidong. Age range: 30-69. Males 100%
Patient characteristics and setting	
Index tests	AFP by reversed passive haemagglutination (R-PHA) and then radioimmunoassay if R-PHA was positive. Participants were considered to be positive when the value for AFP was 20 mg/L.The tests were then repeated within a short interval (usually two weeks to three months). For values of AFP 200 mg/L, the test was repeated at an interval of 2–4 weeks; if the AFP value remained constant or increased, clinical follow-up was undertaken; if the titre of AFP reduced sequentially on two occasions, then the re-examination schedule was as follows: for those with AFP 100–200 mg/L, clinical follow-up was performed every 1–2 months; if the AFP level increased to > 200 mg/L, clinical follow-up was done every 2–4 weeks; if the titre was lower than 100 mg/L at least twice, the case would be followed up at a three-month interval.
Target condition and reference standard(s)	The clinical examination (for participants with AFP > 200 mg/L, ultrasonography examination
Flow and timing	Participants were considered to be positive when the value for AFP was 20 mg/L.The tests were then repeated within a short interval (usually two weeks to three months).
Comparative	
Notes	This research work was partially supported by a grant from Jiangsu Provincial Health Research Project (1989–1990), and from the National 8th Five-year Key ScientiŽc Project of P. R. China (1991–1995).

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Chen 2003 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Chen 2015

Study characteristics	
Patient Sampling	Our study was performed in the Third People's Hospital of Nantong City, and a total of 238 individuals were enrolled in this study. Serum samples from 103 HCC patients in the preoperative state. Age range: 37-80. Males 85%
Patient characteristics and setting	

Chen 2015 *(Continued)*

Index tests	The concentrations of AFP in the serum samples from HCC patients were measured in the Clinical Pathology Laboratory of the third Affiliated Hospital of Nanjing Medical University content of AFP > 20 ng/mL was considered abnormal.
Target condition and reference standard(s)	The diagnosis of all HCC patients was histologically confirmed. The benign group was comprised of 39 patients with liver cirrhosis, 47 patients with chronic hepatitis, and 9 patients with nonalcoholic fatty liver diseases.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Conflicts of interest: none

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Chen 2015 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Chen 2018

Study characteristics	
Patient Sampling	884 eligible participants were consecutively recruited from three hospitals in China (Cancer Hospital of Chinese Academy of Medical Science, Beijing Youan Hospital, and Beijing Friendship Hospital) from November 2013 to December 2014, including 202 HCC patients, 226 patients with liver cirrhosis, 215 patients with chronic HBV infection, and 203 healthy volunteers. Age range not reported. Males 77%
Patient characteristics and setting	
Index tests	AFP: a commercial ELISA kit was used (CanAg, Fujirebio Diagnostics, Göteborg, Sweden), cut-off value 20 ng/mL.
Target condition and reference standard(s)	HCC was diagnosed according to the Chinese guidelines of diagnosis and treatment for HCC; liver cirrhosis was diagnosed according to the guidelines of prevention and treatment for chronic hepatitis jointly proposed by the Chinese Society of Hepatology and the Chinese Society of Infectious Diseases.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Disclosure: "The authors report no potential conflicts of interest in this work."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Chen 2018 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Cheng 2012

Study characteristics			
Patient Sampling	Patients with cirrhosis were followed up for 48 months with up to 8 serial blood tests (among which is AFP). Abdominal ultrasound or CT was performed every 6 months in those without HCC at presentation. Levels of biomarkers, sensitivity, specificity of these biomarkers individually and in combination were investigated for HCV-related HCC and non-HCV-related HCC. Age range not reported. Males 59%		
Patient characteristics and setting			
Index tests	AFP with a cut-off value of 20 ng/mL		
Target condition and reference standard(s)	HCC; US, CT		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No data provided on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Cheng 2012 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		High risk	

Chimparlee 2015

Study characteristics	
Patient Sampling	Serum samples for the measurement of OPN and AFP levels were obtained from patients who were diagnosed with HBV-related HCC for the first time at King Chulalongkorn Memorial Hospital from January 2010 to December 2014. The control groups comprised 3 groups and included healthy volunteers with no apparent liver disease, patients with chronic hepatitis, and patients with liver cirrhosis. All patients with HCC or chronic liver disease included in the current study were positive for serum hepatitis B surface antigen (HBsAg) for the previous 6 months. Patients with hepatitis C virus (HCV) and/or HIV co-infection were excluded. Age range not reported. Males 83%
Patient characteristics and setting	
Index tests	Blood samples were obtained at initial presentation; sera were separated by centrifugation and stored at –70°C until tested. Serum AFP levels were determined using a commercially available ELISA kit according to the manufacturer's recommendations (Cobus'Core, Roche Diagnostics, Basel, Switzerland), using the normal upper limit of AFP (20 ng/mL) as a cut-off point.
Target condition and reference standard(s)	The diagnosis of HCC was based on typical imaging studies and/or histopathology according to American Association for the Study of Liver Diseases (AASLD) practice guidelines.

Chimparlee 2015 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	This study was supported by the Rajadapiseksompoj research grant, Faculty of Medicine, Chulalongkorn University. The study was also supported by National Research University Project, Office of Higher Education Commission (WCU011-HR57) and the Ratchadaphiseksomphot Endowment Fund of Chulalongkorn University (CU-57-001-HR).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Chimparlee 2015 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Choi 2019

Study characteristics	
Patient Sampling	This was a phase 3 biomarker study based on the EDRN definition. Serum samples were collected from four previous prospective studies conducted by our group (Fig. 1): one EDRN biomarker phase 4 HCC surveillance study for cirrhosis (the PRIUS study, clinicaltrials.gov, registration no. NCT 01446666, 407 patients) and three randomised controlled trials (RCTs) to explore the optimal antiviral treatment regimen in patients with chronic hepatitis B (CHB; NCT01639066, 102 patients; NCT01639092, 90 patients; and NCT01023217, 90 patients). Age range not reported. Males 69%
Patient characteristics and setting	
Index tests	AFP was measured using a chemiluminescent microparticle immunoassay (ARCHITECT i2000SR; Abbott, Chicago, IL). All clinical data, including the presence of HCC, were blinded to laboratory technicians to avoid measurement bias. Diagnosis of HCC was triggered only by suspicious nodule on surveillance images (US, CT, and/or MRI). Biomarkers were not involved in the decision making. US: no specification US + AFP: at least one positive, no other specification
Target condition and reference standard(s)	Confirmation of HCC was based on the predefined criteria by study protocols, i.e. results of histologic examination and/or typical imaging features (nodule > 1 cm with arterial hypervascularity and portal/delayed-phase washout) by CT and/or MR.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Supported by grants from the Korean Gastroenterology Fund for Future Development; the Korean National Health Clinical Research project, Ministry of Health & Welfare, Republic of Korea (HC15C3380); the Korean Health Technology R&D Project, Ministry

Choi 2019 *(Continued)*

of Health & Welfare (HI17C1862); the National Research Foundation of Korea (NECA-S-17-008); and the Technology Innovation Program (10079271) funded by
the Ministry of Trade, Industry & Energy of the Republic of Korea.

Potential conflict of interest: Dr. Lim consults, advises, is on the speakers bureau for, and receives grants from Bayer Healthcare and Gilead Sciences.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Choi 2019 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Chuaypen 2018

Study characteristics	
Patient Sampling	Patients and blood samples for the measurement of WFA+-M2BP levels were obtained from patients who were diagnosed with HBV-related HCC for the first time at King Chulalongkorn Memorial Hospital (Bangkok, Thailand) between May 2010 and December 2015. In this study, 150 patients with HCC were recruited. The control group comprised 150 patients without HCC. Exclusion criteria for the control group were: (i) co-infection with HCV and/or HIV; (ii) previous HBV antiviral treatment; and (iii) evidence of HCC or other cancers during follow-up. Age range: 30-82. Males 76%
Patient characteristics and setting	
Index tests	Serum AFP levels were measured by a commercially available ELISA kit (Cobas Core; Roche Diagnostics, Basel, Switzerland.
Target condition and reference standard(s)	Hepatocellular carcinoma was diagnosed on the basis of typical imaging studies and/or histopathology (fine needle aspiration, core liver

Chuaypen 2018 *(Continued)*

	biopsy, or surgical resection) according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	This study supported by the National Research University (NRU59-026-HR), the Thailand Research Fund (RTA5980008), and the Rachadapisek Sompot Fund for Postdoctoral Fellowship, Chulalongkorn University. This study was also supported by the Japan Society for the Promotion of Science (KAKENHI Grant No. JP15H05289).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Chuaypen 2018 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Cottone 1983

Study characteristics			
Patient Sampling	A prospective cohort of 100 participants with cirrhosis and clinically suspected HCC Age range not reported. Males 68%		
Patient characteristics and setting			
Index tests	US with defined positivity criteria for HCC		
Target condition and reference standard(s)	HCC: if focal lesion at US biopsy; if US negative, follow-up 12 months		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or COI		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	

Cottone 1983 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Cottone 1988

Study characteristics	
Patient Sampling	A prospective cohort of participants with compensate liver cirrhosis in a university centre in Italy Age range: 40-77. Males 57%
Patient characteristics and setting	

Cottone 1988 *(Continued)*

Index tests	AFP measurement by radioimmunoassay (Abbot USA) with a cut-off value > 20 ng/mL
Target condition and reference standard(s)	US and histology
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Cottone 1988 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Cui 2002

Study characteristics			
Patient Sampling	60 participants with HCC and 30 with cirrhosis Age range: not reported. Males 71%		
Patient characteristics and setting			
Index tests	Serum AFP measurement with a cut-off value > 20 ng/mL		
Target condition and reference standard(s)	HCC; US, CT, MR, angiography, histology Controls: US, CT, and follow-up 12 months		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Cui 2002 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Cui 2003

Study characteristics			

Cui 2003 *(Continued)*

Patient Sampling	Serum PIVKAII and AFP levels and GGTII activity were determined in 90 patients with cirrhosis and 120 patients with HCC. Patients with vitamin K and antibiotic use in the recent 3 months, with a haemoglobin levels under 490 mg dL and free bilirubin concentrations up to 27 mg dLor conjugated bilirubin concentrations up to 22 mg dL were excluded from this study. Age range: 32-84. Males 70%
Patient characteristics and setting	
Index tests	The serum concentration of AFP was determined by electrochemiluminescence immunoassay (Roche, Elecsys 1010/2010 Systems) according to the manufacturer's instructions. The cut-off level was fixed at 20 ng/mL.
Target condition and reference standard(s)	In all, 58% (70 out of 120) of HCC patients were diagnosed by fine needle biopsy under the guidance of ultrasonography, and in 16% (19), the diagnosis was confirmed after surgery. Ultrasonography, CT, MRI, and selective celiac angiography diagnosed the remaining patients (26%, 31 out of 120). In patients with cirrhosis, HCC was ruled out on the basis of imaging examinations including sonography and CT)performed on a regular basis. Also, patients with cirrhosis who developed HCC within 1 year from getting serum were excluded.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		

Cui 2003 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

da Costa 2015a

Study characteristics	
Patient Sampling	In Thailand, specimens were obtained from patients and hospital-based controls recruited at the cancer control unit of the National Cancer Institute of Thailand, Bangkok (TLCS, Thailand liver cancer study, Case-control 1). The study was conducted from April 2008 to December 2009. All cases of PLC were recruited and matched controls were obtained from outpatient clinics. Age range and % of males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement, with a cut-off value > 20 ng/mL
Target condition and reference standard(s)	Differential diagnosis of HCC versus CC was established by a combination of clinical examination,imaging using ultrasonography, computerised tomography (CT) or MRI, biochemistry (AFP and liver function enzymes testing) and histological confirmation on a

da Costa 2015a *(Continued)*

	small subset of patients from whom needle biopsies were available.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Grant sponsor: European Union Collaborative Project Prolifica, 7th Framework Programme, FP7-AFRICA-2010, Health-F2-2011-265994

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

da Costa 2015a *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

da Costa 2015b

Study characteristics			
Patient Sampling	The Gambia (GLCS, Gambia liver cancer study, Case-control 2), specimens were obtained in the course of a nationwide case-control study performed between 1997 and 2001 in three tertiary referral hospitals as described previously. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum AFP measurement.with a cut-off value > 20 ng/mL		
Target condition and reference standard(s)	Differential diagnosis of HCC versus CC was established by a combination of clinical examination, imaging using ultrasonography, CT, or MRI, biochemistry (AFP and liver function enzymes testing), and histological confirmation on a small subset of patients from whom needle biopsies were available.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Grant sponsor: European Union Collaborative Project Prolifica, 7th Framework Programme, FP7-AFRICA-2010, Health-F2-2011-265994		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

da Costa 2015b *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

da Costa 2015c

Study characteristics	
Patient Sampling	In France (FLCS, French liver cancer study, Case-control 3), specimens were obtained from patients and controls recruited at Hopital Croix-Rousse in Lyon between September 2011 and May 2012. Age range and % of males not reported

da Costa 2015c *(Continued)*

Patient characteristics and setting	
Index tests	Serum AFP measurement, with a cut-off value > 20 ng/mL
Target condition and reference standard(s)	HCC was diagnosed according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines (using AFP, US, CT, MR, and histology)
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Grant sponsor: European Union Collaborative Project Prolifica, 7th Framework Programme, FP7-AFRICA-2010, Health-F2-2011-265994

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

da Costa 2015c *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

da Costa 2015d

Study characteristics			
Patient Sampling	In Korea, a prospective cohort was assembled using specimens obtained from chronic hepatitis or cirrhosis patients and controls recruited in at the Kosin University Hospital in Busan between 1999 and 2001, with subsequent follow-up until 2006 (KLCS, Korean liver cancer study, Cohort 1). Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum AFP measurement with a cut-off value > 20 ng/mL		
Target condition and reference standard(s)	Incident HCC was diagnosed according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Grant sponsor: European Union Collaborative Project Prolifica, 7th Framework Programme, FP7-AFRICA-2010, Health-F2-2011-265994		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

da Costa 2015d *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ding 2020

Study characteristics

Ding 2020 *(Continued)*

Patient Sampling	CHB and HBV-HCC patients were identified in the Ruijin Hospital (Shanghai, China) from December 2007 and March 2019. The exclusion criteria of CHB for the current study were as follows: hepatitis C virus (HCV) infection (n = 303), autoimmune hepatitis (AIH, n = 176), primary biliary cholangitis (PBC, n = 14), HCC (n = 36), and undetermined liver disease (n = 334). Furthermore, the patients without complete data (n = 1675) or who had other liver diseases (n = 16) were excluded. Age range not reported. Males 76%
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. Cut-off value - no definition
Target condition and reference standard(s)	Diagnosis of HCC, based on the American Association for the Study of Liver Diseases (AASLD) practice guidelines for the treatment of HCC. The definition of CHB was based on the 2018 AASLD practice guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared no conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Ding 2020 *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Dodd 1992

Study characteristics			

Dodd 1992 *(Continued)*

Patient Sampling	From July 1990 to November 1991, all patients undergoing evaluation for potential hepatic transplantation at our institution were entered into a prospective sonographic screening study focused on the detection of hepatic malignant tumours. All patients in the study had hepatic sonograms obtained before transplantation. In the patients who underwent subsequent hepatic transplantation, the sonographic results were directly correlated with the resected total hepatectomy specimens. For this report the authors evaluated the sonographic pathologic correlative results from 200 transplant recipients with histologically-proved hepatic cirrhosis. Age range: 18-74. Males 61%
Patient characteristics and setting	
Index tests	Quote: "Each screening sonogram consisted of preliminary scanning of the liver by one of seven technologists. The livers were scanned according to protocol in both transverse and longitudinal planes with anterior and lateral intercostal and subcostal probe placement. The protocol included hard-film documentation of the liver in sequential transverse images from the dome to the caudal tip of the right lobe and sequential longitudinal images from the extreme right to the extreme left margin of the liver. All scans were obtained with similar sonographic equipment (Acuson Corp., Mountain View, CA) using 2.5- or 3.5-MHz phased-sector transducers. The technologist's results were reviewed with a staff radiologist. As per standard procedure at our institution, approximately 65% of the patients were rescanned by a radiologist for either clarification of the technologist's results or quality control. Approximately 90% of the sonograms were interpreted by two radiologists specialising in sonography. The remaining sonograms were interpreted by four radiologists from the division of abdominal imaging who routinely attend in the ultrasound service. The sonographic criteria for the diagnosis of a possible malignant tumour consisted of identification of a discrete focal mass distinguishable from the adjacent hepatic parenchyma or a poorly marginated, focal region of heterogeneous echogenicity. Both types of abnormalities were interpreted as suggestive of malignancy regardless of their intrinsic echogenicity. Diffuse heterogeneous echogenicity was attributed to cirrhosis rather than tumour unless it was associated with mass effect on, or thrombosis of, the intrahepatic vessels. The only focal lesions considered benign and excluded from the study were those meeting the strict criteria for simple cysts."
Target condition and reference standard(s)	At the time of transplantation, the results of the pretransplantation sonograms were directly correlated with serially sectioned fresh total hepatectomy specimens. For ease of correlation, all livers were sliced in the transverse plane at 1 cm intervals. Lesions were matched by location and size. Each sonographically identified lesion was either confirmed or refuted on the basis of pathologic findings. Sonographically missed lesions were recorded.
Flow and timing	The time from sonography to transplantation varied from 1 to 343 days (mean, 63 days), with 86% of the sonograms obtained within 120 days.
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Dodd 1992 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		

Dodd 1992 *(Continued)*

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Dong 2015

Study characteristics			
Patient Sampling	A total of 584 participants who visited Hangzhou First People's Hospital from June 2011 to June 2013 were enrolled in this study. They were divided into four age- and gender-matched groups (HCC, liver cirrhosis, chronic hepatitis B patients, and healthy participants). The participants who presented with other liver diseases, such as autoimmune hepatitis, alcoholic hepatitis, and other types of hepatitis virus infections were excluded from this study. Age range not reported. Males 68%		
Patient characteristics and setting			
Index tests	Serum AFP measurement; cut-off value 20 ng/mL		
Target condition and reference standard(s)	In all, 190 patients with HCC had been diagnosed by serum AFP level, liver ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI). Those who met the diagnostic criteria for HCC, which was confirmed by histological examination, were enrolled.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors declared that they had no competing interests.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Dong 2015 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Durazo 2008

Study characteristics	
Patient Sampling	240 patients with either hepatitis B virus (HBV) or hepatitis C virus (HCV) infection were studied. Age range not reported. Males 80%
Patient characteristics and setting	The study included 144 with HCC, 47 with chronic hepatitis (fibrosis stage I-III on liver biopsy), and 49 with cirrhosis.

Durazo 2008 *(Continued)*

Index tests	AFP was tested using an immunometric assay utilising cheminoluminescence (Wako Diagnostic); no prespecified cut-off value.
Target condition and reference standard(s)	The diagnosis of HCC was based on American Association for the Study of Liver Diseases (AASLD) practice guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest disclosure

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Durazo 2008 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Edis 1998

Study characteristics			
Patient Sampling	The study group was made up of 110 cirrhotic patients who were seen at our hospital before liver transplantation between 1989 and 1997. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum samples were stored at -80°C, avoiding repeated freezing and thawing, and were tested for AFP and anti-p53 prior to any therapeutic intervention. AFP was detected by standard radioimmunoassay (RIA) and levels above 10 ng/L were regarded as positive.		
Target condition and reference standard(s)	Diagnosis of HCC was made by ultrasound investigation, computed tomography (CT) scan, CT portography, angiography, biopsy if possible, and histologic examination of the explanted liver.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		

Edis 1998 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Edoo 2019

Study characteristics	
Patient Sampling	The study included a total of 1362 patients who were admitted for treatment or who had their physical check-up in the First Affiliat-

Edoo 2019 *(Continued)*

	ed Hospital of Zhejiang University (Hangzhou, China) from 2014 to 2018.Totally 1075 participants were patients with primary hepatic cancer (PHC). 237 patients were diagnosed with liver cirrhosis. A retrospective analysis of patients diagnosed with primary hepatic cancer and liver cirrhosis was collected from the hospital database. Age range and % of males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. Cut-off value 20ng/mL
Target condition and reference standard(s)	Confirmed pathological diagnosis
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared no conflict of interests.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Edoo 2019 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Eissa 2013

Study characteristics	
Patient Sampling	In the present study, we measured serum nitric oxide and glutathione reductase levels in patients with HCC, and in cirrhotic patients. From March 2012 to September 2012, 50 patients with HCC (37 males and 13 females; aged 40–90 years with a mean ± SE of 60.7 ± 1.29) were recruited from the Oncology Center, Mansoura University, Mansoura, Egypt. Controls: a group of 30 cirrhotic patients (19 males and 11 females; aged 33–80 years with a mean ± SE of 56.4 ± 1.6), without any evidence of HCC was used and selected from the outpatient clinic. Age range: 40-90. Males 70%
Patient characteristics and setting	
Index tests	Serum a-fetoprotein was measured using a commercially available ELISA kit from (DiaMetra Company), cut-off value > 200 ng/mL.
Target condition and reference standard(s)	All cases were tested for either pathological proof or a typical radiologic pattern on the post-contrast study plus the diagnostic serum AFP.
Flow and timing	No information on interval between index test and reference standard

Eissa 2013 *(Continued)*

Comparative	
Notes	Competing interests: the authors declared no conflict of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Eissa 2013 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

El-Abd 2015

Study characteristics			
Patient Sampling	This study was conducted on 100 Egyptian adults including 40 (27 males and 13 females) newly diagnosed HCC patients with HCV liver cirrhosis with a mean age of 56.5 ± 5.7 years, 40 patients (23 males and 17 females) with chronic HCV with liver cirrhosis without HCC, with a mean age of 56.4 ± 7.7 years, and 20 (12 males and 8 females) apparently healthy participants as a control group, with a mean age of 32.9 ± 2.2 years. Patients with HCC and chronic HCV were recruited from the Gastroenterology and Hepatology Department, Cairo University. All patients and controls were subjected to full history taking. HCC patients were diagnosed by triphasic abdominal CT scan. Data of all participants were obtained from medical records and personal interviews. All laboratory tests were assayed in the Chemical Pathology Unit, Cairo University Hospital. Age range: 40-64. Males 63%		
Patient characteristics and setting			
Index tests	Sera from chronic HCV and HCC patients were used for estimation of serum level of AFP by solid phase two sequential chemiluminescent immunometric assay using IMMULITE 2000 system analyser. The kits were supplied by Siemens (Siemens Healthcare Diagnostics, United States, cat#L2KAP2). Values up to 10 ng/mL were considered normal.		
Target condition and reference standard(s)	HCC: CT		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

El-Abd 2015 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

El-Abd 2016

Study characteristics	
Patient Sampling	This is a case-control study that was conducted over a period of consecutive 6 months from April 2013 to September 2013. Participants

El-Abd 2016 *(Continued)*

	were classified into two groups: group (I) included 50 patients with HCC. Group (II) included 30 cirrhotic patients. The studied patients (group I and II) were recruited from those presented to the outpatient clinic of the Endemic Hepatogastroenterology Department of Kasr El Aini Hospital (Cairo University, Egypt) and National cancer institute (Cairo University, Egypt). Age range not reported. Males 70%
Patient characteristics and setting	
Index tests	AFP was done using Architect, based on the chemiluminescence Immunoassay (CLIA) technology; no prespecified cut-off value.
Target condition and reference standard(s)	HCC was diagnosed in respect to AASLD practice guidelines. The diagnosis of focal lesions was originally detected by ultrasonography. Quote: "We used multidetector CT scan to confirm the presence of hypervascular lesions in the arterial phase that washed out in the portal venous or delayed phases. If lesions showed atypical findings, confirmatory step. No patients needed to be biopsied. The presence of cirrhosis was diagnosed by ultrasonography. All cirrhotic patients underwent regular ultrasonographic screening (every 4–6 months) to exclude the development of any hepatic nodules."
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared that they had no conflict of interest. The study was not sponsored by any organisation.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

El-Abd 2016 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

El Gawad 2014

Study characteristics	
Patient Sampling	This study included 40 newly diagnosed HCC patients, all patients who were presented to the outpatients' clinic at the NCI, Cairo University, as well as the National Liver Institute, Cairo over a period of consecutive 9 months from January to September 2012, and were eligible for the study. Exclusion criteria: prolonged obstructive jaundice, intrahepatic cholestasis with vitamin K deficiency, and intake of warfarin or antibiotics. Age range: 44-77. Males 64%
Patient characteristics and setting	
Index tests	Serum AFP cut-off value 20 ng/mL
Target condition and reference standard(s)	They were proven to be HCC by computed tomography (CT) or magnetic resonance imaging (MRI).
Flow and timing	No information on interval between index test and reference standard

El Gawad 2014 *(Continued)*

Comparative	
Notes	Conflict of interest: none declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

El Gawad 2014 *(Continued)*

Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

el-Houseini 2005

Study characteristics			
Patient Sampling	Blood samples were collected from patients at the National Cancer Institute (NCI) of Cairo during a 1-year period. Patients were diagnosed according to radiological imaging, laboratory tests, and clinical investigation following the institutional protocol. Two groups: 44 patients with HCC and 20 patients with cirrhosis but without HCC. Age range not reported. Males 72%		
Patient characteristics and setting			
Index tests	A commercially available microparticle enzyme immunoassay was used to determine the serum level of AFP expressed in ng/mL, with no prespecified cut-off value.		
Target condition and reference standard(s)	Patients were diagnosed according to radiological imaging, laboratory tests, and clinical investigation following the institutional protocol.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

el-Houseini 2005 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

El Mahdy 2019

Study characteristics	
Patient Sampling	Three groups: Group I (control): 60 apparently healthy individuals Group II (cirrhosis): 75 patients with liver cirrhosis Group III (HCC): 60 patients with HCC Age range and % of males not reported
Patient characteristics and setting	
Index tests	The analysis of serum alpha fetoprotein (AFP) (ng/mL) was done by IMMULITE 1000 system supplied by Siemens kit (SIEMENS Medical Solutions Diagnostics, USA).

El Mahdy 2019 *(Continued)*

	No pre-definition of cut-off value		
Target condition and reference standard(s)	All HCC patients were diagnosed by characteristic vascular enhancement pattern detected by multislice triphasic spiral CT scan or MRI according to established diagnostic criteria. Cirrhosis-control US		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Authors declared no conflict of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			

El Mahdy 2019 *(Continued)*

DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

El Moety 2011

Study characteristics	
Patient Sampling	The study included 80 patients and 15 normal participants. They were grouped as follows: Group (1) 50 patients with hepatocellular carcinoma Group (2) 30 patients with chronic hepatitis C Group (3) 15 normal participants Complete history taking, clinical examination stressing on (liver and spleen size, ascites, jaundice, Encepathalopathy, and liver masses). Laboratory testing after overnight fasting (primary biliary cholangitis (primary biliary cirrhosis; CBP), alanine transaminase (ALT), aspartate aminotransferase (AST), bilirubin, albumin, hepatitis B surface antigen (HBsAg), hepatitis C virus antibody (HCV

El Moety 2011 *(Continued)*

	AB), and alpha-fetoprotein, and nitric oxide. Child Pugh score. Abdominal ultrasound for detecting for hepatic lesions. Triphasic CT for diagnosis of focal hepatic lesions as hepatocellular carcinoma with the characteristic pattern. Age range and % of males not reported
Patient characteristics and setting	
Index tests	Serum AFP with no prespecified cut-off value
Target condition and reference standard(s)	HCC; CT
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No data provided on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

El Moety 2011 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Elnemr 2012

Study characteristics			
Patient Sampling	60 participants with HCC and 60 with cirrhosis were enrolled in a university medical centre in Egypt. Age range not reported. Males 83%		
Patient characteristics and setting			
Index tests	Serum AFP measurement by elechtochemiluminescence immunoassay using a Cobas E411 analyzer (Roche Diagnostics, Tokyo, Japan) with a cut-off value of 20 ng/mL		
Target condition and reference standard(s)	HCC: AFP, US, CT; cirrhosis: US		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No author had any conflict of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Elnemr 2012 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

El-Serag 2017

Study characteristics

El-Serag 2017 *(Continued)*

Patient Sampling	The authors presented interim results from a prospective cohort study (8/14-5/17) at the Houston VAMC. Quote: "We enrolled consecutive patients with cirrhosis irrespective of aetiology and no past or present HCC in a 6-monthly surveillance program consisting of liver imaging (mostly ultrasound) combined with AFP. We limited the analysis to 26 HCC cases with complete information on biomarkers before HCC development and 543 controls with consistently negative liver imaging." Age range not reported. Males 98%
Patient characteristics and setting	
Index tests	Serum AFP measurement; cut-off value 20 ng/mL
Target condition and reference standard(s)	US, CT, MR, histology; follow-up 6 months
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Abstract. No information on funding or conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

El-Serag 2017 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

El Shafie 2012

Study characteristics	
Patient Sampling	66 patients, selected from the hepatology department of National Liver Institute-Menoufyia University, National Cancer Institute-Cairo University, Internal Medicine-Al Zahraa University Hospital and Hepatology Centre - National Medical Centre; 31 of them were diagnosed as HCC according to clinical examination, radiological investigations including abdominal ultrasonography, triphasic CT, and laboratory investigations. The remaining 35 patients had post HBV or HCV liver cirrhosis. Age range: 34-71. Males 84%
Patient characteristics and setting	
Index tests	Serum AFP was measured using automated Elecsys (Roche- Diagnostic, Branchburg, NJ-Germany) with no prespecified cut-off value.
Target condition and reference standard(s)	HCC, CT Controls: US
Flow and timing	No information on interval between index test and reference standard
Comparative	

El Shafie 2012 *(Continued)*

Notes	Conflicts of Interest: there were no conflicts of interest in this study.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

El Shafie 2012 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

El-Shenawy 2012

Study characteristics			
Patient Sampling	In this study, the patients were selected from the Department of Hepatology, National Liver Institute, Minoufiya University and Department of Oncology, Faculty of Medicine, Minoufiya University. There were two groups: Group (1) included 57 patients (48 males and 9 females; mean age: 46.87 ± 6.58 years). The patients were diagnosed as having HCC by the presence of characteristic hepatic masses on liver MRI, CT, and/or hepatic angiography (i.e. enlarged tumours and/or tumours with typical arterial vascularisation). Group (2) [liver cirrhosis (LC)]. The group included 46 patients (37 males and 9 females; mean age: 42.28 ± 9.34 years). The group of patients consisted or people having hepatitis B virus and/or hepatitis C virus–related cirrhosis. Age range not reported. Males 82.5%		
Patient characteristics and setting			
Index tests	Assessment of serum AFP was performed using VIDAS instrument, BioMerieux, France, using the Enzyme Linked Fluorescent Assay (ELFA). The results were expressed as IU/mL with no prespecified cut-off value.		
Target condition and reference standard(s)	HCC, CT, MR		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		

El-Shenawy 2012 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

El-Sherif 2012

Study characteristics	
Patient Sampling	The present study was carried on 80 participants. Patients with either liver cirrhosis or HCC were selected among patients who were admitted in the Department of Tropical Medicine and Gastroenterology and Internal medicine; Assiut University Hospital from

El-Sherif 2012 *(Continued)*

	September 2009 - September 2010. They were 30 people with liver cirrhosis (group I) and 30 people with HCC (group II). Age range not reported. Males 70%
Patient characteristics and setting	
Index tests	Alpha-foetoprotein was performed on IMMULITE analyser, using chemiluminescent assay (Siemens Healthcare Diagnostics, UK) with no predefined cut-off value.
Target condition and reference standard(s)	The diagnosis of HCC was ascertained using a histopathologic examination by liver needle biopsy (30 adults were enrolled in the present study; the patients underwent liver needle biopsy under ultrasound guidance). The diagnosis of cirrhosis was confirmed by abdominal ultrasonography and biochemical findings.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

El-Sherif 2012 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Eltaher 2016

Study characteristics	
Patient Sampling	60 HCV-positive patients either attended or were admitted to the Department of Hepatology and Gastroenterology or Internal Medicine, Benha University Hospital, Egypt from October 2014 to March 2015. The study population was divided as follows. Group I: 30 HCV-positive patients with HCC, aged 32–64 years Group II: 30 HCV-positive patients with liver cirrhosis, aged 34–58 years. Males 55%
Patient characteristics and setting	
Index tests	AFP levels (0.3–1000 ng/mL) were assessed by Axsym using microparticle enzyme immunoassay (MEIA) technology with no prespecified cut-off value.
Target condition and reference standard(s)	We used triphasic CT, and/or MRI to detect characteristic focal lesions of HCC, with or without elevated AFP.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Funding: none. Conflicts of interest: none declared
Methodological quality	

Eltaher 2016 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

El-Tayeh 2012

Study characteristics			
Patient Sampling	Samples taken during routine follow-up of 96 patients with different liver diseases are utilised in the current study. They included 37 patients with HCC, 28 patients with liver cirrhosis. Age range: 41-70. Males 78%		
Patient characteristics and setting			
Index tests	AFP is determined using commercially available microparticle enzyme immunoassay. AFP is expressed in ng/mL with no predefined cut-off value.		
Target condition and reference standard(s)	Patients are diagnosed according to radiological imaging, laboratory tests, and clinical investigations following the institutional protocol.		
Flow and timing			
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

El-Tayeh 2012 *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

El Zefzafy 2015

Study characteristics	
Patient Sampling	This study was conducted on 60 adult patients: 30 patients with CHCV infection and 30 patients with HCC who presented to Tropical, Internal Medicine Department of Al-Zahraa University Hospital, from March to November 2014. Exclusion criteria: patients with history or evidence of other malignancies; patients suffering from any other organ failure; and other causes of cirrhosis e.g. alcohol. Age range: 35-62. Males 60%
Patient characteristics and setting	
Index tests	Serum alpha-foetoprotein (AFP) was detected by COBAS e411 chemiluminescence auto-analyser, using Roche reagents (Roche Diagnostics GmbH, D-68289 Mannheim, Germany) with no predefined cut-off value.
Target condition and reference standard(s)	All patients and controls were subjected to the following: 1. Full history and clinical examination. 2. Abdominal ultrasonography. 3. Abdomino-pelvic triphasic CT scan for suspected cases of HCC.

El Zefzafy 2015 *(Continued)*

	Liver biopsy and histopathological examination were done when needed for patients with hepatic focal lesions, not fulfilling imaging or AFP diagnostic criteria for HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

El Zefzafy 2015 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Erdal 2016

Study characteristics			
Patient Sampling	133 participants were enrolled in our study and were divided into three groups: HCC (n = 40), cirrhosis (n = 54), and control (n = 39). Patients with another malignancy were excluded from the study. Age range: 45-87. Males 73%		
Patient characteristics and setting			
Index tests	AFP was measured by the chemiluminescence method (ARCHITECH system; Abbott Laboratories, Abbott Park; IL, USA). The upper limit of the normal level is 7 ng/mL.		
Target condition and reference standard(s)	The diagnosis of 22 HCC patients was made by histopathology. If histopathology was not present, the diagnosis of HCC was based on the American Association for the Study of Liver Diseases (AASLD) practice guidelines (11), and it was confirmed by imaging modalities (ultrasound, magnetic resonance imaging, or computed tomography) and biochemistry (AFP and liver function test).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Conflict of Interest: no conflict of interest was declared by the authors. Financial disclosure: the authors declared that this study was funded by Scientific Research Projects Office of Gazi University.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Erdal 2016 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ertle 2013

Study characteristics

Ertle 2013 *(Continued)*

Patient Sampling	We performed a prospective single-centre study including 164 cases of HCC-patients and 422 controls seen between 02/2007 and 11/2008. 10 patients had to be excluded due to pregnancy (n = 2), warfarin use (n = 4), or missing data (n = 4). Age range not reported. Males 56%
Patient characteristics and setting	
Index tests	Serum concentrations of AFP and ® DCP were determined using the Wako LiBASys ® clinical auto-analyser by a liquid-phase binding assay [17]. Interassay coefficient of variation for total AFP concentration ranges from 2.6% to 4.6%. The analytical limit of detection is 0.8 ng/mlL and the assay is linear up to 1,000 ng/mL AFP concentration.
Target condition and reference standard(s)	HCC was verified by histological findings or by two different cross-sectional scans as defined by the European Association for the Study of the Liver (EASL) guidelines. Controls consisted of patients with viral hepatitis, cirrhosis, other chronic liver diseases such as nonalcoholic steatohepatitis (NASH), autoimmune hepatitis (AIH), and others. Liver diseases were classified according to clinical, serological, and histological criteria. Liver cirrhosis was diagnosed by histology or typical findings such as portal hypertension in known chronic liver diseases.
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Ertle 2013 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ette 2015

Study characteristics	
Patient Sampling	This was a cross-sectional case–control study. Patients were drawn from referrals to the Liver Unit at the Obafemi Awolowo University Teaching Hospitals Complex, Ile-Ife, from April 2011 to March 2012. The patients were divided into two broad groups: HCC and non-HCC groups. 62 consecutive patients presenting with untreated primary hepatocellular carcinoma. The controls were 57 patients with benign hepatic diseases which comprised of 34 patients with chronic hepatitis B infection, 1 patient with chronic hepatitis C infection, 21 patients with compensated cirrhosis of the liver, and 1 patient with Non-alcoholic Fatty Liver Disease. Patients with a history of the use of warfarin or other dicoumarol or total bilirubin level above 20 mg/dL (340 µmol/L) were excluded from the study. Age range not reported. Males 72%
Patient characteristics and setting	

Ette 2015 *(Continued)*

Index tests	AFP was tested using commercially available immunoenzymometric assay kit manufactured by INTECO Diagnostics, UK Ltd., London with no predefined cut-off value.
Target condition and reference standard(s)	62 consecutive patients with HCC were enrolled. Patients presented with untreated primary HCC diagnosed using the European Association for the Study of Liver Diseases (EASL) and American Association for the Study of Liver Diseases (AASLD) criteria. All the cases and controls were subjected to abdominal ultrasound while CT scan was restricted to those shown on ultrasound to have focal lesions in the liver.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Competing interests: the authors declared no conflict of interests.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Ette 2015 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ezzikouri 2015

Study characteristics	
Patient Sampling	The participants were prospectively enrolled at the Tokyo Metropolitan Komagome Hospital, Showa University Fujigaoka Hospital, and Kanazawa University Hospital, Japan. 651 serum samples, collected from September 2007 to November 2014, were obtained from 395 HCV-positive patients, including 133 moderate chronic hepatitis C (CHC), 85 liver cirrhosis (LCC), and 177 HCC (HCC-C) patients; 232 HBV patients, including 103 chronic HBV (CHB), 56 liver cirrhosis (LCB), and 73 HCC (HCC-B) patients; and 24 healthy controls. Age range not reported. Males 47%
Patient characteristics and setting	
Index tests	Serum AFP concentration, cut-off value 20 ng/mL
Target condition and reference standard(s)	HCC was diagnosed by ultrasonography and computed tomography, and confirmed by liver biopsy. Liver cirrhosis was by presence of ascites and/or gastroesophageal varices, and defined by the aspartate transaminase (AST) to platelet ratio index (APRI) and Fibrosis-4 index.
Flow and timing	No information on interval between index test and reference standard
Comparative	

Ezzikouri 2015 *(Continued)*

Notes	Authors have no conflict of interest. This work was supported by grants from the Ministry of Health Science H24-B-014, H25-009 and Welfare and the Ministry of Education, Science and Culture, Japan, 23590547. The funders of the study had no role in the study design, data collection, analysis, interpretation, or writing of the paper. All authors had access to the raw data.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Ezzikouri 2015 *(Continued)*

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Fabris 1991

Study characteristics			
Patient Sampling	A total of 238 patients were considered in this study. 211 had liver cirrhosis. Age range: 19-86. Males 60%		
Patient characteristics and setting			
Index tests	Serum AFP measurement, no specification. Cut-off value 20 ng/mL		
Target condition and reference standard(s)	The diagnosis of liver cirrhosis was established on laparoscopic and histologic findings specific for the disease. The diagnosis of HCC was established on previously reported criteria: histologic evaluation (n = 16), or specific findings of one or more of the following investigations: an AFP level greater than 1000 pg/L with the ultrasonography, computed axial tomography, and angiography.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Cconflicts of interest not reported		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Fabris 1991 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Fabris 1991 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Fang 2010

Study characteristics			
Patient Sampling	A total of 273 people with HBV-related HCC (n = 145) and fibrosis (n = 128) were recruited during 2007–2008. Patients with HBV-related HCC were recruited from Eastern Hepatobiliary Hospital, Shanghai. Excluded from the patients cohort were patients with hepatitis A virus, HCV, hepatitis D virus, hepatitis E virus, human immunodeficiency virus, Epstein-Barr virus, and cytomegalovirus infection, alcohol consumption > 30 g/day, metastatic liver cancer, autoimmune liver disease, drug-related liver disease, alcoholic hepatitis, obstructive jaundice, other causes of chronic liver disease, renal inadequacy, or blood diseases and insufficient biopsy samples. Age range not reported. Males 87.5%		
Patient characteristics and setting			
Index tests	Serum AFP concentration measurement determined on Roche E170 modular with matched reagents (Roche E170, Germany). The reference values were 0 ng/mL -- 20 ng/mL.		
Target condition and reference standard(s)	All the enrolled HCC were confirmed by histological study after surgical resection. The diagnosis of liver fibrosis was confirmed with percutaneous liver biopsy and histological study, independently inspected by two pathologists.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Grant sponsor: Natural Science Foundation of China		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Fang 2010 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Farid 2014

Study characteristics	
Patient Sampling	This cross-sectional study was conducted on 60 patients. All patients presented to the Endemic Medicine Department, Kasr Al Ainy Hospital, Cairo University, during the period between January 2011 and July 2011. Patients were classified into three groups:

Farid 2014 *(Continued)*

	Control group: 20 apparently healthy people without any evidence of liver disease. Liver cirrhosis (LC) group: 20 patients with HCV-related LC. Cirrhosis was diagnosed based on clinical background (manifestations of liver cell failure), laboratory background (markers of cytolysis, cholestasis, and synthetic function derangement), and imaging (morphological changes and signs of portal hypertension). HCC group: 20 patients with HCC in addition to LC. Patients who received prior interferon therapy, or immunosuppressive therapy, or those who received therapy for HCC lesion or recurrent HCC were excluded from the study. Age range: 28-80. Males 79%
Patient characteristics and setting	
Index tests	The serum AFP level was determined using the enzyme-linked immunosorbent assay (ELISA) (Monobind Inc., Lake Forest, CA, 92630, USA) with a cut-off level of 10ng/mL.
Target condition and reference standard(s)	Liver masses (in HCC group) were diagnosed as HCC according to the Korean Liver Cancer Study Group. - AFP value > 200 ng/mL with a specific imaging pattern, defined by intense contrast uptake during the arterial phase followed by contrast washout during venous or delayed phases in contrast-enhanced study such as computed tomography (CT) scan or magnetic resonance imaging (MRI). - AFP value < 200 ng/mL with two or more positive findings of dynamic contrast enhancement (CT or MRI). - A tumour of 2-cm size with typical characteristics of HCC in dynamic contrast enhancement CT or MRI regardless of the serum AFP levels.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared that there was no conflict of interest. This work has been partly self funded and partly funded by the University without any organisational support and without any interest to the contributing authors.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Farid 2014 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Feng 2016

Study characteristics	
Patient Sampling	A total of 700 patients who were diagnosed with a hepatopancreatobiliary disease and who had undergone surgery were consecutively enrolled from the Institute of Hepatobiliary Surgery, Southwest Hospital, Third Military Medical University, between September 2008 and September 2011. Participants receiving surgery were divided into two groups:

Feng 2016 *(Continued)*

	(i) participants with HCC (n = 329); and (ii) participants without HCC, with either benign or malignant hepatopancreatobiliary disease (n = 371). None of the participants were receiving and/or had received vitamin K therapy. Age range not reported. Males 69%
Patient characteristics and setting	
Index tests	Serum AFP was examined, using a commercially available immunometric assay (ST AIA-PACK AFP, Tosoh, Tokyo, Japan), with enhanced chemiluminescence at the Southwest Hospital Clinical Diagnostic Center with no predefined cut-off value.
Target condition and reference standard(s)	HCC was clinically diagnosed, based on international guidelines. Pathologic findings from resected specimens were confirmed for all 329 participants. No person had received any previous therapy to treat HCC such as TACE, RFA, PEI, or resection, and people underwent surgery for the first time at this hospital. A cohort of people with a hepatopancreatobiliary disease other than HCC, based on enhanced imaging findings, who were undergoing surgery at this hospital, were used. All of these people had been diagnosed by clinical findings as well as pathologic findings.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Competing interests. The authors declared that they had no competing interests.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		

Feng 2016 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Fujii1995

Study characteristics	
Patient Sampling	During the two years, from January 1989 to December 1990, 50 patients with hepatocellular carcinoma plus liver cirrhosis (37 males and 13 females) and 50 patients with liver cirrhosis alone (37 males and 13 females) were included. Age range not reported. Males 74%
Patient characteristics and setting	
Index tests	The serum AFP was measured using an a-Fetoprotein Radioimmunoassay Kit (Dinabot Laboratories, Tokyo, Japan), and its normal level is less than 20 ng/mL.
Target condition and reference standard(s)	In 37 out of the 50 hepatocellular carcinoma patients, diagnosis was made by histological examination (biopsy and necropsy); in the remainder, it was based on markedly elevated serum AFP lev-

Fujii1995 *(Continued)*

	els (400 ng/mL), space-occupied lesions demonstrable by various imaging techniques, and typical computed tomographic and/or angiographic findings [7-10]. In the group of patients with liver cirrhosis, the diagnosis was made on liver biopsy in 40 patients, and on clinical, biochemical, and computed tomographical findings in the remaining 10 patients.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Fujii1995 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Gad 2005

Study characteristics	
Patient Sampling	Quote: "We conducted a cross-sectional study between October 2001 and November 2002. Data were gathered from two affiliations: Shinshu University (Japan) and Suez Canal University (Egypt) Hospitals. A total of 334 consecutive patients with chronic liver disease seen at outpatient liver clinics in the two settings (who met our inclusion/exclusion criteria) were included; of them, 110 patients were diagnosed as HCC. We excluded patients with alcoholic and schistosomal liver diseases from our study populations. We had also excluded patients known from their medical history to have interstitial lung fibrosis, or any other lung disease from our study population." Age range not reported. Males 77%
Patient characteristics and setting	
Index tests	Assessment of alpha foetoprotein (AFP) and protein-induced vitamin K deficiency or absence (PIVKA-II) was performed using commercially available kits. Cut-off points were set at 10 ng/mL for AFP.
Target condition and reference standard(s)	Chronic liver disease and cirrhosis were identified and diagnosed according to liver biopsy findings, clinical and/or radiological evidence of portal hypertension. HCC was excluded by imaging studies (abdominal ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), and/or hepatic angiography), one of which must have been performed at least six months following the measurement of AFP. HCC was diagnosed when meeting the study inclusion criteria of positive cytology and/or histology or by the presence of characteristic hepatic masses on liver CT, MRI, and/or hepatic angiography (i.e. enlarging tumours and/or tumours with typical arterial vascularisation).
Flow and timing	HCC was excluded by imaging studies (abdominal ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), and/or hepatic angiography), one of which was to be performed at least six months following the measurement of AFP.
Comparative	

Gad 2005 *(Continued)*

Notes	Acknowledgments: "We would like to thank Takeda Foundation, Osaka, Japan for their financial support."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Gad 2005 *(Continued)*

Was there an appropriate interval between index test and reference standard?	No
Did all patients receive the same reference standard?	Unclear
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Gambarin-Gelwan 2000

Study characteristics			
Patient Sampling	We analysed retrospectively the charts of 106 consecutive adult patients who underwent OLT for treatment of cirrhosis over a 1-year period at Mount Sinai Hospital. All patients had US, CT, and serum AFP measurements within 6 months of OLT. The results were compared to explant histology. Age range: 24-71. Males 65%		
Patient characteristics and setting			
Index tests	QUOTE: "US: US exams were performed using an ATL UM-9 (Advanced Technology Laboratories, Bothell, WA). Depending upon the patient's physique, either a 2.25-MHz or a 2- to 4-MHz broad bandwidth transducer was used. AFP: AFP < 20 ng/mL, the upper limit of normal at our institution, was defined as low risk. AFP > 20 ng/mL was defined as high risk."		
Target condition and reference standard(s)	Explant histology: a pathologist specialising in the hepatobiliary system reviewed all liver explants. Each liver explant was sectioned every 1 cm. The presence of tumour nodules, their size, and their location were recorded. The underlying liver pathology was evaluated.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No data on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Gambarin-Gelwan 2000 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Gambarin-Gelwan 2000 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Gani 2015

Study characteristics	
Patient Sampling	This is a cross-sectional study. The participants in this study were liver cirrhotic patients, aged 18 years and older. Age range not reported. Males 55%
Patient characteristics and setting	Among 106 patients, 59 patients had cirrhosis with HCC, and the other 47 patients had cirrhosis without HCC as negative control to HCC group.
Index tests	The quantitative measurement of plasma AFP was performed using ADVIA Centaur AFP assay, a two-site sandwich immunoassay using direct chemiluminometric technology, which uses constant amounts of two antibodies. The results were reported in ng/mL, with a cut-off value of ≤15 ng/mL.
Target condition and reference standard(s)	Diagnosis of hepatocellular carcinoma in the patient group were defined according to AASLD guidelines on hepatocellular carcinoma or by presence of liver nodule, AFP > 200 ng/mL and supported with two imaging results with typical features of hepatocellular carcinoma.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest disclosure

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Gani 2015 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Garretti 1988

Study characteristics	
Patient Sampling	Among 600 consecutive people with cirrhosis, 64 HCC (10%) were identified.

Garretti 1988 *(Continued)*

	Age range: 29-75. Males not reported		
Patient characteristics and setting			
Index tests	All patients underwent US evaluation		
Target condition and reference standard(s)	CT scan was performed both in people with positive US and in people with negative case when there was a clinical suspect of HCC (increased AFP levels or hepatic decompensation).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest disclosure		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Garretti 1988 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Ge 2015

Study characteristics	
Patient Sampling	In this study, we evaluated the diagnostic capability of the combination of AFP with two novel potential biomarkers, Dickkopf-1 (DKK1) and osteopontin (OPN), for HCC in 390 participants including 89 people with HCC, 36 people with liver cirrhosis, 65 people with chronic hepatitis B, and 200 healthy controls. Age range and % of males not reported.
Patient characteristics and setting	The HCC patients and healthy controls enrolled in this study were collected from December 2008 to June 2009 and from May to June, 2013, respectively, from the Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, China.
Index tests	Concentrations of serum AFP was measured by the same method with another commercial kit (Raygene Biotechnology Company, Shanghai, China). The assays were conducted according to the manufacturer's instructions, and all specimens were performed blindly and in duplicate.
Target condition and reference standard(s)	The diagnosis of HCC was based on American Association for the Study of Liver Diseases (AASLD) Practice Guidelines, verified by ultrasound, CT scan, or MRI and biochemistry (AFP serology and liver function enzymes) findings, and was confirmed by histopathology.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Ge 2015 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Ge 2015 *(Continued)*

Could the patient flow have introduced bias?	Low risk

Gentile 2017

Study characteristics			
Patient Sampling	160 patients were enrolled, 56 cases and 104 controls. Age range not reported. Males 73%		
Patient characteristics and setting			
Index tests	Blood samples for AFP and PIVKA-II assays were also collected from all the enrolled patients, regardless if they belonged in the case or in the control group. PIVKA-II assay was performed using Lumipulse® G1200 (Fujirebio Inc., Malvern, PA, USA), an enzyme-linked immunoassay, based on chemiluminescence principles (CLEIA – chemiluminescent enzyme immunoassay) with no predefined cut-off value.		
Target condition and reference standard(s)	Cases: proven HCV infection (positive anti-HCV and detectable serum HCV-RNA) plus radiological, histological, or cytological evidence of hepatocellular carcinoma as assessed in American Association for the Study of Liver Diseases (AASLD) hepatocellular carcinoma practice guidelines.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest declared		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Gentile 2017 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Giannelli 2005

Study characteristics	
Patient Sampling	120 patients with HCC and 90 patients with liver cirrhosis Age range 26-85. Males 74%
Patient characteristics and setting	Quote: "Serum samples from 120 patients with HCC (95 men and 25 women, aged 40–84 years) were included in our study. Serum samples from 90 patients affected by cirrhosis were also collected. Serum samples from 41 healthy people (17 men and 24 women, aged 24–45 years) were collected as controls and stored as above."
Index tests	Serum a-FP was measured using an ELISA kit (IMMULITE 2000) based on a solid-phase 2-site sequential chemiluminescent immunometric technique purchased from Diagnostic Products (Los Angeles, CA).

Giannelli 2005 *(Continued)*

Target condition and reference standard(s)	HCC diagnosis was confirmed by ultrasound and, when necessary, by CT.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest disclosure

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Giannelli 2005 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		High risk	

Giannelli 2007

Study characteristics			
Patient Sampling	Between 2001–2005, 961 consecutive patients were observed from two European hospital centres. Inclusion criteria were: age over 18 years and presence of HCC or liver cirrhosis (LC), and exclusion criteria were other concomitant cancers. 499 were classified as HCC according to the EASL Barcelona conference criteria. The remaining 462 patients were classified as LC according to clinical and biochemical parameters. Age range not reported. Males 90%		
Patient characteristics and setting			
Index tests	AFP: AFP and SCCA were measured using ELISA kits purchased from Diagnostic Products (Los Angeles, CA) and from Xeptagen (Naples, Italy), respectively, and following the manufacturer's instructions as previously described. The diagnostic cut-off and the related sensitivity, specificity and 95% confidence intervals were determined.		
Target condition and reference standard(s)	HCC: 499 were classified as HCC according to the EASL Barcelona conference criteria. All these patients underwent US and CT scans, while liver biopsy was performed in 380/499 patients (76%).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest disclosure		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Giannelli 2007 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Gopal 2014

Study characteristics	
Patient Sampling	This study was reported as a retrospective case-control study of cirrhotic patients with and without HCC. They included all patients diagnosed with HCC at Parkland Hospital between January 2005 and June 2012. Patients were identified by a combination of International Classification of Diseases, 9th revision, codes for HCC (155.0 or 155.2), a prospectively maintained list of patients seen in a multidisciplinary liver tumour clinic, and tumour conference presentation lists. In the HCC group, they excluded patients who did not have an AFP level before HCC diagnosis, and in the control group (patients with cirrhosis), they excluded patients with any suspicious liver mass on imaging and those who did not have an AFP test during the study period (January 2010–July 2011). Age range: 49-61. Males 71%
Patient characteristics and setting	
Index tests	AFP. Quote: "We dichotomized AFP at a cut-off value of 20 ng/mL because this is the most commonly reported and used cut-off value in clinical practice."
Target condition and reference standard(s)	HCC: Quote: "Two authors (A.G.S. andA.C.Y.) adjudicated all HCC cases to confirm that they met diagnostic criteria, based on AASLD guidelines". Cirrhosis: "Patients initially were identified using a previously validated combination of International Classification of Diseases, 9th revision, codes. Patients were required to have at least 6 months of follow-up evaluation to confirm the absence of HCC."
Flow and timing	No data on interval between index test and reference standard. Quote: "Between January 2005 and June 2012, there were 457 patients with cirrhosis who were diagnosed with HCC. We excluded 5 patients who did not have an AFP level before HCC diagnosis. Between January 2010 and July 2011, there were 914 patients with cirrhosis who were seen in an outpatient setting at Parkland Hospital, of whom 238 patients were excluded for a lack of AFP level or insufficient follow-up duration."
Comparative	
Notes	Conflicts of interest: the authors disclosed no conflicts.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Gopal 2014 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Grazi 1995

Study characteristics	
Patient Sampling	227 patients were included in this retrospective study; 111 had HCC, and 85 of these were also with liver cirrhosis. The remaining 116 patients, defined as the control group, included 23 patients

Grazi 1995 *(Continued)*

	with liver metastases from colorectal cancer, 26 with benign hepatic lesions, 20 with tumours other than HCC without hepatic metastases, and 47 with other liver diseases. Age range:15 to 74 yeas. Males 88%
Patient characteristics and setting	
Index tests	The assays for AFP (AFP Reagen Pack, Abbott, North Chicago, IL, USA), were carried out at the Central Laboratory of the S. Orsola Hospital. The cut-off value was 20 ng/mL.
Target condition and reference standard(s)	The diagnosis of HCC was confirmed histologically in all cases.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest disclosure

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Grazi 1995 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Guan 2020

Study characteristics			
Patient Sampling	A total of 581 cases of serum samples including 302 cases of HCC, 105 cases of liver cirrhosis, 59 cases of chronic hepatitis B (CHB), and 115 cases of healthy controls. Age range not reported. Males 63%		
Patient characteristics and setting			
Index tests	Alpha-foetoprotein (AFP) was measured using standard methods and matched reagents (HITACHI 7600, Hitachi Koki Co. Ltd., Hitachinaka City, Japan. No pre-definition of cut-off value		
Target condition and reference standard(s)	HCC was confirmed by histological study after surgical resection. The diagnosis of each patient was confirmed by laboratory, pathological, and imageological examination.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors declared that there were no conflicts of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Guan 2020 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Guan 2020 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Hallager 2018

Study characteristics			
Patient Sampling	Patients enrolled in DANHEP before 31 December 2012 were eligible for inclusion if they fulfilled the following criteria: (i) a positive HCV-RNA test, (ii) a valid PIN and address recorded in the Danish Civil Registration System, (iii) ≥ 18 years of age, and (iv) cirrhosis before 31 December 2013. Age range and % of males not reported		
Patient characteristics and setting	1075 patients with CHC and cirrhosis at risk of HCC were enrolled.		
Index tests	AFP measurement with a cut-off value of 20 ng/mL		
Target condition and reference standard(s)	SNOMED and ICD-codes used in the definition of all inclusion criteria, outcomes and covariates are provided in the supplementary material.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest declared		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Hallager 2018 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Han 2014

Study characteristics

Han 2014 *(Continued)*

Patient Sampling	160 HBV-infected patients with HCC, 88 CHB patients without HCC from April 2012 to April 2013 in the Department of Hepatology, Qilu Hospital of Shandong University. Exclusion criteria included co-infection with human immunodeficiency virus (HIV) or hepatitis C virus (HCV), alcoholic liver diseases, autoimmune liver diseases, non-alcoholic fatty liver diseases (NAFLD), and other causes of chronic liver diseases. Age range: 46-61. Males 77%.
Patient characteristics and setting	The current study enrolled a total of 84 patients with HBV/HCV-related HCC (69 males and 15 females), 74 patients with HBV/HCV-associated liver cirrhosis (42 males and 32 females), and 29 patients with chronic hepatitis B/C (14 males and 15 females).
Index tests	AFP was also measured by an automatic analyser (COBAS e 601, Roche Diagnostics, Germany). Cut-off value 20 ng/mL
Target condition and reference standard(s)	HCC patients were diagnosed according to the 2010 update of the American Association for the Study of Liver Diseases (AASLD) Practice Guidelines for Management of hepatocellular carcinoma. Chronic HBV infection was defined as a positive hepatitis B surface antigen (HBsAg) for at least 6 months prior to the beginning of this study. Within all the 88 CHB patients, 33 were accompanied by cirrhosis.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared that no competing interest existed.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		

Han 2014 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Han 2018

Study characteristics	
Patient Sampling	The current study enrolled a total of 84 patients with HBV/HCV-related HCC (69 males and 15 females), 74 patients with HBV/HCV-associated liver cirrhosis (42 males and 32 females), and 29 patients with chronic hepatitis B/C (14 males and 15 females). These patients were admitted to Shengjing Hospital of China Medical University (Shenyang, China) from September 2012 to October 2014. Age range: 28-78. Males 82%.
Patient characteristics and setting	
Index tests	Serum AFP with no predefined cut-off value
Target condition and reference standard(s)	The diagnostic criteria employed were based on the guidelines for the prevention and treatment for chronic HBV (2010 version) (39) and diagnosis, management, and treatment of HCC (2011) (1).

Han 2018 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared that they had no competing interests.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Han 2018 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Hu 2018

Study characteristics			
Patient Sampling	Patients diagnosed with HCC and liver disease were enrolled at the three centres (Peking University 1st Hospital, Xi'an Jiaotong University 1st Hospital and The Second Hospital of Nanjing, Affiliated to Medical School of Southeast University) between July 2013 and July 2016. HCC was diagnosed according to the Asian Pacific Association for the Study of the Liver (APASL) consensus recommendations on HCC. Only newly diagnosed and treatment-naive patients with HCC were enrolled. Liver disease samples were mainly from patients infected with hepatitis B virus (HBV) or hepatitis C virus (HCV) and include samples from patients with hepatitis and cirrhosis, which were diagnosed according to APASL guideline. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum AFP concentration: no specification. Predefined cut-off value 20 ng/mL		
Target condition and reference standard(s)	The Asian Pacific Association for the Study of the Liver (APASL) HCC guidelines Control: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Yijie Zheng is employee of Abbott Laboratories. The other authors declared that they had no competing interests.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Hu 2018 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Hu 2018 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Hu 2019

Study characteristics			
Patient Sampling	A total of 565 patients with pathologically diagnosed HCC were enrolled in this study. Quote: "Patients in HCC group need to meet the following 5 criteria to be included: Barcelona clinic liver cancer (BCLC) stages A, B, or C; Edmondson–Steiner grades I, II, or III; Child-Pugh grades A, B, or C. Patients will be excluded if they meet one of the following: past history of HCC; blood system diseases; immune-related diseases; organic disease outside liver; presence of other types of cancers." The control group comprised 441 age- and sex-matched individuals diagnosed with cirrhosis. Age range not reported. Males 86%.		
Patient characteristics and setting			
Index tests	Serum AFP: no specification. No predefinition of the cut-off value		
Target condition and reference standard(s)	HCC pathologically diagnosed as HCC; surgical resection treatment Cirrhosis - control: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Hu 2019 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Hu 2019 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Huo 2007

Study characteristics			
Patient Sampling	Quote: "Between April 1996 and January 2001, 248 consecutive patients who underwent curative surgical resection for HCC in our institution were included as the index patients in this study." Age range not reported. Males 73%		
Patient characteristics and setting	Quote: "Between April 1996 and January 2001, 248 consecutive patients who underwent curative surgical resection for HCC in our institution were included as the index patients in this study. Their clinical and pathological profiles were prospectively collected and retrospectively analysed."		
Index tests	Serum AFP levels were measured by using a radioimmunoassay kit (ELSA2-AFP, CIS, Cedex, France) at the time of diagnosis.		
Target condition and reference standard(s)			
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest declared		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		

Huo 2007 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Ibrahim 2013

Study characteristics	
Patient Sampling	105 patients were included, 70 of them were diagnosed with HCC.

Ibrahim 2013 *(Continued)*

	Age range and % of males not reported
Patient characteristics and setting	105 patients were included, 70 of them were diagnosed as HCC. These patients were divided into group 1 early stage HCC (n = 37; male = 23, female =14), and group 2 late stage HCC (n = 43; male = 29, female = 14), and the other 35 patients (Group 3) were with cirrhosis (20 males and 15 females). In addition, there were about 20 healthy controls (Group 4).
Index tests	AFP and abdominal ultrasonography. Computed tomography (CT) was done for patients groups only, except when the previous investigations suggested a possible diagnosis of HCC. Liver biopsy was done to prove the diagnosis of hepatocellular carcinoma (group 3 only). Although the diagnosis of HCC was confirmed by any of the following criteria: histological evidence, demonstration of a focal lesion > 2 cm in size and with arterial hypervascularisation by two imaging techniques showing this morphological aspect with an AFP level of ≥ 400 ng/mL, the study research depended on histological evidence through liver biopsy as a golden standard test to confirm the diagnosis of all HCC patients (Group 1 and Group 2).
Target condition and reference standard(s)	Details on the diagnosis of HCC not provided
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest disclosed

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Ibrahim 2013 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Iizuka 2010a

Study characteristics	
Patient Sampling	The abilities of quantitative analyses of 7 genes hypermethylation in serum DNA, α-fetoprotein (AFP) and prothrombin-induced vitamin K absence II (PIVKA-II), and various combinations to detect HCC were evaluated in a training cohort of 164 HCV-infected patients (108 HCCs; 56 non-HCCs). Age range not reported. Males 69%
Patient characteristics and setting	"Our training cohort (Table 1) included 164 patients positive for HCV antibody, all of whom were treated at Yamaguchi University Hospital between May 1998 and April 2006, and were subjected to analyses of AFP and PIVKA-II, routine radiography, US, computed tomography (CT), magnetic resonance imaging (MRI), and, if necessary, hepatic angiography, dynamic CT, or dynamic MRI before and after treatment."
Index tests	AFP with a cut-off value of 20 ng/mL
Target condition and reference standard(s)	"On the basis of those imaging techniques, 108 of the 164 patients were diagnosed with HCC. Subsequently, 95 of these 108 patients (88.0%) bearing HCC underwent hepatic surgery or biopsy; and all tumours from the 95 patients were pathologically confirmed as HCC. Moreover, none of the 108 HCC patients showed any other malignancies at enrolment. We confirmed that none of the

Iizuka 2010a *(Continued)*

	remaining 56 patients developed HCC during the follow-up period of more than 2 years. Using the results of imaging techniques and pathological examinations, we judged that 79 of the 164 patients (48.2%) had liver cirrhosis."
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest disclosure. Grant sponsors: the Ministry of Education, Culture, Sports, Science and Technology (No. 18390366, No. 17591406 and Knowledge Cluster Initiative); the Venture Business Laboratory of Yamaguchi University; the New Energy and Industrial Technology Development Organization (Grant number: 03A02018a)

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Iizuka 2010a *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Iizuka 2010b

Study characteristics	
Patient Sampling	"The abilities of quantitative analyses of 7 genes hypermethylation in serum DNA, α-fetoprotein (AFP) and prothrombin-induced vitamin K absence II (PIVKA-II), and various combinations to detect HCC were evaluated in a validation cohort comprised 262 consecutive HCV-infected patients who were enrolled in 4 distinct institutes between May 2006 and April 2008. Out of the 262 patients, 1 was excluded due to daily intake of warfarin, which may affect serum levels of PIVKA-II, and 3 were excluded because of small amounts of extracted cell-free DNA (cfDNA)." Age range not reported. Males 69%
Patient characteristics and setting	
Index tests	AFP with a cut-off value of 20 ng/mL
Target condition and reference standard(s)	The detection program for HCC in individual institutes was performed according to the nationwide follow-up survey conducted by the Liver Cancer Study Group of Japan (LCSGJ). On the basis of findings from multiple imaging modalities (US, CT, MRI, hepatic angiography, dynamic CT, and dynamic MRI), hepatologists from the individual institutes diagnosed 112 of the 258 patients (43.4%) as HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	

Iizuka 2010b *(Continued)*

Notes	Grant sponsors: the Ministry of Education, Culture, Sports, Science and Technology (No. 18390366, No. 17591406 and Knowledge Cluster Initiative); the Venture Business Laboratory of Yamaguchi University; the New Energy and Industrial Technology Development Organization (Grant number: 03A02018a). No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			

Iizuka 2010b *(Continued)*

If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ishii 2000

Study characteristics	
Patient Sampling	This prospectively designed, cooperative study was performed from November 1992 to March 1994. Patients previously diagnosed to have chronic hepatitis or liver cirrhosis were registered consecutively in this study if the following criteria were satisfied: 1) HCC was not detected by ultrasonography at the time of entry; and 2) patients agreed to close follow-up for more and/or equal to 1 year, and had already been followed for 6 months before entry. Patients were excluded if they had lack of sufficient clinical data and because of loss to follow-ups during the observation period. Age range: 24-84. % of males not reported
Patient characteristics and setting	
Index tests	AFP. Levels of AFP were measured by an enzyme immunoassay with an anti-AFP (Abott AFP-EIA kit, Dainabott Laboratory, Tokyo, Japan). No pre-specified threshold.

Ishii 2000 *(Continued)*

Target condition and reference standard(s)	HCC: once HCC was suspected through abdominal ultrasonography or by serum AFP and PIVKA-II levels, CT with contrast medium and/or hepatic angiography were performed to establish the diagnosis of HCC. In a few cases in which the diagnosis of HCC was still equivocal. Despite the CT and hepatic angiography, percutaneous liver biopsy was performed.
Flow and timing	Out of 918 patients, 184 were excluded for lack of sufficient clinical data (61 patients) and because of loss to follow-up during the observation period (123 patients with < 1 year of follow-up). No data on interval between index test and reference standard
Comparative	
Notes	No conflict of interest disclosure. Only acknowledgment quote: "The authors thank Ezai Industries Inc. (Tokyo, Japan) for measuring PIVKA-II levels and gratefully acknowledge the help of Mr. N. Magario (Eizai Industries Inc.)."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Ishii 2000 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Ismail 2017a

Study characteristics	
Patient Sampling	A case-control study of 305 cases was conducted between 2012 and 2014; 128 Egyptian (E) participants were enrolled from the National Cancer Institute (NCI), Cairo University, and 177 Saudi (S) participants who were enrolled from King Abdullah Medical City, Holy Makkah. Age range not reported. Males 67%
Patient characteristics and setting	Group I: a total of 57 healthy volunteers as normal control. Group II: a total of 62 people designated as the cancer control group who had malignancies of the gastrointestinal system other than HCC: 41 colorectal carcinomas, 8 pancreatic cancers, 7 stomach cancers, 4 bile duct carcinomas, and 2 peritoneal neoplasms. Group III: 21 cases with benign hepatic lesions: 11 haemangiomas, 8 focal nodular hyperplasias, 1 hepatocellular adenoma, and 1 hepatic cyst. All cases proved to be free from malignant liver disease by imaging and fine needle biopsy. Group IV: a total of 99 chronic viral hepatitis cases: 34 HBV, 60 HCV, and 5 combined HBV and HCV. On the basis of the calculated APRI, 41 (41.1%) cases had low APRI indicating absence of advanced fibrosis or cirrhosis, 36 (36.4%) had APRI values indicating advanced fibrosis, and 22 (22.2%) had high values indicating cirrhosis. Group V: 66 HCC cases. The diagnosis of HCC was based on histopathology. If histopathology was not available, diagnosis was based on two imaging modalities; MRI, CT, or contrast- enhanced ultrasound showing an enhancing vascular mass of more than 2 cm.
Index tests	Serum AFP (AxSYM, Abbott Laboratories) with a cut-off value of 20 ng/mL

Ismail 2017a *(Continued)*

Target condition and reference standard(s)	The diagnosis of HCC was based on histopathology; if histopathology was not available, diagnosis was based on two imaging modalities; magnetic resonance imaging, computed tomography or contrast-enhanced ultrasound showing an enhancing vascular mass of more than 2 cm.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Ismail 2017a *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Ismail 2017b

Study characteristics	
Patient Sampling	The studied patients and controls were divided into the following groups: Group I (GI): included 30 patients with liver cirrhosis Group II (GII): included 30 cirrhotic patients with HCC fulfilling HCC criteria on tri-phasic CT scan Group III (GIII): included 30 healthy individuals Exclusion criteria: a past history or evidence of other malignancies, autoimmune disorders, organ failure, and other causes of cirrhosis (e.g. alcoholic and non-alcoholic fatty liver diseases). Age range not reported. Males 55%
Patient characteristics and setting	
Index tests	Serum AFP: no specification. No definition of the cut-off value
Target condition and reference standard(s)	All studied groups were subjected to abdominal ultrasound, while GI and GII were subjected to Tri-phasic CT scan abdomen.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Ismail 2017b *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Ismail 2017b *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Iyer 2018

Study characteristics			
Patient Sampling	A retrospective observational study with data analysis from all patients with diagnosis of HCC or those with chronic liver disease based on standard clinical, biochemical and US criteria with clinical suspicion of HCC. The control group included patients who had cirrhosis and who were under regular bi-anual surveillance with US and AFP. Age range: 44-69. Males 78%.		
Patient characteristics and setting			
Index tests	AFP. The study authors had taken two cut-off values for suspecting HCC in cirrhotic patients based on previous studies in India: 16 ng/mL and 200 ng/mL.		
Target condition and reference standard(s)	HCC: CT Control group: no criteria mentioned		
Flow and timing	No data on interval between index test and reference standard		
Comparative			
Notes	No data on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Iyer 2018 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Izzo 1999

Study characteristics	
Patient Sampling	From 1993 to 1997 patients all patients with chronic hepatitis B or C virus infections of at least 5-year duration were offered a screening programme to detect HCC. Patients with Child class B or C cirrhosis, a history of hepatic encephalopathy, bleeding gastroesophageal varices, ascites, or a prior diagnosis of any type of malignancy were excluded from the study. The diagnostic sensitivity and specificity of serum sIL-2R levels for the 457 patients with histologically severe liver injury were 99.0% and 95.6%, respectively, compared with 80.0% and 94.7% for serum AFP levels. Age range: 29-80. Males 61%
Patient characteristics and setting	
Index tests	AFP with a prespecified cut-off value of 10 ng/mL
Target condition and reference standard(s)	HCC: when a mass lesion in the liver was detected by ultrasonography, the serum AFP level exceeded 10 ng/mL, and/or the serum sIL-2R level exceeded 850 U/mL, further diagnostic evaluation was performed using CT scanning (with intravenously administered bolus contrast agent) or MRI of the abdomen. Confirmed liver tumours were biopsied under ultrasonographic or CT guidance, and the histological diagnosis of HCC was based on routine hematoxylin and eosin staining. Patients without HCC: follow-up after negative ultrasound findings
Flow and timing	Out of 1520 patients analysed, diagnostic sensitivity and specificity of AFP levels are given for a subgroup of patients with histologically-severe liver injury (457 patients). No data given on interval between index test and reference standard All CLD patients, including HCC received biopsy.
Comparative	
Notes	No conflicts of interest declared

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Izzo 1999 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Jalli 2015

Study characteristics	
Patient Sampling	This study included 96 cirrhotic patients who were referred to the gastroenterologist for follow-up. 30 of them had concomitant hepatocellular carcinoma (HCC) proved by pathology, and were selected. Non-cooperative cases, severe ascites, and contraindications for MRI were excluded from the study. Age range and % of males not reported

Jalli 2015 *(Continued)*

Patient characteristics and setting	
Index tests	US: US of the liver was performed for each patient by Logic 7, GE, USA, ultrasound machine, with a 3.5 MHz curve transducer and 7.5 MHz linear probe for surface evaluation. US was done by a radiologist with 10 years of experience in abdominal US. He determined whether the lesion suspected of HCC existed or not. Radiologists were blinded for definite diagnosis of the patients. US criteria for lesion assessment as HCC: radiologist's opinion
Target condition and reference standard(s)	HCC: histopathological results of the lesion biopsies were considered as reference standard.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No data on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Jalli 2015 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Jang 2016

Study characteristics	
Patient Sampling	Using 401 stored plasma samples obtained from 208 HCC patients and 193 liver cirrhosis control patients, plasma AFP, PIVKA-II, OPN, and DKK-1 levels were measured by ELISA.
Patient characteristics and setting	
Index tests	AFP was measured using an automated quantitative enzyme linked fluorescent assay (ELFA) with mini-VIDAS1 AFP (Biomerieux, Marcy-L'Etoile, France) and with a cut-off value of 20 ng/mL.
Target condition and reference standard(s)	HCC was diagnosed based on histological findings or typical imaging characteristics as defined by the Korean Liver Cancer Study Group guidelines, which are similar to the American Association for the Study of Liver Diseases (AASLD) practice guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Funding: this study was supported by grant from bioMérieux. The funder provided support in the form of salaries for authors [PL and CB] but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Competing Interests: PL and CB were employed by bioMérieux. The remaining authors had no conflicts of interest.

Jang 2016 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Jang 2016 *(Continued)*

Could the patient flow have introduced bias?	Unclear risk

Jeon 2016

Study characteristics			
Patient Sampling	We enrolled 157 consecutive patients with newly diagnosed HCC and 156 patients with liver cirrhosis (LC) as the control group. Age range not reported. Males 69%.		
Patient characteristics and setting			
Index tests	Plasma AFP was measured using an automated enzyme-linked chemiluminescent immunoassay (ELICA) with a cut-off value of 20 ng/mL.		
Target condition and reference standard(s)	HCC was diagnosed by histological and imaging findings outlined by the American Association for the Study of Liver Disease (AASLD) practice guidelines.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest declared		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Jeon 2016 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Ji 2016

Study characteristics	
Patient Sampling	A total of 1034 patients were enrolled, of whom 521 were in the cohort for differential diagnosis (cohort A), 447 were in the cohort for high-risk population surveillance (cohort B), and 66 were in the treatment-monitoring cohort (cohort C). Cohort B comprised individuals with HCC, chronic hepatitis B (CHB), and LC and HCs who were recruited from EHBH, CZH, and RMH of Wuhan University in Hubei Province and from NFH of Southern Medical University in Guangdong Province from January 2013 to February 2014. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP was measured by the electrochemiluminescence immunoassay (ECLIA) (Roche E170 Analyzer, Roche, Tokyo, Japan). Predefined cut-off value 20 ng/mL.
Target condition and reference standard(s)	The diagnosis of liver cirrhosis was based on the histopathology of a liver biopsy or clinical, laboratory, and imaging evidence when possible. Patients with cirrhosis who had elevated AFP concentrations were required to have undergone imaging by multiple methods (ul-

Ji 2016 *(Continued)*

	trasonography, CT, or MRI) and to have had no evidence of a hepatic mass for at least 3 months before enrolment. The diagnosis of HCC was made by abdominal ultrasonography, dynamic CT scanning, or MRI characteristics and AFP, and it was confirmed by histopathology.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Ji 2016 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Jiao 2018

Study characteristics			
Patient Sampling	In this study, a total of 443 serum samples including 180 patients with HCC, 61 patients with liver cirrhosis (LC), 99 patients with chronic hepatitis, and 103 healthy individuals were enrolled from November 2011 to April 2013.		
Patient characteristics and setting			
Index tests	Tumor markers (AFP, carcino-embryonic antigen [CEA], carbohydrate antigen 19-9 [CA19-9]), and liver function parameters (total protein [TP], serum total bilirubin [STB], alanine aminotransferase [ALT], and aspartate aminotransferase [AST]) were tested using commercially available electrochemiluminescence immunoassay (Roche Diagnostics Ltd., Shanghai, China)		
Target condition and reference standard(s)	The HCC diagnosis was based on histopathology, and if histopathology was not available, it was performed on two imaging modalities (magnetic resonance imaging, computed tomography, or contrast enhanced ultrasound).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest declared		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Jiao 2018 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Johnson 1978

Study characteristics	
Patient Sampling	50 patients with histologically-confirmed primary hepatocellular carcinoma were investigated at diagnosis.

Johnson 1978 *(Continued)*

	Age range: 53-74. % of males not reported
Patient characteristics and setting	30 patients, all men, aged 53-74 years, had developed the tumour on the basis of underlying cirrhosis. In the other 20 cases (11 men and 9 women, aged 27-72 years), the tumour had arisen in an otherwise normal liver.
Index tests	Quote: "AFP concentrations were estimated using a sensitive radioimmunoassay technique capable of detecting concentrations of 2 IU/ml (2-1 ng/mL). In contrast, positive results with the immunodiffusion technique may be obtained only at concentrations above about 5000 IU/mL (5250 ng/mL). All samples were run in duplicate, and in those in which the concentration was above normal (as established from 50 healthy controls from the unit staff) the assay was repeated at least once."
Target condition and reference standard(s)	50 patients with histologically-confirmed primary hepatocellular carcinoma were investigated at diagnosis.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Johnson 1978 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Kanmura 2007

Study characteristics	
Patient Sampling	153 male patients with chronic liver disease attributable to HCV infection were selected. 77 of the patients were negative for HCC, which was confirmed by US or CT of the abdomen. Samples from 64 patients with HCC were obtained before treatment. Patients were randomly divided into two groups; the second analysis group (group of interest) consisted of 29 and 33 patients with and without HCC. Age range: 64-81. Males 100%
Patient characteristics and setting	
Index tests	AFP: prespecified cut-off at 20 ng/mL
Target condition and reference standard(s)	HCC: US or CT
Flow and timing	No information of interval between index test and reference standard Reference standard: CT or US
Comparative	
Notes	No data on conflicts of interest

Kanmura 2007 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		

Kanmura 2007 *(Continued)*

Could the patient flow have introduced bias?	High risk

Khairy 2015

Study characteristics			
Patient Sampling	Serum levels of AFP and AFP-L3 were determined in 47 patients with HCC and 17 patients with liver cirrhosis admitted to Kasr Al-Aini Hospital Cairo University. Age range and % of males not reported.		
Patient characteristics and setting			
Index tests	AFP was assessed by ELISA technique in all patients. No pre-defined cut-off		
Target condition and reference standard(s)	HCC: all HCC patients were diagnosed by non-invasive criteria applied to cirrhotic patients according to the 2012 European Association for the Study of the Liver (EASL) guidelines.		
Flow and timing	No data on interval between index test and reference standard		
Comparative			
Notes	No data on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Khairy 2015 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Kim 2001

Study characteristics	
Patient Sampling	From May 1996 to November 1999, a total of 52 consecutive patients with liver cirrhosis underwent whole liver transplantation at our institution. Age range and % of males not reported
Patient characteristics and setting	
Index tests	Ultrasound. Quote: "Experienced radiologists (J.H.L. and W.J.L.), retrospectively reviewed pre-transplantation ultrasonographic studies. ATL HDI-3000 (Advanced Technology Laboratories, Bothell, WA) and Acuson XP (Acuson Corp, Mountain View, CA) scanners with 2.5- or 3.5-MHz transducers. All nodular lesions—hyperechoic, hypoechoic, isoechoic, and mixed echogenic lesions larger than 1.0 cm not explainable by normal structures and different from general normal echoes of the liver parenchyma—were interpreted as potential HCCs. A hypoechoic or mixed echogenic lesion with or without a peripheral hypoechoic rind or an isoechoic lesion with a peripheral hypoechoic rind was regarded as HCC. A hyperechoic lesion without a peripheral hypoechoic rind was regarded as a dysplastic nodule."

Kim 2001 *(Continued)*

Target condition and reference standard(s)	Explanted livers were serially sectioned in the transverse or coronal plane at 5 mm to 10 mm intervals depending on the location of hepatic masses. All nodular lesions seen at ultrasonography were matched with corresponding lesions based on their segment locations on the ultrasonograms versus their counterparts in the explanted livers. The ultrasonographic diagnosis was considered correct if the mass identified on ultrasonography coincided with the anatomic location in the pathologic specimen.
Flow and timing	The range of duration between ultrasonography and transplantation was 7 to 100 days (mean, 56 days).
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Kim 2001 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Kim 2006a

Study characteristics			
Patient Sampling	62 HCC patients, 60 patients with chronic liver diseases, and 60 healthy controls Age range not reported. Males 56%		
Patient characteristics and setting			
Index tests	AFP levels were measured by cheminoluminescence method using Elecsys kit (Roche Diagnostic) with no predefined cut-off value.		
Target condition and reference standard(s)	HCC was diagnosed according to EASL diagnostic criteria. Cirrhosis: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Kim 2006a *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Kim 2006b

Study characteristics

Kim 2006b *(Continued)*

Patient Sampling	A case-control study was conducted in patients with hepatitis C antibody-positive liver cirrhosis and liver cancer who visited the hospital between March 2000 and December 2004. Patients co-infected with hepatitis B virus were excluded. Age range not reported. Males 69%
Patient characteristics and setting	
Index tests	AFP: Alpha-fetoprotein was measured by Electrochemiluminescence Assay (Elecsys AFP, Roche, Basel, Switzerland) (normal value < 7.0 ng/mL).
Target condition and reference standard(s)	HCC: either histology, AFP > 400 ng/mL, or hypervascular liver mass on imaging
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Kim 2006b *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Kim 2006c

Study characteristics			
Patient Sampling	A total of 227 consecutive patients with HCC (42) or chronic liver disease (185) were enrolled. Age range not reported. Males 74%		
Patient characteristics and setting			
Index tests	Serum AFP measurement by electrochemilumino-immunoassay. Predefined cut-off value 20 ng/mL		
Target condition and reference standard(s)	HCC: histology, CT, MRI; chronic liver disease: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Kim 2006c *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Kim 2006c *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Kim 2012

Study characteristics			
Patient Sampling	Serum AFP levels were collected in 354 patients with liver disease and 196 patients with HCC. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	AFP: the serum AFP was measured using a routine automated method in chemiluminescent microparticle immunoassay (ARCHITECT i2000SR, Abbott). The cut-off value for the AFP level was set at 20 ng/mL according to the manufacturer's instruction.		
Target condition and reference standard(s)	HCC: all cases of HCC were diagnosed by fine-needle biopsy under the guidance of ultrasonography and surgery.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No data on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		

Kim 2012 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Kim 2014

Study characteristics	
Patient Sampling	Liver cirrhosis group (LC): 35 patients with compensated hepatitis B virus (HBV) cirrhosis and no HCC. The cirrhosis group had at least 1 year of follow-up from the time that serum was obtained for these studies. Patients were diagnosed with cirrhosis, based

Kim 2014 *(Continued)*

	on established clinical, laboratory, and imaging criteria with ultrasound examination. HCC group: 60 patients before HCC treatment who were infected with HBV were also enrolled, from whom serum samples were collected and defined as the HCC group. To reduce causal heterogeneity, HCC patients who had other types of chronic liver disease, except for chronic hepatitis B, such as chronic hepatitis C and alcoholic hepatitis, were excluded. A subgroup of patients with stage I HCC was compared to the patients in the LC group. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP. Predefined cut-off value of 20 ng/mL
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made per the AASLD practice guidelines by a hepatologist with more than 20 years of experience.
Flow and timing	No data on interval between index test and reference standard
Comparative	
Notes	The authors have declared that no competing interests existed.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Kim 2014 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Kim 2016

Study characteristics	
Patient Sampling	During a study period of 10 years, 2074 adult liver transplant (LT) recipients were identified. They were divided into 2 groups: HCC (n = 970; 46.8%) and non-HCC (n = 1104; 53.2%). Age range and % of males not reported
Patient characteristics and setting	A total of 2074 patients underwent living-donor LT (n = 1825) or deceased-donor LT (n = 249) with a mean MELD score 17.0 ± 9.3.
Index tests	AFP and DCP were measured at the time of pretransplant workup. "The upper normal ranges of AFP and DCP in our institution are 7.5 ng/mL and 40 mAU/mL, respectively."
Target condition and reference standard(s)	Patients were divided into 2 groups: HCC (n = 970; 46.8%) and non-HCC (n =1104; 53.2%), according to the presence or absence of viable HCC at the explant liver.

Kim 2016 *(Continued)*

Flow and timing	No data on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Kim 2016 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Kim 2018

Study characteristics			
Patient Sampling	54 Korean patients with HCC and 26 Korean patients with liver cirrhosis were obtained from Samsung Medical Center (Seoul, South Korea). Age range not reported. Males 63.5%.		
Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification. Predefined cut-off value 20 ng/mL		
Target condition and reference standard(s)	HCC histology Cirrhosis: no definition		
Flow and timing	No data on interval between index test and reference standard		
Comparative			
Notes	The authors declare no conflict of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Kim 2018 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Kim 2018 *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Kim 2019

Study characteristics			
Patient Sampling	A prospective cohort the Seoul National University Hospital (Seoul, Republic of Korea). The training set comprised 53 patients with very early or early HCC based on the Barcelona Clinic Liver Cancer staging system [1], 47 patients with cirrhosis and 50 healthy controls enrolled between January 2014 and August 2017 as part of an ongoing study. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification. No pre-definition of the cut-off value		
Target condition and reference standard(s)	HCC was mostly diagnosed based on the noninvasive criteria of an international guideline [1,2]. Cirrhosis was diagnosed based on either histological or clinical findings.		
Flow and timing	No data on interval between index test and reference standard.		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Kim 2019 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Kim 2019 *(Continued)*

Could the patient flow have introduced bias?	High risk

Kim 2019a

Study characteristics			
Patient Sampling	The test set for the validation of the biomarker signatures consisted of 82 patients with very early or early HCC and 80 patients with cirrhosis from an independent study evaluating the metagenomics profiling of HCC between April 2017 and October 2018.		
Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification. No pre-definition of the cut-off value		
Target condition and reference standard(s)	HCC was mostly diagnosed based on the noninvasive criteria of an international guideline [1,2]. Cirrhosis was diagnosed based on either histological or clinical findings. Age range and % of males not reported		
Flow and timing	No data on interval between index test and reference standard.		
Comparative			
Notes	The authors declare no conflict of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Kim 2019a *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Kim 2019b

Study characteristics			
Patient Sampling	We prospectively enrolled patients with compensated liver cirrhosis at Severance Hospital Yonsei University from Jannuary 2007 to June 2010. The exclusion criteria were as follows: HCC at enrolment or past history of it, HCC development within 6 months after enrolment, decompensated cirrhosis, co-infection with human immunodeficiency virus, and loss to follow-up. Age range not reported. Males 63.5%		
Patient characteristics and setting			
Index tests	Serum AFP measurement. no specification. Predefined cut-off value 20 ng /mL and 7 ng/mL		
Target condition and reference standard(s)	Patients received HCC surveillance using US and AFP measurement at least every 6 months. The diagnosis of HCC was established based on the guideline of the AASLD.		
Flow and timing	No data on interval between index test and reference standard.		
Comparative			
Notes	The authors disclose no conflicts of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Kim 2019b *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Krygier 2011

Study characteristics			

Krygier 2011 *(Continued)*

Patient Sampling	Patients have been routinely in surveillance for HCC and registered from 2006 -2009 on eHEPAR III database. 89 with cirrhosis and 29 with HCC The data were prospectively collected. Age range not reported. Males 65%
Patient characteristics and setting	
Index tests	AFP determined by VITROS AFPimmunoassay; the optimal cut-off was derived from data.
Target condition and reference standard(s)	In all participants, US, CT, MRI were performed.
Flow and timing	No data on interval between index test and reference standard.
Comparative	
Notes	No conflicts of interest reported; funded by a prophylactic program eHEPAR III

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Krygier 2011 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Kudo 2019

Study characteristics	
Patient Sampling	A total of 656 Japanese patients with HBV- or HCV-related liver cirrhosis considered at very high risk for HCC development were enrolled. Patients were included if they were aged > 20 years; had HBV- or HCV-related liver cirrhosis (confirmed by liver biopsy or radiologically); portal hypertension or platelet count < 130,000/mL; and no history of HCC; and if they provided informed consent. Patients were excluded if they had a history of hypersensitivity to egg yolk, severe liver dysfunction (AST, ALT, or bilirubin > 10× ULN), cirrhosis associated with HCC, and treatment with interferon, and were aged < 20 years or judged inappropriate for inclusion by the study investigator. 38 participants discontinued the follow-up and were not included in the analyses. Age range: 58-74. Males 42%
Patient characteristics and setting	
Index tests	B mode US. No predefinition of positivity criteria
Target condition and reference standard(s)	CT/MRI every 8 months
Flow and timing	No data on interval between index test and reference standard.
Comparative	
Notes	Masatoshi Kudo received honoraria from Daiichi-Sankyo and GE HealthCare. The remaining authors had no conflicts of interest.

Kudo 2019 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Kudo 2019 *(Continued)*

Could the patient flow have introduced bias?	Unclear risk

Kumada 2014

Study characteristics	
Patient Sampling	Out of 2830 patients positive for hepatitis B surface antigen (HBsAg) or anti-hepatitis C virus (HCV) antibody, who visited the Department of Gastroenterology and Hepatology, 1214 patients met the eligibility criteria: HBsAg- or HCV RNA-positive for more than 6 months, follow-up period of > 3 years before HCC diagnosis, availability of sera sampled at least twice at 12-month intervals, maximal tumour diameter < 3 cm, and 3 nodules or less at diagnosis, and no oral intake of warfarin which is a DCP-inducing agent. Of these 1214 patients, 114 patients had HCC and 1100 patients had no evidence of HCC during the follow-up period. To reduce the confounding effects of covariates between HCC and control patients, we selected patients using propensity score matching. We were able to match 104 patients with developed HCC to 104 non-HCC developing patients. Age range: 14-84. Males 56%
Patient characteristics and setting	
Index tests	AFP. No explicit info on AFP cut-off being predefined
Target condition and reference standard(s)	HCC: 45 patients were diagnosed as HCC histologically (surgical specimen, 39 patients; US-guided needle biopsy specimens, 6 patients). The remaining 59 patients were diagnosed as patients with HCC, showing typical findings of dynamic MRI including hypervascular in the arterial phase with washout in the portal venous or delayed phase. Patients with liver cirrhosis (LC): US, MRI.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	All authors declared that there were no conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Kumada 2014 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Lee 2004

Study characteristics	
Patient Sampling	From January 1988 to December 1997, 253 patients diagnosed with hepatic cirrhosis with hepatitis B virus infection were examined with hepatitis B markers, biochemical tests, serum α-FP, screening tests, and ultrasound.

Lee 2004 *(Continued)*

	Age range: 33-55. Males 68%
Patient characteristics and setting	
Index tests	AFP: serum α-FP levels were measured using the Medgenix α-FPIRMA kit (Biosource, Nivelles, Belgium). AFP cut-off predefined. Quote: "When we defined cut-off values of serum α-FP as 20, 100 and 500 ng/mL, the corresponding sensitivity and specificity for HCC were 62.9% and 24.0%, 7.4% and 54.2%, 77.3% and 91.9%, respectively."
Target condition and reference standard(s)	HCC: patients who had elevated serum α-FP underwent liver ultrasonography (US) and abdominal computed tomography (CT) was performed to confirm the presence of liver cancer. Hepatic angiography and hepatic biopsy were performed if necessary.
Flow and timing	No information in interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Lee 2004 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Lee 2014

Study characteristics	
Patient Sampling	120 patients, diagnosed with HCC for the first time at Korea University Guro Hospital between July 2007 and March 2011, was recruited for this study. The diagnosis of HCC was based on typical imaging patterns and/or histologic examinations conducted according to the AASLD practice guidelines, proposed in 2005. Age range not reported. Males 79%
Patient characteristics and setting	
Index tests	The optimal cut-off values were calculated using the maximum sum of sensitivity and specificity.
Target condition and reference standard(s)	The diagnosis of HCC was based on typical imaging patterns and/or histological examinations conducted according to the AASLD practice guidelines proposed in 2005. Blood samples from 40 patients with CLD, but without HCC, were obtained.
Flow and timing	No information on interval between index test and reference standard.
Comparative	
Notes	No potential conflicts of interest relevant to this article was reported.

Lee 2014 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Lee 2014 *(Continued)*

Could the patient flow have introduced bias?	High risk

Li 2016a

Study characteristics	
Patient Sampling	In this retrospective study, 95 patients with chronic hepatitis B who were diagnosed with a small HCC (≤ 2 cm) or a cirrhotic nodule between July 2014 and September 2015 were involved. The patient inclusion criteria were as follows: (1) ElastPQ measurements were performed on lesion and background liver, (2) lesion nature was confirmed by pathology or at least 2 of the three contrast-enhanced imaging modalities (CEUS, CECT, or CEMRI), and (3) the cirrhotic nodule had follow-up for more than 6 months with no malignancy changes observed in physical examinations. Age range: 38-61. Males 87%
Patient characteristics and setting	
Index tests	AFP: threshold pre-specified: cut-off ≤ 20 ng/mL
Target condition and reference standard(s)	In the small HCC group, 53 patients were included, with 26 patients confirmed by pathology and 27 patients confirmed by at least 2 of the 3 contrast enhanced imaging modalities (CEUS, CECT, CEMRI). Case-control study
Flow and timing	Reference standard: pathology, contrast-enhanced ultrasound, contrast-enhanced CT, contrast-enhanced MRI No data on interval between index test and reference standard
Comparative	
Notes	The authors declared no conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Li 2016a *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Li 2016b

Study characteristics	
Patient Sampling	This study included 435 chronic hepatitis B patients (G1) and 195 preclinical patients (G2) defined as samples longitudinally collected from the same patients as G1, but at an average of 6 months prior to diagnosis. They were divided into 3 cohorts: discovery, training, and validation cohort. Data for accuracy of AFP is provided in training and validation cohorts. Inclusion criteria:

Li 2016b *(Continued)*

	G1 group: (A) No HCC was diagnosed at least one year after G3 time point; (B) Traditional ultrasound and AFP tests were performed on that patient for cancer screening and the data are available. G2 group: (A) No tumours and chronic diseases unrelated to the liver. (B) Traditional ultrasound and alpha-fetoprotein (AFP) tests was performed on that patient once every 6 months for HCC screening, and the data are available to allow assessment of sensitivity and specificity for the biomarkers. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP: AFP cut-off not pre-specified. Quote: "The optimal cutoff value was determined by following criteria: A). maximizing the sum of sensitivity and specificity; B). minimizing the overall error (square root of the sum [1-sensitivity]2+[1-specificity]2) ; C). minimizing the distance of the cutoff value to the top-left corner of the ROC curve"
Target condition and reference standard(s)	HCC: NCCN guidelines (CT, MRI)
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest exist. The authors have no financial relationship to disclose.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Li 2016b *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Li 2016c

Study characteristics	
Patient Sampling	This study included 435 chronic hepatitis B patients (G1) and 195 preclinical patients (G2) defined as samples longitudinally collected from the same patients as G1, but at an average of 6 months prior to diagnosis. They were divided into 3 cohorts: discovery, training and validation cohort. Data for accuracy of AFP is provided in training and validation cohorts. Inclusion criteria: G1 group: (A) No HCC was diagnosed at least one year after G3 time point; (B) Traditional ultrasound and AFP tests were performed on that patient for cancer screening and the data are available. G2 group: (A) No tumours and chronic diseases unrelated to the liver. (B) Traditional ultrasound and alpha-fetoprotein (AFP) tests was performed on that patient once every 6 months for HCC screening, and the data are available to allow assessment of sensitivity and specificity for the biomarkers. Age range and % of males not reported

Li 2016c *(Continued)*

Patient characteristics and setting	
Index tests	AFP: AFP cut-off not pre-specified. Quote: "The optimal cutoff value was determined by following criteria: A). maximizing the sum of sensitivity and specificity; B). minimizing the overall error (square root of the sum [1-sensitivity]2+[1-specificity]2); C). minimizing the distance of the cut-off value to the top-left corner of the ROC curve"
Target condition and reference standard(s)	HCC: NCCN guidelines (CT, MRI)
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"No conflicts of interest exist. The authors have no financial relationship to disclose."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Li 2016c *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Li 2017a

Study characteristics	
Patient Sampling	All study participants were enrolled at 302 Military Hospital of China, Beijing, China, and were followed up during the study period of 36 months, until confirming HCC diagnosis or the date of study end (December 31, 2016). Age range: 40-52. Males 100%
Patient characteristics and setting	A total of 109 patients met the inclusion criteria and were analysed. All participants had a mean age of 53.9 (SD = 9.7) years, were 60.6% male, 94.5% were with a history of HBV infection (see Table 1). During 36 months of follow-up, 34 out of 109 cirrhotic patients were confirmed to have HCC eventually (31.2%).
Index tests	AFP and AFPL3 were measured by Automated Immunoassay Analyzer (COBAS6000, ROCHE, Switzerland) with no predefined cut-off value.
Target condition and reference standard(s)	Guidelines of the Ministry of Health of the People's Republic of China
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Li 2017a *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Li 2017b

Study characteristics	
Patient Sampling	Participants were recruited from a previous cohort (Long An cohort), which comprised of 2258 HBsAg-positive study participants. This group included a subgroup of participants with basal core promotor (BCP) double mutations (1261) and another subgroup of wild type BCP (997). Participants in this study were recruited from the male mutant group from the Long An cohort, and included a total of 300 participants. A prospective cohort study to determine the accuracy of AFP for HCC in those infected with HBV. Age range not reported. Males 61%
Patient characteristics and setting	
Index tests	AFP: AFP was tested using Diagnostic Kit for the Quantitative Determination of Alpha-feto-protein (ELISA) (Beijing Wantai Biological Pharmacy Enterprise Co., Ltd., Beijing, China) according to the manufacturer's instructions. The cut-off value of AFP for HCC was set at 20 ng/mL.
Target condition and reference standard(s)	From the Long An cohort: The diagnosis was made by one or more of the following: (i) surgical biopsy; (ii) elevated serum AFP (levels ≥400 ng/mL), excluding pregnancy, genital cancer and other liver diseases including metastasis of tumours from other organs, plus clinical symptoms or one image (US or computed tomography, CT); (iii) elevated serum AFP (levels <400 ng/mL), excluding pregnancy, genital cancer and other liver diseases including metastasis of tumours from other organs, plus two images (US and CT) or one image (US or CT) and two positive HCC markers such as DCP, GGT II, AFU, CA19-9.
Flow and timing	No data on interval between index test and reference standard
Comparative	
Notes	No info on conflicts of interest

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Li 2017b *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Li 2019a

Study characteristics	
Patient Sampling	Between October 2017 and March 2019, a total of 411 consecutive patients with early HCC and LC were enrolled in the current case-control study. Age range not reported. Males 72.55%

Li 2019a *(Continued)*

Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification. No definition of a predefined cut-off value		
Target condition and reference standard(s)	HCC diagnosis was based on pathology or typical radiological results of HCC on two dynamic image examinations or one dynamic technique with serum AFP level ≥ 200 ng/mL. Liver cirrhosis was diagnosed by histology or on clinical findings with abdominal US, CT, and MRI features of blunted, nodular liver edge accompanied by splenomegaly and oesophageal varices.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors reported no conflicts of interest in this work.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Li 2019a *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Liao 2012

Study characteristics	
Patient Sampling	The study assessed 111 serum samples obtained from individuals in China with no liver disease (n = 26), chronic hepatitis B without cirrhosis (n = 21), HBV-infected cirrhosis (n = 32), or HBV-infected HCC (n = 32). Age range non reported. Males 68%
Patient characteristics and setting	

Liao 2012 *(Continued)*

Index tests	AFP with a cut-off value of 20 ng/mL		
Target condition and reference standard(s)	The diagnosis of HCC was based on histopathology.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest disclosed		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Liao 2012 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		High
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Libbrecht 2002

Study characteristics	
Patient Sampling	Between January 2000 and July 2001, a total of 52 patients with liver cirrhosis underwent liver transplantation at our institution. Age range not reported. Males 46%
Patient characteristics and setting	
Index tests	On US, hyperechoic, hypoechoic, and mixed echogenic nodular lesions were interpreted as HCC.
Target condition and reference standard(s)	"After explantation, the cirrhotic liver was fixed in formalin for 24 to 48 hours. Subsequently, the liver was sectioned at 5-mm intervals, and each section was carefully inspected. Every lesion that was macroscopically different from the surrounding liver tissue in terms of size, colour, texture, or degree of bulging beyond the cut surface was removed and embedded in paraffin. Maximal diameter and segmental localization of each focal lesion was noted. Liver segments were defined according to Couinaud. Four-micron thick sections were made from the paraffin-embedded material and routinely stained with hematoxylin and eosin and Gordon and Sweet reticulin.Microscopic examination of sections was performed by two hepato-pathologists (L.L. and T.R.) in consensus using a multi-headed microscope.After explantation, the cirrhotic liver was fixed in formalin.Each focal lesion was classified according to internationally accepted criteria as low-grade dysplastic nodule (LGDN), high-grade dysplastic nodule (HGDN), HCC, or other type of lesion."
Flow and timing	Mean intervals between imaging examination and transplantation were 70+/- 50 days (range, 2 days to 166 days)
Comparative	
Notes	No conflicts of interest disclosed
Methodological quality	

Libbrecht 2002 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Lim 2015

Study characteristics			
Patient Sampling	This retrospective study enrolled 361 cirrhotic patients with HCC, and 276 cirrhotic patients without HCC occurrence. Age range not reported. Males 68%		
Patient characteristics and setting			
Index tests	The measurements of AFP and AFP-L3 were performed using the mTAS assay (Wako Pure Chemical Industries, Ltd, Osaka, Japan), and PIVKA-II was analysed by an enzyme immunoassay (Fujirebio Inc., Tokyo, Japan). The cut-off values of AFP (20 and 200 ng/mL) and PIVKA-II (40 and 100 mAU/mL) were used.		
Target condition and reference standard(s)	The diagnosis of HCC was made either histologically or non-invasively, and based on the guidelines of the American Association for the Study of Liver Diseases (AASLD) or the European Association for the Study of the Liver (EASL).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest disclosed		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Lim 2015 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Lin 2000

Study characteristics	
Patient Sampling	A total of 198 consecutive patients entered the study from August 1996 to December 1998. This included 122 previously untreated patients with cirrhosis and HCC. The remaining 76 patients with cirrhosis alone were studied as controls. Quote: "As only a few patients with HCC (9/122) were classified as Child C, they were excluded from the differential diagnosis, which included all other patients with HCC and 50 controls classified as Child A and B to avoid a possible bias caused by higher incidence of Child C-classified controls." Age range: 24-89. Males 66.2%
Patient characteristics and setting	
Index tests	AFP: the serum AFP levels were detected by radioimmunoassay (Abbott Laboratories). The normal range was < 20 ng/mL.

Lin 2000 *(Continued)*

Target condition and reference standard(s)	HCC: the diagnosis was made by ultrasound-guided percutaneous aspiration cytology or biopsy.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No info on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Lin 2000 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Lin 2015

Study characteristics	
Patient Sampling	In total, we collected 1416 serum samples from five groups of participants: healthy controls, inactive HBsAg carriers, patients with chronic hepatitis B, patients with HBV-induced liver cirrhosis, and patients with diagnosed hepatocellular carcinoma. Age range: 39-57. Males 83%
Patient characteristics and setting	The recruited participants were defined as healthy individuals, inactive HBsAg carriers, patients with chronic hepatitis B, patients with HBV-induced liver cirrhosis, or patients with hepatocellular carcinoma by medical doctors, according to eligibility criteria listed in the Appendix.
Index tests	The miRNA classifier established in the training stage was initially validated in two cohorts of patients with hepatocellular carcinoma and controls. These two validation cohorts were independent of the discovery cohort and training cohort and were also independent of each other. They were recruited at different times or different hospitals. We compared the ability of the classifier to diagnose hepatocellular carcinoma with the performance of α-fetoprotein at two commonly used cut-offs of 20 ng/mL (AFP20) and 400 ng/mL (AFP400).
Target condition and reference standard(s)	Patients with hepatocellular carcinoma were diagnosed based on at least two imaging technologies (i.e. hepatic ultrasound together with CT, or MRI, or both), and most cases were further confirmed histopathologically according to the AASLD guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Lin 2015 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Lin 2016

Study characteristics			
Patient Sampling	335 participants including early-stage HCC, liver cirrhosis (LC), chronic hepatitis B (CHB), and healthy people as controls (HC) were enrolled in two cohorts. Age range not reported. Males 70%		
Patient characteristics and setting	Cohort I: 26 patients with early-stage HCC, 22 with liver cirrhosis (LC), 23 with chronic hepatitis B (CHB), and 22 healthy controls Cohort II: 96 patients with early-stage HCC, 39 with LC, 51 with CHB, and 56 healthy controls In both cohorts, HCC patients were complicated with both cirrhosis and HBV infection, and LC patients were also HBV-positive.		
Index tests	Serum samples were collected from all participants in both cohorts.		
Target condition and reference standard(s)	HCC: histopathology		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest declared		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Lin 2016 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Liu 2007

Study characteristics	
Patient Sampling	Patients were recruited from 4 hospitals in Beijing (Youan Hospital, Wujing Hospital, Ditan Hospital, and Beida Hospital), from the Nanjing 2nd Hospital in Nanjing, and from Shanghai Hospital in Shanghai, China. In total, 497 HBV-infected patients with chronic liver diseases were recruited. The study included 227 cases with HCC and cirrhosis, and 80 cases with cirrhosis. 47 were excluded because of metastasis, autoimmune liver disease, drug related hepatitis, alcoholic hepatitis, or obstructive jaundice. Age range: 39-64. Males 83%
Patient characteristics and setting	
Index tests	AFP: no explicit data on AFP cut-off value being prespecified
Target condition and reference standard(s)	HCC: 227 patients with cirrhosis and HCC were diagnosed histologically by biopsy, autopsy, and surgical specimens, and clinically by ultrasonography and/or computed tomography scanning in a reg-

Liu 2007 *(Continued)*

	ular examination, and this was combined with the measurement of AFP (cut-off of 20 ng/mL).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Potential conflict of interest: nothing to report

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Liu 2007 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Liu 2010a

Study characteristics			
Patient Sampling	107 patients admitted to Beijing Youan Hospital, Capital Medical University from January 2006 to December 2008 were recruited in this study and divided into HCC group (n = 75) and liver cirrhosis group (n = 32). Age range: 45-65. Males 74%		
Patient characteristics and setting			
Index tests	AFP: serum AFP levels were measured by electrochemiluminescence (Abbott, USA) and GPC3 levels were measured by enzyme-linked immunosorbent assay (ELISA; BioMosaics Company, USA), following their manufacturer's instructions. The cut-off value was set at 400 µg/L according to the guidelines of clinical diagnosis and staging criteria for primary hepatocellular carcinoma (HCC) established by Chinese Society of Liver Cancer in 2001.		
Target condition and reference standard(s)	HCC: diagnostic criteria: 1) hepatic space-occupying lesion with a serum AFP level ≥ 400 µg/L; and 2) serum AFP level < 400 µg/L, but with new hepatic space occupying lesions, arterial phase enhancement on computed tomography or magnetic resonance imaging. Liver cirrhosis (LC): patients in LC group were followed up at least for 2 years.		
Flow and timing	No data on interval between index test and reference standard		
Comparative			
Notes	No data on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Liu 2010a *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Liu 2017

Study characteristics			
Patient Sampling	The study enrolled 360 participants, including 240 patients with HCC, 29 patients with liver cirrhosis (LC), 66 patients with chronic hepatitis B (CHB), and 25 with hepatic cirrhosis, at the Department of Hepatology, Qilu Hospital of Shandong University from July 2014 to July 2015. A history of other tumours, human immunodeficiency virus or autoimmune liver diseases, alcoholic liver disease, nonalcoholic fatty liver disease, and other causes of chronic liver disease were the exclusion criteria. Age range: 18-83. Males 69%		
Patient characteristics and setting			
Index tests	AFP: AFP concentrations higher than 20 ng/mL were considered abnormal. Patients with AFP ≥ 20 ng/ml or methylated SOX1 and VIM were regarded as positive. Patients with AFP < 20 ng/mL and unmethylated SOX1 and VIM were regarded as negative.		
Target condition and reference standard(s)	HCC: all patients with HCC were diagnosed according to the criteria of the American Association for the Study of Liver Diseases (AASLD) practice guidelines, updated in 2010.		
Flow and timing	No data on interval between index test and reference standard		
Comparative			
Notes	No data on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		

Liu 2017 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Liu 2018

Study characteristics	
Patient Sampling	From April 2016 to July 2017, blood samples were obtained from the Zhongnan Hospital of Wuhan University, including 4 groups: 80 preoperative samples of HCC, 83 samples of cirrhosis, 60 samples of chronic hepatitis B, and 83 healthy control (samples collected from the Physical Examination Center of the Zhongnan Hospital of Wuhan University). Age range not reported. Males 78%
Patient characteristics and setting	
Index tests	AFP: to assess whether plasma SNHG18 could be used as a potential diagnostic marker for HCC, receiver–operating characteristic curve (ROC) was constructed by 5 models: HCC versus the healthy control, HCC with AFP levels below 200 ng/mL versus the healthy

Liu 2018 *(Continued)*

	control, HCC versus cirrhosis, HCC with AFP below 200 ng/mL versus cirrhosis with AFP also below 200 ng/mL, and hepatitis B versus the healthy control.
Target condition and reference standard(s)	HCC: all patients had been pathologically diagnosed as HCC, none of whom had previously undergone radiotherapy or chemotherapy treatment.
Flow and timing	No data on interval between index test and reference standard
Comparative	
Notes	The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		

Liu 2018 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Liu 2019

Study characteristics	
Patient Sampling	"In our retrospective study, the participants, including HCC group (newly diagnosed), and control patients with chronic hepatitis B infections (CHB), chronic hepatitis C infections (CHC), non-viral liver diseases, cirrhosis, cholangiocarcinoma were enrolled from the First Hospital of Jilin University (Changchun, China) between March 2012 and March 2017. Inorder to determine whether AFP was associated with the liver inflammation, only patients with abnormal liver function (defined as AST and ALT exceeding the upper limit of normal value at the same time) were included in the study." The exclusion criteria were: 1. Unavailable AFP value. 2. Undergoing extrahepatic acute diseases. 3. Any types of malignancy for patients with the exception of hepatobiliary system. Age range not reported. Males 68%
Patient characteristics and setting	
Index tests	Serum AFP was measured quantitatively by electrochemiluminescence (Cobas e601, Roche). No predefinition of a cut-off value
Target condition and reference standard(s)	The HCC diagnosis was confirmed with histological findings or typical imaging characteristics according to the guidelines of the European Association for the Study of the Liver (EASL). The diagnosis of cirrhosis or other liver diseases were based on clinical indicators and imageological examination in accordance with the international guidelines.

Liu 2019 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Liu 2019 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Liu 2020

Study characteristics	
Patient Sampling	"Participants were recruited from May 2016 to July 2018 at the Department of Hepatology, Qilu Hospital of Shandong University including 105 patients with HBV-HCC, 54 patients with chronic hepatitis B (CHB), and 32 healthy controls. The following inclusion criteria were set: (1) patients > 18 years old; (2) patients with measurable, histologically-proven hepatocellular carcinoma; (3) patients with the clear history of chronic HBV infection. The following exclusion criteria were set: (1) age > 80 years; (2) metastatic disease; (3) patients with a history of other tumours; (4) coinfection with hepatitis virus other than HBV or autoimmune hepatitis (AIH); (5) patients with drug-induced liver injury (DILI); (6) patients with alcoholic liver disease (ALD) or non-alcoholic fatty liver disease (NAFLD); (7) patients previously received surgery, chemotherapy or radiotherapy." Age range 51-64. Males 86%
Patient characteristics and setting	

Liu 2020 *(Continued)*

Index tests	Serum AFP measurement: no specification. Predefined cut-off value: 400 ng/mL
Target condition and reference standard(s)	Patients were diagnosed with HBV-HCC based on the finding from ultrasound, enhanced computed tomography (CT), magnetic resonance imaging (MRI), alpha-fetoprotein (AFP) serology, and needle biopsy of the liver.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			

Liu 2020 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Loglio 2018

Study characteristics	
Patient Sampling	This single centre cohort study was conducted on 64 consecutive patients with nucleotid analogs (NUCs) suppressed liver cirrhosis with HCC (HCC group) and 148 HBV NUC suppressed cirrhotic patients (control group) who remained HCC-free for 84 months after serum collection. Patients in anticoagulant therapy were excluded. Age range not reported. Males 80%
Patient characteristics and setting	

Loglio 2018 *(Continued)*

Index tests	AFP: AFP levels were tested by standard techniques with no predefined cut-off value.
Target condition and reference standard(s)	HCC: reference standard not specified
Flow and timing	No data on interval between index test and reference standard
Comparative	
Notes	Galli C - Abbott diagnostics - employment

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Loglio 2018 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Loglio 2019

Study characteristics	
Patient Sampling	All consecutive HCC-free Caucasian HBsAg-positive mono-infected patients with compensated cirrhosis starting tenofovir (TDF) or enticavir (ETV) between October 31, 2006 and April 1, 2014 at two tertiary Liver Centers in Milan, Italy, were evaluated for inclusion in this longitudinal cohort study, having a normal AFP levels at baseline, or within the first year of therapy. 28 patients were excluded from the study: 17 patients received TDF/ETV for less than one year, 3 developed HCC within the first year of treatment, 2 did not have regular monitoring of serum AFP, 3 had significant alcohol abuse (> 60 g/day for men and > 40 g/day for women assessed by patient's clinical interviews), and 5 who did not normalise AFP levels within the first year despite virological suppression. Age range: 21-83. Males 82%
Patient characteristics and setting	
Index tests	Serum AFP levels were determined by ImmunoAssay in Electrochemistry Luminescence 'ECLIA'. Predefined cut-off value 7 ng/mL
Target condition and reference standard(s)	Contrast-enhanced computed tomography (CT) or magnetic resonance imaging techniques were performed if liver ultrasound could not carefully evaluate the whole hepatic parenchyma during surveillance. As for internal protocol, whenever serum AFP increased > 7 ng/mL in patients with normal ALT levels and permanent undetectable HBV-DNA, with no lesion detected by US, a CT scan or an MRI was performed within 3 months together with a new AFP determination. In patients with negative CT or MRI but still serum AFP levels persistently above the upper normal limit, a CT or MRI was repeated every 3 months.
Flow and timing	The recurrence standard was performed within three months after the index test
Comparative	
Notes	Conflicts of interest: Massimo Iavarone: Speaking and Teaching: Bayer, Gilead Science, Janssen, BTG, Abbvie; Consultant for BTG. Mauro Viganò: Speaking and Teaching: Roche, Gilead Sciences, BMS. Mariagrazia Rumi: Speaking and Teaching: MSD, Abbvie, Gilead; Advisory board: Abbvie.

Loglio 2019 *(Continued)*

	Angelo Sangiovanni: speaker bureau for Bayer, Gilead Science, Janssen, BTG, Abbvie, Novartis, Advisory board for Tiziana science. Massimo Colombo: Grant and research support: BMS, Gilead Sciences; Advisory Committees: Merk, Roche, Novartis, Bayer, BMS, Gilead Sciences, Tibotec, Vertex, Janssen Cilag, Achillion, Lundbeck, GSK, GenSpera, AbbVie, Alfa Wasserman; Speaking and teaching: Tibotec, Roche, Novartis, Bayer, BMS, Gilead Sciences, Vertex, Merk, Janssen, Abbvie. Pietro Lampertico: Speaking bureau/advisory boards: BMS, Roche, Gilead Sciences, GSK, MSD, Abbvie and Janssen, Eiger, Myr Pharma

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			

Loglio 2019 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Lok 2010

Study characteristics	
Patient Sampling	A nested case-control study was used to compare the accuracy of AFP and DCP in the detection of HCC during a 12-month period before the diagnosis of HCC. For this study, all 39 HCC cases (33 definite [32 histologically confirmed] and 6 presumed) diagnosed between randomisation and 3.8 years after randomisation were included. For each case, 2 controls without HCC — matched for treatment assignment, presence of cirrhosis on baseline biopsy, and length of follow-up — were selected. Age range: 43-60. Males 78%
Patient characteristics and setting	
Index tests	AFP levels at enrolment and every 3 months were tested at the local clinical laboratories. Cut-off values 20 ng/mL and 200 ng/mL

Lok 2010 *(Continued)*

Target condition and reference standard(s)	Two definitions of HCC were adopted as described previously: 1 for “definite” HCC and 1 for “presumed” HCC. Definite HCC was defined by histologic confirmation or a new mass lesion on imaging with AFP levels increasing to 1000 ng/mL. Presumed HCC was defined as a new mass lesion on ultrasound in the absence of histology and AFP 1000 ng/mL in conjunction with 1 of the following characteristics: (1) 2 liver imaging studies showing a mass lesion with characteristics of HCC (arterial enhancement wash out), (2) progressively enlarging lesion on ultrasound leading to death, or (3) 1 additional imaging study showing a mass lesion with characteristics of HCC that either increased in size over time or was accompanied by AFP level 200 ng/mL and more than tripling of baseline value. All cases of HCC were adjudicated by an outcomes review panel to ascertain that they met predefined diagnostic criteria and to determine the date when these criteria were first met. Ultrasound examinations were repeated 6 months after enrolment and again every 12 months. Patients with an elevated or rising AFP and those with new lesions on ultrasound were evaluated further by CT or MRI.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Conflicts of interest The authors disclose the following: "Financial relationships of the authors with Hoffmann-La Roche, Inc.,are as follows: A.S. Lok is a consultant; R.K. Sterling is a consultant, receives research support, and is on the speaker’s bureau; J.C. Hoefs is on the speaker’s bureau; T.R. Morgan is a consultant, on the speaker’s bureau, and receives research support; A.M. Di Bisceglie is a consultant, on the speaker’s bureau, and receives research support; W.M. Lee receives research support; and H.L. Bonkovsky receives research support. Financial relationships of the authors with Eisai Co, Ltd, are as follows: A.S. Lok receives research support. The remaining authors disclose no conflicts. Funding Supported by the National Institute of Diabetes & Digestive & Kidney Diseases (contract numbers are listed below); the National Institute of Allergy and Infectious Diseases (NIAID); the National Cancer Institute; the National Center for Minority Health and Health Disparities; by General Clinical Research Center and Clinical and Translational Science Center grants from the National Center for Research Resources and National Institutes of Health (grant numbers are listed below); by Eisai Co, Ltd, through a Materials Cooperative Research and Development Agreement (M-CRADA) with the National Institutes of Health for testing of des--carboxy prothrombin; and by Hoffmann–La Roche, Inc, through - a Cooperative Research and Development Agreement (CRADA) with the National Institutes of Health."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		

Lok 2010 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Lok 2010 *(Continued)*

Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Long 2011

Study characteristics	
Patient Sampling	Blood samples were obtained from 111 patients with HCC (Beijing Youan Hospital and Beijing Chaoyang Hospital), 36 patients with liver cirrhosis (LC) (Beijing Youan Hospital; 26 with Child-Pugh class A and 10 with Child-Pugh class B), 33 patients with chronic hepatitis B (CHB) (Beijing Youan Hospital), 9 patients with colorectal carcinoma, 11 patients with lung carcinoma, 10 patients with breast carcinoma, 8 patients with cerebral vascular accident, 7 patients with myositis, and 100 healthy blood donors. Age range: 29-65. Males 69%
Patient characteristics and setting	
Index tests	AFP: the cut-off value considered positive for AFP was 20 ng/mL.
Target condition and reference standard(s)	HCC: no information regarding reference standard
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors reported no conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		

Long 2011 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Luo 2018a

Study characteristics	
Patient Sampling	In the present study, a total of 1448 participants, including normal controls (healthy volunteers, NC) and patients with chronic hepatitis B (CHB) infections, liver cirrhosis, HCC, gastric cancer, or intrahepatic cholangiocarcinoma were enrolled between September 2008 and May 2014. The exclusion criteria were abnormal liver biochemistry, a history of liver disease or other systematic diseases for the healthy controls, and a history of acute diseases or other types of malignant diseases for patients with liver disease. The discovery cohort consisted of 108 participants. Fasting serum samples were collected. The test cohort consisted of 684 participants. The validation cohort 1 consisted of 572 participants. Age range: 45-69. Males 78%

Luo 2018a *(Continued)*

Patient characteristics and setting	
Index tests	AFP, with a cut-off value of 20 ng/mL
Target condition and reference standard(s)	HCC: the HCC diagnosis was confirmed with ultrasound, computed tomography, or magnetic resonance imaging; and most participants were further diagnosed by histopathology according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines. Cirrhosis was diagnosed based on clinical evidence of portal hypertension or hepatic decompensation according to the same guidelines. Chronic hepatitis B was defined as the presence of hepatitis B surface antigen for > 6 months, concentrations of hepatitis B virus DNA > 105 copies/ mL, and elevated aspartate aminotransferase or alanine aminotransferase levels.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Potential conflict of interest: nothing to report

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Luo 2018a *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Luo 2018b

Study characteristics	
Patient Sampling	In the present study, a total of 1448 participants, including normal controls (NC) (healthy volunteers) and patients with chronic hepatitis B (CHB) infections, liver cirrhosis, HCC, gastric cancer or intrahepatic cholangiocarcinoma were enrolled between September 2008 and May 2014. The exclusion criteria were abnormal liver biochemistry, a history of liver disease or other systematic diseases for the healthy controls, and a history of acute diseases or other types of malignant diseases for patients with liver disease. The discovery cohort consisted of 108 participants. Fasting serum samples were collected. The test cohort consisted of 684 participants. The validation cohort 1 consisted of 572 participants. Age range: 43-65. Males 72%
Patient characteristics and setting	
Index tests	AFP with a cut-off value of 20 ng/mL
Target condition and reference standard(s)	HCC: the HCC diagnosis was confirmed with ultrasound, computed tomography, or magnetic resonance imaging; and most cases were further diagnosed by histopathology according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines.

Luo 2018b *(Continued)*

	Cirrhosis was diagnosed based on clinical evidence of portal hypertension or hepatic decompensation according to the same guidelines. Chronic hepatitis B was defined as the presence of hepatitis B surface antigen for > 6 months, concentrations of hepatitis B virus DNA > 105 copies/ mL, and elevated aspartate aminotransferase or alanine aminotransferase levels.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Potential conflict of interest: nothing to report

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Luo 2018b *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Luo 2018c

Study characteristics	
Patient Sampling	In the present study, a total of 1448 participants, including normal controls (NC) (healthy volunteers) and patients with chronic hepatitis B (CHB) infections, liver cirrhosis, HCC, gastric cancer or intrahepatic cholangiocarcinoma were enrolled between September 2008 and May 2014. The exclusion criteria were abnormal liver biochemistry, a history of liver disease or other systematic diseases for the healthy controls, and a history of acute diseases or other types of malignant diseases for patients with liver disease. The discovery cohort consisted of 108 participants. Fasting serum samples were collected. The test cohort consisted of 684 participants. The validation cohort 1 consisted of 572 participants. Age range: 34-64. Males 72.5%
Patient characteristics and setting	
Index tests	AFP with a cut-off value of 20 ng/mL
Target condition and reference standard(s)	HCC: the HCC diagnosis was confirmed with ultrasound, computed tomography, or magnetic resonance imaging; and most cases were further diagnosed by histopathology according to the guidelines of the American Association for the Study of Liver Diseases. Cirrhosis was diagnosed based on clinical evidence of portal hypertension or hepatic decompensation according to the same guidelines. Chronic hepatitis B was defined as the presence of hepatitis B surface antigen for > 6 months, concentrations of hepatitis B virus DNA > 105 copies/mL, and elevated aspartate aminotransferase or alanine aminotransferase levels.
Flow and timing	No information on interval between index test and reference standard

Luo 2018c *(Continued)*

Comparative			
Notes	Potential conflict of interest: nothing to report		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Luo 2018c *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Ma 2018

Study characteristics			
Patient Sampling	From January 2012 to December 2013, 368 HCC patients were recruited. Enrolment criteria were as follows: (1) definitive HCC diagnosis; (2) no prior anticancer treatment; (3) complete resection of all tumour nodules, with the cut surface being free of cancer by histological examination; TACE treatments targeted intrahepatic lesions; and (4) availability of complete clinicopathologic and follow-up data. A total of 150 HDs and 152 patients with chronic hepatitis B (CHB) and/or liver cirrhosis (LC) without a history of malignancy were enrolled as negative controls. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	AFP: cut-off 400 ng/mL		
Target condition and reference standard(s)	HCC: for the 295 patients who underwent curative resection, HCC diagnosis was based on histopathology, while for the 73 patients who underwent TACE, HCC diagnosis was based on imaging scans and AFP according to the American Association for the Study of Liver Disease (AASLD) practice guidelines.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors declared that there were no conflicts of interest to disclose regarding funding from industrial sources or other disclosures with respect to this study.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Ma 2018 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Mao 2017

Study characteristics			

Mao 2017 *(Continued)*

Patient Sampling	HCC patients with HBV infection who were admitted to Affiliated Nantong No. 3 Hospital of Nantong University between October 2014 and June 2016 were included in this study. 3 mL heparinised blood was collected from each participant of the four groups, including 31 healthy volunteers (NC group), 82 HCC patients with hepatitis B virus infection (HBV) infection (HCC group), 29 patients with HBV-related liver cirrhosis (LC group), and 28 patients with chronic HBV (HBV group). Age range: 26-77. Males 77%.
Patient characteristics and setting	
Index tests	AFP: "plasma AFP was measured using I2000 automatic chemiluminescence immunoassay analyser (Abbott Architect i2000SR, USA). ROC curves were plotted for each biomarker to investigate their capability to distinguish between HCC and non-HCC, and moreover define the cut-off value of each biomarker for HCC diagnosis by maximum sensitivity and specificity."
Target condition and reference standard(s)	HCC: HCC was diagnosed based on histological findings or typical imaging characteristics as defined by the Diagnosis, Management and Treatment of Hepatocellular Carcinoma (V2011) issued by the Ministry of Health of the Chinese People's Republic of China.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The funding organisation played no role in designing the study, collecting, analysing and interpreting the data, writing the report, or decision-making in submitting the report for publication.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		

Mao 2017 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Maringhini 1988

Study characteristics	
Patient Sampling	A total of 363 patients with histologically proven cirrhosis and a clinical suspicion of neoplastic degeneration (pain, fever, weight loss, and increased alkaline phosphatase levels) were prospectively investigated from January 1980 to October 1984. Age range not reported. Males 75%
Patient characteristics and setting	
Index tests	AFP: cut-off 500 ng/mL US: echographic diagnosis of HCC was made when discrete areas of increased, decreased, mixed sonodensity, or a focal dilatation of intrahepatic bile ducts were identified.
Target condition and reference standard(s)	HCC: HCC was diagnosed by blind biopsy, laparoscopic biopsy, sonographic-guided fine aspiration cytology or US-guided microcore biopsy in 74 patients, and by AFP > 500 ng/mL and a clinical

Maringhini 1988 *(Continued)*

	follow-up or selective angiography in 72 patients. All patients with a final diagnosis of either cirrhosis (C) + HCC or C only were followed for at least one year or until death.
Flow and timing	No information on interval between index test and reference standard. At US, there were 28 patients with non-evaluable results; estimates according to intention-to-diagnose principle.
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Maringhini 1988 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Marrero 2003

Study characteristics	
Patient Sampling	All the patients were enrolled from the liver and liver transplantation clinics at the University of Michigan Medical Center between September 2001 and May 2002. Four groups of consecutive participants were enrolled. Group 1 (G1): normal healthy individuals with no history of liver disease, alcohol consumption less than 40 g/week, and no risk factors for viral hepatitis. All participants were documented to have normal liver biochemistry Group 2 (G2): patients with histologically-confirmed noncirrhotic chronic hepatitis Group 3 (G3): patients with histologically-proven cirrhosis and compensated liver disease (i.e. Child-Turcotte-Pugh [CTP] score 7) Group 4 (G4): patients with histologically-proven HCC Age range: 43-69. Males 54%.
Patient characteristics and setting	
Index tests	AFP: AFP was tested using commercially available immunometric assays utilising enhanced chemiluminescence at the University of Michigan Hospital Clinical Diagnostic Laboratory. To determine the optimal cut- off value for DCP and AFP in the diagnosis of HCC, receiver operating characteristic (ROC) curves were constructed using all possible cut-offs for each assay.
Target condition and reference standard(s)	HCC: histology

Marrero 2003 *(Continued)*

	Computed tomography and magnetic resonance imaging studies of patients with HCC were reviewed by a radiologist who was not aware of the serum marker results.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Marrero 2003 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Marrero 2005

Study characteristics	
Patient Sampling	Consecutive patients with HCC, and patients with cirrhosis that were age, sex, and race/ethnicity matched to the HCC patients were enrolled from the Liver Clinic between September 2001 and August 2004. Age range: 46-66. Males 65%
Patient characteristics and setting	
Index tests	AFP: AFP was tested using commercially available immunoassays utilising enhanced chemiluminescence at the University of Michigan Hospital Clinical Diagnostic Laboratory. To determine the optimal cut-off value for GP73 and AFP in the diagnosis of HCC, ROC curves were constructed using all possible cut-offs for each assay.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made by histopathology (n = 107, including all T1 lesions), and if histopathology was not available by two imaging modalities (ultrasound [US], magnetic resonance imaging [MRI], or computed tomography) showing a vascular enhancing mass > 2 cm (n = 37). Liver cirrhosis [LC] control group: the people with cirrhosis in the control group were followed for a median of 12 months (range 7–18 months) after enrolment, and no one had developed HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No info on conflicts of interest

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Marrero 2005 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Marrero 2009

Study characteristics	
Patient Sampling	We performed a large, EDRN-defined, phase-2 biomarker case-control study. Cases included consecutive adult patients with HCC seen between February 2005 and August 2007 at 7 medical centres in the USA. Patients with HCC were excluded if they were younger than 18 years of age, had prior treatment of their tumour, or history of other solid tumours. Controls were excluded if there was clinical evidence of significant hepatic decompensation (refractory ascites, grade 3 or 4 encephalopathy, or hepatorenal syndrome), Child–Pugh class C or Model for End-Stage Liver Disease (MELD) score 15, detection of HCC at initial evaluation or at follow-up, need for long-term immunosuppressive therapy, prior solid organ transplant, previous or current cancer history (excluding nonmelanoma skin cancer), and significant medical comorbidities in which survival was predicted to be less than a year. Age range 46-71. Males 75%
Patient characteristics and setting	
Index tests	AFP: "Peripheral blood was collected from each participant at the time of the office visit prior to treatment. All aliquots were shipped to a centralised storage facility at the University of Michigan. One aliquot was sent to a centralised laboratory at the University of California, Los Angeles, where the des-γ carboxyprothrombin (DCP), AFP, and AFP-L3% assays were performed blinded to clinical data and identifiers. Sera from 10% of the participants were assayed at a different facility for quality control purposes. AFP and AFP-L3% were simultaneously determined in serum by automated systems (Wako, Mountain View, CA). All samples were performed in duplicates. Samples with AFP value exceeding 1000 ng/mL (upper limit of standard curve) were diluted 10-, 100-, and 1000-fold and remeasured. Samples from 554 (65%) patients had total AFP values of 10 ng/mL with non-detectable AFP-L3%. For this analysis, non-detectable values were assigned a value of 0.5%, which is the lower limit of detection. Another 66 (8%) samples had AFP-L3% values that were nonreportable, and the results of these samples were nonreliable because the total AFP was between 10 ng/mL and 20 ng/mL. For this analysis, nonreportable values were set as missing data. The study was designed to have above 90% power at 1-sided 5% type 1 error for comparing the joint sensitivity and specificity for differentiating early stage HCC from cirrhotic patients between DCP and AFP at current clinical cutoff points of 20 ng/mL for AFP, 10% for AFP-L3%, and 150 mAU/mL for DCP."
Target condition and reference standard(s)	HCC: HCC was defined by histologic examination or by the appropriate imaging characteristics as defined by accepted guidelines. Control group: controls were patients with cirrhosis seen during the same period as the cases at these centres. To assure that controls did not have HCC, all controls were assessed by AFP and an imaging test (US, CT, or MRI) 6 months after enrolment.
Flow and timing	No data on interval between index test and reference standard
Comparative	
Notes	The authors disclosed no conflicts.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Marrero 2009 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Marrero 2009 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Mashaly 2018

Study characteristics	
Patient Sampling	The study included 75 patients (44 HCC group, 31 liver cirrhosis group). Patients with other malignancies or organ dysfunction were precluded. Age range: 40-70. Males 68%
Patient characteristics and setting	
Index tests	AFP: serum AFP was measured using human AFP ELISA kit from (DiaMetra Company, Spello, Perugia Italy). Cut-off value: 200 ng/mL
Target condition and reference standard(s)	HCC was suspected clinically by elevated AFP levels, further imaging studies done to detect focal hepatic lesions using abdominal ultrasound then confirmed by computed tomography or magnetic resonance.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared that there was no conflict of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Mashaly 2018 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Matievskaya 2003

Study characteristics	
Patient Sampling	159 consecutive patients with chronic HBV and HCV infection referred to liver centre were enrolled. Patients with autoimmune hepatitis or alcoholic liver disease were excluded. Age range: 17-82. Males 67%

Matievskaya 2003 *(Continued)*

Patient characteristics and setting	
Index tests	Serum AFP by immunoenzimatic assay; cut-off value 200 ng/mL
Target condition and reference standard(s)	HCC were confirmed by histology, imaging and AFP
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Matievskaya 2003 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Matsuda 2008

Study characteristics	
Patient Sampling	Blood was collected from patients with chronic liver disease (hepatitis or cirrhosis defined by platelet count of 100 x 10^3/mL or over, and less than 100 x 10^3/mL, respectively), those with HCC before treatment for the first lesions, and from healthy volunteers from June 2005 to February 2008. Patients with other malignancies or who die from causes related to the operations were excluded. Age range: 38-75. Males 70%
Patient characteristics and setting	
Index tests	The serum alpha-fetoprotein (AFP) concentrations were determined by chemiluminescent enzyme immunoassay (Lumipulse AFP-N; Fujirebio, Japan). The cut-off values calculated by the ROC curve were 11.35 ng/mL in AFP (75.00% sensitivity, 80.95% specificity).
Target condition and reference standard(s)	HCC: no information regarding reference standard
Flow and timing	No information regarding reference standard. No information on interval between index test and reference standard. Tissue specimens were obtained from 52 HCC patients, and 51 patients were analysed for biomarker accuracy.
Comparative	
Notes	No information on conflicts of interest

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Matsuda 2008 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Mauduit Astolfi 1987

Study characteristics			
Patient Sampling	80 patients with cirrhosis and clinically-suspected hepatocarcinoma were retrospectively analysed. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Standard US examination Positivity criteria: single nodule or multiple nodules with different echogenicity of surrounding parenchyma or diffuse alteration of hepatic parenchyma with neoplastic pattern		
Target condition and reference standard(s)	Histology: laparoscopy with biopsy or US-guided biopsy		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Mauduit Astolfi 1987 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

McMahon 2000

Study characteristics	
Patient Sampling	This is a 16-year longitudinal follow-up study of total 1487 patients chronically infected with HBV (HBsAg-positive for 12 months or longer). Blood samples were taken every 6 months. Age range not reported. Males 59%
Patient characteristics and setting	
Index tests	AFP: "Between 1982 and 1993, a level of above 25 ng/mL was used as a cut-off level for further evaluation for the presence of HCC. Before 1993, all persons diagnosed with HCC had an AFP at diagnosis of greater than 25 ng/mL. After 1993 the cut-off level was lowered to 15 ng/mL because a carrier with an AFP of 15 ng/mL had been found to have a large nonresectable tumour."
Target condition and reference standard(s)	Carriers whose AFP levels were elevated were contacted by phone. Men or nonpregnant women were requested to have another blood sample drawn 1 month later for testing of their AFP level. If the second AFP level was elevated, the patient was sent to the Alaska Native Medical Center for evaluation with an ultrasound (US) examination of the liver, clinical evaluation, and liver function tests, including serum transaminase levels and total bilirubin. Computed tomography (CT) of the liver was performed in selected individuals if the US examination was unsatisfactory or suggested a lesion. If no lesion was evident on

McMahon 2000 *(Continued)*

	imaging studies, the AFP and US were repeated every 3 to 6 months until the AFP level returned to normal or a lesion suggestive of HCC was found.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

McMahon 2000 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Mehinovic 2018

Study characteristics			
Patient Sampling	50 patients with liver cirrhosis and 50 patients with HCC were included in this study. Age range: 29-81. Males 57%		
Patient characteristics and setting			
Index tests	Chemiluminescent microparticle immunoassay ARCHITECT AFP assay (CMIA, Ireland) was used for AFP detection; no prespecified cut-off value.		
Target condition and reference standard(s)	Unclear: ECHO and computerised tomography (CT) were used to detect and measure the size of HCCs.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"Funding: This research did not receive any financial support. Competing Interests: The authors have declared that no competing interests exist."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Mehinovic 2018 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Min 2014

Study characteristics	
Patient Sampling	500 patients with HBV-related liver cirrhosis received surveillance for HCC by AFP and ultrasound examination every 3–6 months. During the median follow-up of 49 months (range: 6–88 months), 76 patients developed HCC. Age range and % of males not reported

Min 2014 *(Continued)*

Patient characteristics and setting			
Index tests	AFP: cut-off value 10 ng/mL		
Target condition and reference standard(s)	HCC: no clear data on reference standard. All patients were followed-up with US and AFP.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Min 2014 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Minami 2015a

Study characteristics	
Patient Sampling	Four cohorts were enrolled based on the presence of HCC or hepatitis C virus (HCV) status. Patients with positive serology for hepatitis B surface antigen were excluded. Inclusion criteria were as follows: Cohort 1: patients who developed HCC after HCV eradication using interferon (IFN)-based therapy. These patients were enrolled from January 1990 to December 2012 at the Department of Gastroenterology of the University of Tokyo Hospital. Of the 37 patients who developed HCC after HCV eradication, 29 were defined as early stage HCC. Cohort 2: patients who did not develop HCC after HCV eradication using IFN-based therapy. These 179 patients, who were enrolled from January 1990 to December 2012, achieved SVR, confirmed as the absence of HCC during follow-up for more than 1 year. Cohort 3: patients who developed HCC without HCV eradication, consisting of 1185 chronic hepatitis C patients who developed HCC, treated initially with radical therapies (percutaneous ethanol injection therapy, percutaneous microwave coagulation therapy, or radiofrequency ablation) from January 1990 to December 2009 at the same institution, excluding those who achieved sustained virological response (SVR) before HCC development. Cohort 4:patients without either HCC or HCV eradication. These patients were extracted from the follow-up cohort, which was analysed for hepatitis C-related HCC development. A matched case–control study was conducted to compare the diagnostic accuracy of AFP between SVR cohorts 1 and 2 and non-SVR cohorts 3 and 4 to minimise the influence of non-HCC factors on AFP levels. Age range: 54-73. Males 79%
Patient characteristics and setting	
Index tests	AFP: to analyse the diagnostic accuracy of AFP for differentiating HCC cases from controls, an area under the receiver operator characteristic curve (AUROC) was calculated. The optimal cut-off value was validated by calculating the Youden index.

Minami 2015a *(Continued)*

Target condition and reference standard(s)	HCC: HCC was diagnosed by dynamic computed tomography (CT) or magnetic resonance imaging (MRI) with hyperattenuation in the arterial phase and washout in the late phase. As the diagnosis of HCC was not definite by CTor MRI, an ultrasound-guided tumour biopsy was performed and pathological diagnosis made based on the Edmondson-Steiner criteria. Early stage HCC was defined as a tumour number < 3 with each lesion < 3 cm in diameter with no evidence of vascular invasion or extrahepatic metastasis.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors have no conflicts of interest to disclose."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Minami 2015a *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Minami 2015b

Study characteristics	
Patient Sampling	Four cohorts were enrolled based on the presence of HCC or hepatitis C virus (HCV) status. Patients with positive serology for hepatitis B surface antigen were excluded. Inclusion criteria were as follows: Cohort 1: patients who developed HCC after HCV eradication using interferon (IFN)-based therapy. These patients were enrolled from January 1990 to December 2012 at the Department of Gastroenterology of the University of Tokyo Hospital. Of the 37 patients who developed HCC after HCV eradication, 29 were defined as early stage HCC. Cohort 2:patients who did not develop HCC after HCV eradication using IFN-based therapy. These 179 patients, who were enrolled from January 1990 to December 2012, achieved SVR, confirmed as the absence of HCC during follow-up for more than 1 year. Cohort 3: patients who developed HCC without HCV eradication, consisting of 1185 chronic hepatitis C patients who developed HCC treated initially with radical therapies (percutaneous ethanol injection therapy, percutaneous microwave coagulation therapy, or radiofrequency ablation) from January 1990 to December 2009 at the same institution, excluding those who achieved SVR before HCC development. Cohort 4: patients without either HCC or HCV eradication. These patients were extracted from the follow-up cohort, which was analysed for hepatitis C-related HCC development. A matched case–control study was conducted to compare the diagnostic accuracy of AFP between SVR cohorts 1 and 2 and non-SVR cohorts 3 and 4 to minimise the influence of non-HCC factors on AFP levels.

Minami 2015b *(Continued)*

	Age range: 56-74. Males 79%
Patient characteristics and setting	
Index tests	AFP: to analyse the diagnostic accuracy of AFP for differentiating HCC cases from controls, an area under the receiver operator characteristic curve (AUROC) was calculated. The optimal cut-off value was validated by calculating the Youden index.
Target condition and reference standard(s)	HCC: HCC was diagnosed by dynamic computed tomography (CT) or magnetic resonance imaging (MRI) with hyperattenuation in the arterial phase and washout in the late phase. As the diagnosis of HCC was not definite by CTor MRI, an ultrasound-guided tumour biopsy was performed and pathological diagnosis made based on the Edmondson-Steiner criteria. Early stage HCC was defined as a tumour number < 3 with each lesion < 3 cm in diameter with no evidence of vascular invasion or extrahepatic metastasis.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors have no conflicts of interest to disclose."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Minami 2015b *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Miura 2007

Study characteristics	
Patient Sampling	104 consecutive patients [64 patients with HCC, 20 with liver cirrhosis (LC), and 20 with chronic hepatitis (CH)] were enrolled in this study. All HCC patients had LC as underlying liver disease. 66 patients were infected with HCV, 30 with HBV, 3 with both viruses, and 5 with no viral markers. 50 healthy individuals including 12 females served as controls. To assess the accuracy of diagnostic tests, the matched data sets (chronic liver diseases patients and HCC patients) regarding biomarkers were analysed by using receiver operator characteristic (ROC) curve analysis. Age range: 22-83. % of males not reported
Patient characteristics and setting	
Index tests	AFP
Target condition and reference standard(s)	No data on reference standard

Miura 2007 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared that they had no financial conflict of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear

Miura 2007 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Miura 2010

Study characteristics		
Patient Sampling	437 consecutive patients ((303 patients with HCC, 89 with chronic hepatitis (CH), and 45 with liver cirrhosis (LC)), who were admitted at Tottori University related Hospitals, Osaka Red Cross Hospital, and Fukuoka University Chikushi Hospital, in Japan, between November, 2002 and December, 2006, were enrolled in this study. Age range and % of males not reported	
Patient characteristics and setting		
Index tests	Serum AFP. No specification. Cut-off value 10 ng/mL	
Target condition and reference standard(s)	The patients were diagnosed by blood chemistry, US, computed tomography (CT), AFP and/or biopsy under US. HCC was diagnosed according to the the American Association for the Study of Liver Diseases (AASLD) practice guidelines.	
Flow and timing	No information on interval between index test and reference standard	
Comparative		
Notes	The authors declared that they had no competing interests.	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Miura 2010 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Mohamed 2020a

Study characteristics	
Patient Sampling	Serum samples were collected from a total number of 200 participants. All patients were recruited from the Department of Gastroenterology and Hepatology, Theodor Bilharz Research Institute, in Egypt, during the period from October 2017 to November 2018.

Mohamed 2020a *(Continued)*

	40 healthy volunteers were involved in the current study as a control group. Participants were divided into 3 categories: Control group (n = 40), liver cirrhosis group without HCC (LC) who had chronic hepatitis C (CHC) more than 6 months of infection (n = 80) and HCC patients who had cirrhosis and were currently infected by HCV, but did not start the treatments (n = 80). Age range not reported. Males 66%
Patient characteristics and setting	
Index tests	Serum AFP level was determined using sandwich Enzyme Linked Immunosorbent Assay (ELISA). No predefinition of a cut-off value
Target condition and reference standard(s)	HCC: CT Control: no specification
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare no conflicts of interests."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Mohamed 2020a *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Mohamed 2020b

Study characteristics	
Patient Sampling	The study included 70 patients with chronic liver disease, divided into two groups:

Mohamed 2020b *(Continued)*

	Group (I): 40 patients with HCC Group (II): 30 patients with liver cirrhosis and without any evidence of HCC (Group III): 30 healthy adults, recruited as controls Age range not reported. Males 64%
Patient characteristics and setting	
Index tests	Serum AFP was assayed via an enzyme Immunoassay. No predefinition of a cut-off value
Target condition and reference standard(s)	Hepatocellular carcinoma was suspected by the abdominal US and confirmed by triphasic CT scan with contrast.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The author declares no conflicts of interest, financial or otherwise."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Mohamed 2020b *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Mok 2004

Study characteristics	
Patient Sampling	A single-centre prospective screening study was initiated in October 1997. The authors recruited study candidates from the hepatology clinic at the Prince of Wales

Mok 2004 *(Continued)*

	Hospital in Hong Kong who were hepatitis B virus carriers between 40 and 70 years old. They excluded patients with non-hepatitis B–related cirrhosis, a known history of malignancy, or a medical condition associated with a life expectancy of less than 2 years. For the purpose of the present study, eligibility was confined to patients with elevated serum α-fetoprotein levels who had at least one abdominal sonogram and had undergone hepatic angiography with a post-Lipiodol CT scan (Lipiodol CT) within 2 months of the abdominal sonogram. Recruitment for the study was completed in March 2000 when the predetermined sample size of 1,018 participants had been attained. Age range: 40-69. Males 78%
Patient characteristics and setting	
Index tests	US: abdominal sonography was performed by one of two designated radiologists, each of whom had more than 10 years of experience, using an electronic curvilinear 3.5-MHz real-time transducer, scanning subcostally and intercostally with the patient in a supine, and then left decubitus, position. A focal lesion was defined as a well-defined solid nodule (mass) with hypoechoic, hyperechoic, or mixed sonographic pattern. Results were categorised as positive, probable, or negative.
Target condition and reference standard(s)	The gold standard for the diagnosis of hepatocellular carcinoma is histology. Tissue for histological assessment was obtained from hypervascular tumours either at surgery or percutaneous needle biopsy. Patients who declined surgery or histological assessment by biopsy were followed every 3 months with repeated α-foetoprotein level measurements and abdominal sonography. We also followed patients with elevated α-foetoprotein levels or focal lesions but normal findings on Lipiodol CT every 3 months with repeated α-foetoprotein level measurements and abdominal sonography for 2 years and then every 6 months thereafter. Patients with serum α-foetoprotein levels above 20 ng/mL on two occasions at least 1 week apart or focal lesions on abdominal sonograms were further evaluated with Lipiodol CT.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Mok 2004 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Montaser 2012

Study characteristics

Montaser 2012 *(Continued)*

Patient Sampling	The study was conducted on 80 patients (40 patients with HCC and 40 patients with chronic liver disease (CLD) as diseased controls) in addition to 40 apparently healthy individuals who served as a healthy control group. Age range not reported. Males 82%
Patient characteristics and setting	
Index tests	Serum AFP was measured by ELISA technique using commercially available immunometric assay using enhanced cheminolumines-cence (EQUIPAR Disgnostica) with no predefined cut-off value.
Target condition and reference standard(s)	HCC was diagnosed on the base of imaging techniques (US and CT)
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest disclosure

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Montaser 2012 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		High risk	

Moriya 2013

Study characteristics	
Patient Sampling	Between April 2011 and March 2012, 300 patients with chronic liver disease for HCC lesions were screened using Gd-EOB-DTPA enhanced MRI. Reduced uptake of Gd-EOB-DTPA on T1-weighted hepatobiliary phase images 20 minutes after contrast medium injection indicated the presence of novel tumours in 15 patients. The diagnosis was histopathologically-confirmed on biopsies from the 15 tumours, with concurrent measurement of serum levels of AFP and AFP-L3%. Frozen blood samples were obtained from a control cohort (n = 183) of patients with chronic hepatitis or cirrhosis caused by hepatitis B or C virus between January 2010 and August 2010. None of the patients in the control cohort demonstrated HCC during the 2-year follow-up period. The aim of this study was to evaluate the clinical usefulness of the highly sensitive assay of AFP-L3% as a marker for the early diagnosis of HCC. Age range: 49-76. Males 53%
Patient characteristics and setting	
Index tests	AFP: with a cut-off value of 20ng/mL. They measured the serum levels of AFP and AFP-L3% using a commercially available automatic measurement system based on a combined LBA–EATA (Wako Pure Chemical Industries Ltd., Osaka, Japan) of blood samples collected at the time of imaging.
Target condition and reference standard(s)	HCC: the study authors performed a sonography-guided or fusion image-guided percutaneous fine-needle biopsy on the newly-detected tumours. Blinded histopathological diagnoses were performed according to the new histological criteria defined by the ICGHN in 2009 with the consensus of two pathologists special-

Moriya 2013 *(Continued)*

	ising in liver lesions (4). Tumours were diagnosed as early HCC if they had the following characteristics: increased cell density with little cell atypia, architectural alterations of a thin trabecular structure an acinus in some areas, and stromal invasion. Control group: none of the patients in the control cohort demonstrated HCC during the 2-year follow-up period.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors have no conflicts of interest to declare."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Moriya 2013 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Moriyama 2000

Study characteristics	
Patient Sampling	This study included 39 patients with HCC based on chronic liver disease (CLD) and 50 CLD patients without HCC. CLD included clinicopathologically-proven chronic hepatitis (CH) and compensated liver cirrhosis (LC). In order to avoid the overestimation of the markers, they excluded cases with advanced HCC. Therefore, they selected 39 HCC patients with the tumour size < 3 cm and the number of tumours that were < 3. Age range: 21-78. Males 60%
Patient characteristics and setting	
Index tests	AFP: serum AFP concentrations were assayed using a solid phase immunoassay analyzer (ARC 1000, Aloka, Tokyo) with a detection limit of 1.0 ng/mL. According to the ROC curve analysis, the optimal cut-off value for AFP was 18.0 ng/mL.
Target condition and reference standard(s)	HCC: HCC was diagnosed by characteristic findings from ultrasonography, computed tomography, and hepatic angiography which are compatible with HCC or in combination with the histological examinations of a tumour biopsy. No information on diagnosis of control group
Flow and timing	No information on interval between index test and reference standard
Comparative	

Moriyama 2000 *(Continued)*

Notes	No information on conflicts of interest		

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Moriyama 2000 *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Mukozu 2013

Study characteristics			
Patient Sampling	87 adult Japanese patients who had hepatitis C virus (HCV)-related liver cirrhosis with or without HCC were treated between 2004 and 2011. The control group was composed of 37 adult Japanese patients with chronic hepatitis C, diagnosed by examination of liver biopsy specimens. Age range 42-83. Males 78%		
Patient characteristics and setting			
Index tests	Measurements of AFP, AFP-L3, and DCP was performed by lectin-affinity electrophoresis coupled with antibody-affinity blotting method or a microchip capillary electrophoresis and liquid-phase binding assay using a μTSAWako i30 auto-analyser (Wako Pure Chemical Industries, Ltd., Osaka, Japan). The cut-off values for serum AFP, AFP-L3, and des-gamma-carboxyprothrombin (DCP) were obtained from the guideline of the Japanese Society of Hepatology. AFP cut-off value 15 ng/mL		
Target condition and reference standard(s)	HCC: the diagnosis of HCC was performed using clinical criteria and the findings obtained by B-mode ultrasonography (US), computed tomography (CT) angiography, or magnetic resonance imaging (MRI). The control group was composed of 37 adult Japanese patients with chronic hepatitis C, diagnosed by examination of liver biopsy specimens.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Mukozu 2013 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Mustika 2019

Study characteristics	
Patient Sampling	Participants were 41 patients with chronic hepatitis B and/or C infection, either with or without liver cirrhosis and HCC. Participants were divided into 3 groups: a chronic hepatitis group, a liver cirrhosis group, and a HCC group based on a series of physical examinations and investigations.

Mustika 2019 *(Continued)*

	Age range not reported. Males 82.5%		
Patient characteristics and setting			
Index tests	AFP test using ELISA method to Prodia Laboratory in Malang pre-defined cut-off value 200 ng/mL		
Target condition and reference standard(s)	HCC was established when serum AFP levels were = 200 ng/mL and nodules were present in the liver on ultrasound examination or abdominal CT scan.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Mustika 2019 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Na 2013

Study characteristics	
Patient Sampling	A total of 260 individuals visiting the Yonsei University Health System from July 2008 to December 2009 were enrolled. All participants were classified in the following groups: patients with HCC (HBV-positive, 57), liver cirrhosis (LC; HBVpositive, 27), chronic hepatitis (CH; HBV-positive, 37), cholangiocarcinoma (CC; 22), gastric cancer (GC; 31), and pancreatic cancer (PC; 34), along with 52 HDs

Na 2013 *(Continued)*

	having no liver-related diseases when examined at the Severance Hospital of Yonsei University. Age range: 33-65. Males 67%
Patient characteristics and setting	
Index tests	Both hCE1 and AFP proteins in plasma samples were quantified by ELISA. A commercially available AFP ELISA kit was purchased from Panomics (Fremont, CA), and protocols recommended by the manufacturer were used. Clinically recommended AFP cut-off values 20 ng/mL and 100 ng/mL
Target condition and reference standard(s)	The diagnosis of HCC was made by the pathologists at the Severance Hospital of Yonsei University.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare there are no conflicts of interest."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Na 2013 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Nabih 2014

Study characteristics	
Patient Sampling	The present study included 69 Egyptian HCV-related cirrhotic patients recruited from the Kasr EL-Aini Hospital, Internal Medicine Department. All patients were subjected to triphasic contrast computed tomography (CT) of the liver and were categorised according to the imaging characteristics into two groups: the ‘‘HCC group’’ and the ‘‘LC group’’. Age range not reported. Males 74%
Patient characteristics and setting	
Index tests	AFP kits (Roche Diagnostic GmbH, Mannheim, Germany
Target condition and reference standard(s)	HCC: triphasic CT of the liver was used to detect the presence of focal lesions and assess site, size, and multiplicity of the focal lesions. The absence of focal hepatic lesions was confirmed by triphasic contrast CT. Liver cirrhosis was defined by evidence of affection of synthetic or excretory functions of the liver in the presence of clinical and sonographic findings of chronic liver disease.
Flow and timing	
Comparative	

Nabih 2014 *(Continued)*

Notes	Conflict of interest: the authors declared that there was no conflict of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Nabih 2014 *(Continued)*

Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Nakamura 2006

Study characteristics	
Patient Sampling	Between June 1997 and September 2003, 1377 consecutive patients who were diagnosed with HCC for the first time via typical CT imaging patterns or biopsies during periodical examination, with AFP, DCP, and US at Okayama University hospital or affiliate hospitals were enrolled. Of these 1377 HCC patients, 16 were excluded from the study because the tumour sizes were not available. So, a total of 1361 of these HCC patients were included in this study. "We also examined 355 consecutive patients with chronic hepatitis or cirrhosis (non-HCC), who visited Okayama University Medical School Hospital between June 1997 and September 2003 for regular follow-up. 7 of these patients were excluded because of the absence of either serum des-gamma-carboxyprothrombin (DCP) or AFP data. Age range: 55-62. Males 68%
Patient characteristics and setting	
Index tests	AFP: serum des-gamma-carboxyprothrombin (DCP) and AFP concentrations were measured either when patients were initially diagnosed with HCC, or confirmed not to have HCC by imaging methods. The serum AFP concentrations were measured using a commercially available EIA kit. AFP cut-off values were set at 20 ng/mL, 100 ng/mL, and 200 ng/mL.
Target condition and reference standard(s)	HCC: HCC diagnosis was confirmed histologically in 616 patients who had undergone hepatic resection or US-guided biopsy. The remaining 745 patients were diagnosed by typical HCC image patterns, using angiography, CT, and MRI. Diagnostic criteria of HCC by imaging modalities were based on previous reports of hyperattenuation at the arterial phase, hypoattenuation at the portal phase in dynamic CT or MRI, and tumour stain on angiography. Control group: HCC was excluded by imaging methods, using CT, MRI, and US, and was confirmed via periodical examination using the same methods, every 3 to 4 months for 1 year. All patients without HCC in this study were followed up for at least 1 year, and no HCC developed in this period. No incidental HCC was found in either the HCC or non-HCC groups.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared no conflicts of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Nakamura 2006 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Nakamura 2006 *(Continued)*

Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Nguyen 2002

Study characteristics	
Patient Sampling	"We conducted a case-control study of patients with HCC and chronic HCV infection and patients with hepatitis C-related cirrhosis whose clinical records were available for retrospective reviews at Stanford University Medical Center (SUMC), Veterans Administration Medical Center in San Francisco (VASF), San Francisco General Hospital (SFGH), and University of California Medical Center in San Francisco (UCSF) between 1995 and 2001. All patients in this study had positive anti-HCV and/or HCV-RNA and at least 1 serum AFP measurement. Patients with positive hepatitis B surface antigen (HBsAg), anti-HIV, active nonhepatic malignancies, and hereditary or autoimmune liver diseases were excluded. All patients were sampled consecutively from liver transplant and liver clinic records." Age range: 35-84. Males 80.5%
Patient characteristics and setting	
Index tests	AFP: laboratory tests including serum AFP, aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TB), platelet count, prothrombin time and international normalised ratio (INR), creatinine, HBsAg, anti-HCV, and HCV RNA were determined using standard, commercially available assays. Sensitivity and specificity of various AFP cut-off values were determined for: AFP greater than 10 ng/mL (normal upper limit for most commercial laboratories), AFP greater than 20 ng/mL (recommended threshold for further investigation), and AFP greater than 100 ng/mL and 200 ng/mL (suggested confirmatory values for HCC in patients with hepatic masses.
Target condition and reference standard(s)	HCC: "In total, 163 cases of HCV-related HCC meeting inclusion criteria were identified and confirmed by cytology and/or histology (87 patients, 53.4%) or by the presence of characteristic (i.e. enlarging tumours and/or tumours with typical arterial hypervascularisation) hepatic masses on liver computed tomography (CT), and/or magnetic resonance imaging tests (MRI), and/or hepatic angiography tests (76 patients, 46.6%). Control group: In total, 149 control patients with chronic HCV infection and cirrhosis were identified and confirmed by liver biopsies (89 patients, 59.7%) and/or clinical or radiographic evidence of portal hypertension (60 patients, 40.3%). HCC was excluded by imaging studies [US, CT, MRI, and/or hepatic angiography], one of which must have been performed at least 6 months following the measurement of AFP."
Flow and timing	The median time between diagnostic imaging studies and AF tests was 14 days (range 0-300 days).
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Nguyen 2002 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Nguyen 2002 *(Continued)*

Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Nomair 2019

Study characteristics	
Patient Sampling	"44 individuals with hepatitis C virus (HCV)-related liver cirrhosis with or without HCC were recruited from the Hepatology Department of the Medical Research Institute Hospital, Alexandria University, Egypt during the period from December 2017 to April 2018. Two groups: 22 patients with liver cirrhosis due to HCV infection and 22 patients with HCC complicating HCV-related cirrhosis Patients had negative serum markers of active infection with hepatitis B virus (HBV), human immunodeficiency virus (HIV) and schistosomiasis. Also, patients with a history of alcohol consumption > 30 g/day, autoimmune diseases, malignancies diabetes mellitus, and non-HCV related liver cirrhosis were excluded from the study." Age range not reported. Males 52%.
Patient characteristics and setting	
Index tests	Serum AFP level was measured for all patients with cirrhosis and HCC (using the automated IMMULITE 1000 immunoassay analyzer; Siemens Medical Solutions Diagnostics Corporation, Erlangen, Germany. No predefinition of a cut-off value
Target condition and reference standard(s)	"Hepatocellular carcinoma cases were diagnosed according to the guidelines of the American Association for the Study of Liver Disease (AASLD) published in 2011, which comprised the presence of a hepatic focal lesion on ultrasound, verified by either a contrast-enhanced triphasic CT-scan study or dynamic contrast-enhanced MRI that showed characteristic criteria for HCC diagnosis (arterial uptake of contrast material followed by washout). Liver cirrhosis was diagnosed based on clinical, laboratory, and imaging criteria (coarse echo pattern of the liver on ultrasound), with reporting of the presence/absence of portal hypertension and splenomegaly. Ascites was graded as none, mild/moderate or severe. Child-Turcotte-Pugh (CTP) score and class were used for assessing the severity of liver disease."
Flow and timing	No information on interval between index test and reference standard
Comparative	

Nomair 2019 *(Continued)*

Notes	The authors reported no conflict of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			

Nomair 2019 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Nomura 1996

Study characteristics			
Patient Sampling	27 patients HCC with diameter < 3 cm and 101 controls; 69 with cirrhosis and 32 with chronic hepatitis. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum AFP by latex agglutination immunoassay, cut-off value 20 ng/mL		
Target condition and reference standard(s)	Histology		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Nomura 1996 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Nomura 1999

Study characteristics			
Patient Sampling	The study included 36 patients with solitary small-sized (< 3 cm in diameter) HCC and 49 patients with posthepatitic cirrhosis carrying no HCC. Patients who had been taking antibiotics containing N-methylthiotetrazole (NMTT) were excluded. Also, cirrhotic patients who subsequently developed HCC within 1 year were excluded. Age range not reported. Males 71%		
Patient characteristics and setting			
Index tests	AFP: serum AFP levels were determined by latex agglutination immunoassay (IATROMATE AFP, Diatron, Tokyo, Japan). Values of 20 ng/mL were considered upper limit of the reference interval.		
Target condition and reference standard(s)	HCC: a diagnosis of HCC was made histologically in all cases. Control group: in patients with cirrhosis, HCC was ruled out on the basis of the results of imaging studies including sonography and CT performed on a regular basis. Also, patients with liver cirrhosis who subsequently developed HCC within 1 year were excluded.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		

Nomura 1999 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Nomura 2012

Study characteristics	
Patient Sampling	Serum samples were obtained prior to treatment from 58 consecutive patients with early or relatively early hepatitis C virus (HCV)-related HCC and 137 patients with HCV-related liver cirrhosis without evidence of HCC. 58 consecutive patients with early (Stage I, n = 28) and relatively early (Stage II, n = 30) HCV-related HCC (29 males and 29 females, 69.7 ± 8.6 years old) hospitalised in the Gastroenterology Unit of Chiba University Hospital between January, 2008 and December, 2010 were included in the study. For comparison, 137 people with liver cirrhosis (56 males and 81 females, 65.8 ± 11.0 years old) encountered during the same period were also included. Serum samples were obtained prior to initial treatment. Age range: 55-78. Males 44%
Patient characteristics and setting	

Nomura 2012 *(Continued)*

Index tests	AFP: serum levels of AFP and PIVKA-II were measured using commercial enzyme immunoassay kits (Fujirebio Inc., Tokyo, Japan), with cut-off values set at 40 ng/mL and 40 mAU/mL, respectively, to give 90% specificity in patients with liver cirrhosis.
Target condition and reference standard(s)	HCC: The diagnosis of HCC was based on typical findings in three-phase dynamic CT or MRI. In cases with inconclusive imaging findings, the diagnosis was confirmed histopathologically. Control group: no information on how they excluded HCC
Flow and timing	In 12 cases, serum samples were obtained just before and 2 months after surgical resection of the tumours in the Department of Hepato-biliary Surgery at Chiba University Hospital.
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Nomura 2012 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Oka 1994

Study characteristics	
Patient Sampling	During the follow-up period of 60 months from 1985, 260 outpatients with cirrhosis were studied. The diagnosis was histological or clinical (or both); all patients were monitored for serum levels of AFP and checked for space-occupying lesions of the liver by several imaging modalities. When participants entered the study, ultrasonography (US) did not show HCC, and the level of serum AFP was less than 200 ng/mL. All patients were prospectively monitored by measurement of serum levels of AFP every 2 months and by US scanning every 3 months, as a rule. HCC was found in 62 patients. 7 patients found to have HCC within 6 months of entry were excluded because their tumours probably already existed at the time of enrolment. Age range not reported. Males 63%
Patient characteristics and setting	
Index tests	AFP: all patients were prospectively monitored by measurement of serum levels of AFP every 2 months. Cut-off values predefined at 20 ng/mL and 100 ng/mL
Target condition and reference standard(s)	HCC: HCC was found in 62 patients. Seven patients found to have HCC within 6 months of entry were excluded because their tumours probably already existed at the time of enrolment. Of the remaining 55 patients, the diagnosis of HCC was confirmed histologically in 18. It was confirmed clinically in 37 by typical tumour stains in hepatic angiography in 25 patients; by increases in AFP or abnormal prothrombin protein induced by vitamin K antagonist (PIVKA II, also called des-y-carboxy prothrombin), a specific marker of HCC; and by an increase in tumour size (detected by US) in the 12 other patients.

Oka 1994 *(Continued)*

	When participants entered the study, ultrasonography (US) did not show HCC, and the level of serum AFP was less than 200 ng/mL. The diagnosis was histological or clinical (or both); all patients were monitored for serum levels of AFP and checked for space-occupying lesions of the liver by several imaging modalities.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		

Oka 1994 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Oka 2001

Study characteristics	
Patient Sampling	Between 1996 and 1997, 663 patients with HCC were admitted to the nine participating hospitals; of these patients, 388 were newly diagnosed and were enrolled in the prospective study. As a control group, 212 patients (138 males and 74 females) from the same nine hospitals who had chronic hepatitis (CH) or liver cirrhosis (LC) caused by the hepatitis B and/or C viruses were enrolled in that period. Age range: 53-72. Males 70%
Patient characteristics and setting	
Index tests	AFP: the serum AFP concentrations were determined at each hospital by using commercially available kits. Cut-off values predefined at 20 ng/mL
Target condition and reference standard(s)	HCC: 102 participants were diagnosed histologically by a percutaneous liver tissue needle biopsy (26.3%). Of the remaining 286 participants (73.7%), comprehensive diagnoses were made based on ultrasonography, computed tomography scanning, angiography, and the other imaging techniques.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Oka 2001 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Oka 2001 *(Continued)*

Could the patient flow have introduced bias?	High risk

Okazaki 1984

Study characteristics			
Patient Sampling	245 outpatient asymptomatic for HCC with chronic liver disease. Patients with elevated AFP were included. Age range: 17-82. Males 70%		
Patient characteristics and setting			
Index tests	US equipment Toshiba Sonolayer SAL 20 A. No definition of positivity criteria; serum AFP with a cut-off value 20 ng/mL		
Target condition and reference standard(s)	CT angiography, follow-up		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Funded by a grant from Ministery of Health; no information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Okazaki 1984 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Omar 2017

Study characteristics	
Patient Sampling	Data of 2363 Egyptian patients with HCV-related chronic liver disease were reviewed. 1291 patients were diagnosed with HCC, while 1072 had HCV-related liver cirrhosis with no HCC on top. Focal hepatic lesions detected on US and/or rising levels of AFP were evaluated by CT or MRI. Age range and % of males not reported

Omar 2017 *(Continued)*

Patient characteristics and setting			
Index tests	AFP: it was found that serum AFP was able to diagnose HCC at the cut-off level of 11.9 ng/mL with sensitivity 68% and specificity 80.6%.		
Target condition and reference standard(s)	HCC: diagnosed by CT or MRI. Lesions showing hyperenhancement in arterial phase were diagnosed as HCC. Rising AFP assays were further evaluated by CT or MRI.		
Flow and timing	No information on timing between index test and reference standard		
Comparative			
Notes	All authors declared no conflicts of interest.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		

Omar 2017 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Omran 2016

Study characteristics	
Patient Sampling	A total of 88 consecutive Egyptian individuals attending the Tropical Medicine Department, Mansoura University hospitals, Mansoura, Egypt during the period from May 2012 to April 2013 were enrolled in this study. They were classified into 3 groups. The first group included 53 patients with hepatocellular carcinoma (HCC), the second group included 20 patients with liver cirrhosis, and the third group of 15 apparently healthy participants serving as control group were included. Patients with heart failure, kidney failure, rheumatoid arthritis, autoimmune liver diseases, hepatitis B virus, metabolic disorders, or other malignancies were excluded. Age range: 42-70. Males 71%
Patient characteristics and setting	
Index tests	AFP: the level of serum alpha fetoprotein was estimated by chemiluminescence, with IMMULITE (1000) AFP kit (Diagnostic Products Corporation; Los Angeles, CA, USA). AFP cut-off value 400 ng/mL
Target condition and reference standard(s)	HCC: the diagnosis of HCC was done according to American Association for the Study of Liver Diseases (AASLD) practice guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest
Methodological quality	

Omran 2016 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Omran 2020

Study characteristics	
Patient Sampling	The present case-control study examined 196 patients with chronic hepatitis C for an estimation study and 122 patients for the validation study. The estimation group included 104 patients with HCC and 92 with non-malignant liver diseases recruited from the Tropical Medicine Department at Mansoura University Hospi-tals, Mansoura, Egypt, between December 2016 and October 2017. Patients with kidney failure, cardiovascular disease, rheumatoid arthritis, autoimmune liver diseases, hepatitis A or B viruses, bilharzial infection, or other causes of liver diseases were excluded. In addition, patients with other causes of thrombocytopenia, such as typhoid, leukaemia, and deficiency of vitamin B12, as well as other causes of HCC or the presence of other malignancies were also excluded from this study. Age range not reported. Males 70%
Patient characteristics and setting	
Index tests	AFP and HCV antibody were evaluated using an immunofluorescence assay (IFA) by auto-analyser (Mini-Vidas;bioMérieux, Marcy L'Etoile, France). The predefined cut-off value: 400 ng/mL
Target condition and reference standard(s)	HCC: non-invasive methods were used for HCC diagnosis, according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines [20]. Tumours were measured using triphasic CTand/or dynamic MRI. Controls: the non-malignant group of patients included those with liver cirrhosis (n = 52) and liver fibrosis (n = 40). Liver cirrhosis was diagnosed based on biochemical, ultrasonographic and computed tomography imaging findings of splenomegaly or macronodular liver.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors have no conflicts of interest to declare."

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Omran 2020 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Omran 2020 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Ozkan 2011

Study characteristics	
Patient Sampling	158 participants were enrolled in the study between May and October 2009. The participants were divided into three groups: Group 1: 75 cirrhotic patients with HCC Group 2; 55 cirrhotic patients without HCC Group 3: 28 healthy controls Age range not reported. Males 67%
Patient characteristics and setting	
Index tests	AFP was tested using commercially available immunoassays utilising enhanced chemiluminescence at our hospital central laboratory. The upper limit of the normal level was 13 ng/mL.
Target condition and reference standard(s)	The diagnosis of HCC was made by histopathology. If histopathology was not available, the diagnosis was reached by two imaging modalities (ultrasound, magnetic resonance imaging or computed tomography) showing a vascular-enhancing mass. Diagnosis of cirrhosis was based on liver histology or clinical, laboratory, and imaging evidence of hepatic decompensation or portal hypertension.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest or funding

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Ozkan 2011 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Park 2017a

Study characteristics	
Patient Sampling	The participant cohort consisted of 298 HCC cases from the Digestive Disease Center at the Soonchunhyang University Seoul Hospital, which were newly diagnosed between

Park 2017a *(Continued)*

	October 2013 and March 2016 by retrospective design. Among them, 79 HCC patients were selected for inclusion in this study after applying the following exclusion criteria: the baseline serum level of AFP, AFPL3, or PIVKA-II was not obtained; presence of extrahepatic malignancy when HCC was diagnosed; previously treated for any type of malignancy before HCC was diagnosed; all other conditions with elevated AFP rather than liver disease; or fibrolamellar HCC which can show normal AFP. A further 77 patients with liver cirrhosis (LC) were selected in this study as a control group. LC was diagnosed based on a histological examination or clinical findings of portal hypertension. The LC patients in the control group had undergone imaging studies to exclude HCC. Age range: 48-70. Males 85%
Patient characteristics and setting	
Index tests	AFP: "alpha-fetoprotein, AFP-L3, and PIVKA-II were measured in the same serum samples using microchip capillary electrophoresis and a liquid-phase binding assay on an automatic analyser (mTAS Wako i30, Wako Pure Chemical Industries, Osaka, Japan). The measurement range was 0.3 to 2000ng/mL for AFP. All testing was conducted at the Soonchunhyang University Seoul Hospital by the same group of laboratory technicians, and none of the technicians was informed of the participant's status before testing. We defined positivity for the 3 biomarkers alone as follows: AFP > 10ng/mL, PIVKA-II > 40mAU/mL, and AFP-L3 > 10%. The cut-off value for serum AFP was 10 ng/mL since this is the setting used by our laboratory automatic analyser (Wako i30). Because the cut-off value of other devices in our hospital is 20ng/ mL, we also determined whether the diagnostic performance of the biomarkers changed for a AFP cut-off value of 20 ng/mL, and we also analysed the diagnostic performance of biomarkers for different cut-off values of AFP-L3 to verify the reproducibility of our study results."
Target condition and reference standard(s)	HCC: HCC was diagnosed based on histological findings or typical imaging characteristics as defined by the Korean Liver Cancer Study Group Guideline. Control group: the liver cirrhosis (LC) patients in the control group had undergone imaging studies to exclude HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors reported no conflicts of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Park 2017a *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Park 2017b

Study characteristics			
Patient Sampling	The study was aimed to investigate the role of AFP for HCC in patients with advanced liver cirrhosis waiting for liver transplantation. During 10 years, 2074 adult liver-surgery recipients were identified. They were divided into two groups as HCC and non-HCC. Age range not reported. Males 71%.		
Patient characteristics and setting			
Index tests	AFP: ROC curve analysis showed that AUC of AFP was 0,693 having a cut-off at 6,8 ng/mL with sensitivity of 64.5% and specificity of 64.5%.		
Target condition and reference standard(s)	HCC: all patients underwent orthotopic liver transplantation (OLT).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Park 2017b *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Park 2020

Study characteristics	
Patient Sampling	This study was a retrospective analysis of prospectively collected data All included patients: i) were aged ≥ 20 years; ii) were diagnosed histologically or radiologically with cirrhosis, with an estimated annual HCC risk > 5%; iii) had an Eastern Cooperative Oncology Group performance status of 0 or 1; and iv) had no previous history or current suspicion of HCC Age range: 29-77. Males 57%
Patient characteristics and setting	
Index tests	US examinations in the original study were performed by 4 board-certified abdominal radiologists specializing in liver imaging (So Yeon Kim, So Jung Lee, Hyung Jin Won, and Jae Ho Byun) using a convex probe (SC6-1, Supersonic Imagine SA; Aixplorer, France). The patient stay duration in the US room ranged from 15 to 20 min.

Park 2020 *(Continued)*

Target condition and reference standard(s)	Patients were evaluated by 1 to 3 rounds of surveillance tests, consisting of paired US and gadoxetic acid-enhanced MRI performed within 7 days at 6 month intervals.
Flow and timing	The reference standard was performed within seven days after the index test.
Comparative	
Notes	"The authors declare no conflicts of interest that pertain to this work."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Park 2020 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Passos-Castilho 2015

Study characteristics			
Patient Sampling	A total of 87 patients with chronic hepatitis B (CHB) were enrolled from 2012 to 2014 at the Hospital das Clínicas of the University of Sao Paulo School of Medicine, including 32 patients with HBV-HCC, 30 patients with HBV-liver cirrhosis (LC), and 25 patients with CHB. Age range: 19-85. Males 72%		
Patient characteristics and setting			
Index tests	AFP: cut-off values prespecified at 20 ng/mL and 200 ng/mL		
Target condition and reference standard(s)	HCC: HCC was diagnosed using imaging or histopathology techniques, in accordance with guidelines of the Brazilian Society of Hepatology.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare that they have no competing interests."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Passos-Castilho 2015 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Pateron 1994

Study characteristics	
Patient Sampling	Between January 1986 and December 1987, all patients with histologically-proven cirrhosis, hospitalised in the Hepatogastroenterology Unit of Jean Verdier Hospital were considered for inclusion in the study. At the end of hospitalisation, all patients with Child-Pugh's class A or B cirrhosis without detectable HCC (no focal lesions at US, serum AFP < 1 5 ng/mL and plasma DCP (di-

Pateron 1994 *(Continued)*

	rect current plasma) < 15 mU/ram) were prospectively included if voluntary consent was given and follow-up appeared feasible. Included patients were followed up until death or January 1990 (endpoint of the study). Screening protocol included clinical examination, determination of serum AFP and plasma DCP and US. Age range not reported. Males 58%
Patient characteristics and setting	
Index tests	AFP: determination of AFP was performed by immunoenzymology using a commercial kit (Roche Diagnostica Laboratory, Neuilly, France) (cut-off value < 15 ng/mL). US: US was performed by one of two experienced operators with convex-array real-time scanners (3.5 MHz, model EUB 410, Hitashi Medical Corp, Tokyo, Japan and Model SDR I SSOXP, Philips Ultrasound, Santa Ana, USA). A focal mass was searched for. The examination specified tumour echoic pattern, as well as thrombosis of the portal trunk or branches of the portal vein.
Target condition and reference standard(s)	HCC: when an anomaly in test results was detected, additional explorations were performed, in particular CT scan with injection of a contrast medium. When US showed a focal mass, guided biopsy was performed when possible.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Pateron 1994 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Paul 2007

Study characteristics	
Patient Sampling	A cross-sectional study was conducted at the Liver Clinic of the All India Institute of Medical Sciences (AIIMS), a tertiary care teaching hospital of India, between 2001 and 2004. Patients with cirrhosis (Child's A and B) of all aetiologies were eligible for enrolment into the study. The exclusion criteria were: patients with

Paul 2007 *(Continued)*

	Child's C cirrhosis, terminally ill patients, those unable to undergo diagnostic tests, those having a history of allergy to iodinated contrast media, or those with asthmatic bronchitis or severely deranged renal functions. A total of 301 patients with cirrhosis were enrolled into the study and subjected to detailed clinical evaluation and diagnostic work up. Of these, 195 were found to have only cirrhosis with no HCC, while 107 had cirrhosis with HCC. Out of these 107 HCC patients, triple-phase CT (TPCT) and AFP estimation could be done in 101, while US was done in 97 patient only. Age range not reported. Males 71%
Patient characteristics and setting	
Index tests	AFP: serum AFP was estimated by the Axsym system (Abbott Laboratory, Abbott Park, Ill., USA) based on the microparticle enzyme immunoassay technology. The best mix of sensitivity and specificity (77.2 and 78.1%, respectively) was seen at a level of 10.7 ng/mL. Other cut-off values included: 20, 50, 100, 200, 300, 400, 500 ng/mL.
Target condition and reference standard(s)	HCC: The gold standard for the diagnosis of HCC was either a positive fine needle aspiration cytology (FNAC) or any two of the following: AFP > 300 ng/mL, arterialisation on any of the imaging techniques, i.e. TPCT or MRI.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared that they had no financial conflict of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Paul 2007 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Piciocchi 2013

Study characteristics	
Patient Sampling	The study included a total of 142 consecutive patients recruited at the Department of Surgery, Oncology and Gastroenterology (DiSCOG) of Padua University, Italy (66 HCC patients, 35 liver cirrhosis patients, 41 patients with chronic hepatitis). Ongoing interferon treatment was an exclusion criteria; previous treatment with no response or relapse was accepted. Age range: 45-78. Males 70%
Patient characteristics and setting	
Index tests	AFP: AFP levels were determined by immunoenzymatic chemiluminescence; the cut-off value for normal AFP levels (20 ng/mL) was chosen on the basis of the European Association for the Study of the Liver (EASL) guidelines and on the data reported in the majority of the studies on the topic. Regarding AFP, the cut-off value for discriminating HCC from CH and CIRR, taken together, was 14 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was confirmed by either histology with the Edmonson grading system [29] or based on the European Association for the

Piciocchi 2013 *(Continued)*

	Study of the Liver (EASL) guidelines for HCC management published in 2001 with typical hypervascular lesions identified using one imaging technique and increased AFP levels. In the absence of diagnostic AFP, hypervascular lesions > 2 cm in patients with liver cirrhosis were confirmed by two imaging techniques (spiral computed tomography or nuclear magnetic resonance). After 2010, the revised AASLD guidelines were adopted, and diagnosis was made for lesions > 1 cm in the presence of one imaging technique showing typical arterial enhancement and venous washout.
Flow and timing	No information between index test and reference standard
Comparative	
Notes	Authors declared no conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		

Piciocchi 2013 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Pinero 2015

Study characteristics	
Patient Sampling	A total of 1502 adult (> 17 years of age) liver transplantations (LTs) were performed in Argentina between 1 June 2005 and 31 December 2011. During the same period, 763 adult LTs were consecutively performed at four LT Argentine centres. From this cohort, 643 adult patients with liver cirrhosis who had a first LT were included in the analysis. Five transplanted patients had HCC on a noncirrhotic liver and were excluded from the final analysis. As established by international guidelines, pre-LT monitoring of HCC was performed in all patients using US with or without a serum α-fetoprotein (AFP) assay every 6 months (= 180 days). US performance during waiting list was analysed after excluding those patients in whom HCC was diagnosed before being included in the waiting list or during transplant pre-evaluation (n = 71). Of 572 patients with liver cirrhosis, 58 had HCC. Age range: 51-67. % of males not reported
Patient characteristics and setting	
Index tests	US. No detailed description
Target condition and reference standard(s)	"All participant LT programs had a common standardised method to examine the explanted liver and were sliced at 5 mm to 10 mm thickness. In addition, the senior liver pathologist from each centre performed macroscopic and microscopic evaluation of each nodule of all the explants to characterise tumour biology including background fibrosis and inflammation, number and diameters (cm) of HCC nodules, presence of microvascular invasion (Mvi), and nuclear grade using the modified Edmonson Steiner grading system."

Pinero 2015 *(Continued)*

Flow and timing	The median time from the last screening image (US) to transplantation was 2.1 months (IR 0.6–4.6 months).
Comparative	
Notes	Authors reported no conflicts of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Pinero 2015 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		High
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Pompili 2003

Study characteristics	
Patient Sampling	131 patients with liver cirrhosis and HCC consecutively observed in our institution between 1995 and 2000 were included in the present study (HCC group). From 1998 to 2000, we also enrolled 59 cirrhotic patients without HCC (CIR group). Age range: 24-84. Males 68%
Patient characteristics and setting	
Index tests	AFP: serum levels of AFP were assessed by using a microparticle enzyme immunoassay performed with commercially available kits (AxSYM AFP system; Abbott Laboratories, Abbott Park, IL, USA). Normal values for adults ranged from 5 ng/mL to 15 ng/mL. The ROC curve analysis identified 20 ng/mL as the best discriminator between HCC and cirrhotic patients.
Target condition and reference standard(s)	HCC: the definitive diagnosis of HCC was based on cytology and/or histology of ultrasound-guided fine-needle biopsies in 99 cases, and unequivocal CT findings in 32 cases. Control group: the 59 cirrhotic patients without HCC (CIR group) had been biopsy-diagnosed in 21 cases; the other 38 had ultrasound signs of cirrhosis and/or ultrasound or endoscopic evidence of portal hypertension with laboratory findings indicative of chronic liver disease. None presented focal hepatic lesions on ultrasound examination.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest
Methodological quality	

Pompili 2003 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Poon 2001

Study characteristics			
Patient Sampling	Chinese patients who attended the Joint Hepatoma Clinic at the Prince of Wales Hospital were enrolled in this study. Serum levels of AFP, albumin, A1AT, A2MG, thyroxine-binding globulin (TBG), and transferrin were determined in 65 patients with HCC (HCC group) and 51 patients with liver cirrhosis only (LC group). Age range: 16-82. Males 79%		
Patient characteristics and setting			
Index tests	AFP: serum levels of AFP were measured by microparticle EIA (MEIA, Abbott Laboratories, Chicago, Ill., USA). Cut-off values were prespecified: 200 ng/mL and 500 ng/mL.		
Target condition and reference standard(s)	All HCC cases were histologically confirmed. For the liver cirrhosis (LC) group, all the patients were followed for 18 months for any sign of HCC to exclude participants with asymptomatic HCC.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Poon 2001 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Porta 2008

Study characteristics	
Patient Sampling	Three groups were studied: Group 1: 30 patients (21 males and 9 females, median age: 62.3 years, range: 50–74) affected with histologically-proven HCC. Group 2: 30 age- and sex-matched hepatitis B virus and/ or hepatitis C virus-related cirrhotic patients with no histologic evidence of cancer Group 3: 30 age- and sex-matched healthy volunteer controls, with no evidence of liver disease and/or of neoplasm. Age range: 50-74. Males 70%
Patient characteristics and setting	
Index tests	AFP: serum A1FP was evaluated using a commercially available kit (ADVIA Centaur System, Bayer Healthcare, Tarrytown, NJ), and the results were expressed as Ul/mL. The 'optimal' (closest to the upper left corner) cut-off value was 14 UI/mL for AFP titers.

Porta 2008 *(Continued)*

Target condition and reference standard(s)	HCC was histologically proven. LC patients had no histological evidence of HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Porta 2008 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Pote 2015

Study characteristics	
Patient Sampling	Patients, who underwent liver resection (LR) or liver transplantation (LT) for HCC between 2004 and 2011 at Beaujon Hospital, and for whom preoperative serum samples were available, were retrospectively included. Staging was determined according to the Barcelona Clinic Liver Cancer (BCLC) system. Very early stage HCC (BCLC stage 0) was defined as a single lesion 62 cm and was histologically sub-classified into early and progressed HCC, according to pathological criteria established by the international consensus group for hepatocellular neoplasia. Controls were patients with advanced chronic liver disease (CLD) at the stage of cirrhosis (F4 according to the METAVIR classification) established by liver biopsy, and enrolled during the same period as HCC cases. The study included a total of 128 participants: 43 controls and 85 HCC patients. Patients who received vitamin K or warfarin were excluded. Age range: 43-66. Males 89%
Patient characteristics and setting	
Index tests	AFP: AFP was determined by an automated system (Elecsys 2010, Roche). To determine the optimal cut-off value for PIVKA-II and AFP in diagnosis of HCC, receiver operating characteristic (ROC) curves were constructed, and the area under the curve (AUC) was calculated.
Target condition and reference standard(s)	HCC: histology and pathology Control group: control patients were followed in the Department of Hepatology (Beaujon Hospital), and US or computed tomography was performed every six months to exclude HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Pote 2015 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Pote 2015 *(Continued)*

Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Powell-Jackson 1987

Study characteristics	
Patient Sampling	The studied patients comprised of four groups: Group 1: 14 patients with HCC and cirrhosis Group 2: 13 with HCC but no evidence of chronic liver disease Group 3: 53 with cirrhosis and no HCC Group 4: 31 with neither cirrhotic nor malignant liver disease Histological confirmation of each diagnosis was obtained by liver biopsy within 4 weeks of the scanning procedures. Age range not reported. Males 71%
Patient characteristics and setting	
Index tests	US: the ultrasound scans were carried out and reported on by several radiologists without knowledge of the histology or other scan result. Ultrasound was carried out using a real-time sector scanner (Diasonics DRF1) using a 3.5 MHz probe. Hepatocellular carcinoma was suggested on ultrasound by the presence of single or multiple space occupying lesions with altered reflectivity in comparison with the remainder of the liver parenchyma.
Target condition and reference standard(s)	HCC: histological confirmation of each diagnosis was obtained by liver biopsy within 4 weeks of the scanning procedures. Control group: the absence of HCC was confirmed in each case by prolonged follow-up (minimum 9 months) or at autopsy in those dying earlier.
Flow and timing	Histological confirmation of each diagnosis was obtained by liver biopsy within 4 weeks of the scanning procedures.
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		

Powell-Jackson 1987 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Qi 2020

Study characteristics	
Patient Sampling	Patients with HCC treated in Gansu Provincial Hospital from 2016 to 2018 were included.

Qi 2020 *(Continued)*

	The inclusion criteria for patients with HCC were as follows: (a) 18-85 years old; (b) patients with pathologically-confirmed HCC; (c) patients meeting the Chinese guidelines Standardization of Diagnosis and Treatment for Hepatocellular Carcinoma (2017 Edition) The exclusion criteria for patients and controls were as follows: (a) participants with missing laboratory detection data; (b) participant with missing clinical and medical history key data; (c) participant with severe haemolysis, microbial contamination or jaundice; (d) participant that did not meet the requirements for sample collection or treatment; and (e) participant withdrawing from the trial based on the medical consideration by investigators patients with non-viral liver diseases (including autoimmune liver disease, drug-induced liver injury, and fatty liver) and hepatitis (mainly hepatitis B and hepatitis C) were included in the chronic disease group Age range: 35-66. Males 80%
Patient characteristics and setting	
Index tests	AFP levels were measured in microparticle chemiluminescence instrument (Abbott). No pre-definition of a cut-off value
Target condition and reference standard(s)	HCC: Chinese guidelines 'Standardization of Diagnosis and Treatment for Hepatocellular Carcinoma' (2017 Edition): (a) according to CT, MRI or ultrasound results, typical imaging lesions of HCC are seen, and typical blood flow changes occur in the lesions (b) CT, MRI, or ultrasound suggest suspected small nodules, which are confirmed by positron emission tomography (PET) examination Controls: no definition
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors have no conflicts of interest to be declared."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Qi 2020 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Qi 2020 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Raedle 1995

Study characteristics			
Patient Sampling	Between April 1993 and July 1994, 147 consecutive patients (88 men and 59 women) with HCV-related chronic hepatitis (CH) attending a specialised out-patient clinic at the University Hospital of Frankfurt/Main, Germany. Patients with any evidence of other causes of liver cirrhosis (LC) or HCC, such as alcoholic liver disease, haemochromatosis, autoimmune hepatitis, primary biliary cirrhosis, or alpha-1 antitrypsin deficiency were not included into the study. A histologically-confirmed hepatocellular carcinoma was diagnosed by routine ultrasound and CT scan in 7/147 patients (4.8%). All patients with HCC had coexisting liver cirrhosis. Age range: 18-74. Males 60%		
Patient characteristics and setting			
Index tests	AFP: quantitative determination of a-fetoprotein (AFP) was performed by a commercially available standard ELISA kit (Enzymun-Test, Boehringer, Mannheim, Germany). In this two-step sandwich assay AFP levels > 20 ng/mL are considered elevated.		
Target condition and reference standard(s)	HCC: a histologically-confirmed hepatocellular carcinoma was diagnosed by routine ultrasound and CT scan in 7/147 patients (4.8%).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		

Raedle 1995 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Raedle 1998

Study characteristics	
Patient Sampling	Blood samples were drawn from 711 consecutive patients with chronic liver disease of various aetiology referred to our outpatient clinic between June 1994 and May 1996. 75 cases of HCC were found.

Raedle 1998 *(Continued)*

	Age range: 14-87. Males 59%
Patient characteristics and setting	
Index tests	AFP: quantitative determination of AFP in all patients was performed by an enzyme immunological assay (Boehringer Mannheim, Germany). In this two-step sandwich assay, AFP levels > 20 ng/mL were considered elevated.
Target condition and reference standard(s)	HCC: in 52 of 75 cases (69.3%) with HCC the diagnosis was histologically proven. No tissue samples were obtained from patients presenting with clinically advanced cancer and HCC typical AFP elevations or imaging findings.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Raedle 1998 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Raff 2014

Study characteristics	
Patient Sampling	The cross-sectional study included 366 patients with cirrhosis, of whom supposedly all had AFP sampled. 163 patients received US as screening test. The AFP analysis included 356 patients - 10 patients excluded without explanation. Medical charts of patients with cirrhosis seen at a single tertiary referral centre (2007-2011) were reviewed. Among other data, use of and findings from CT or MRI scan within 6 months of receiving US were recorded for patients who had US as the initial imaging. Age range not reported. Males 71%
Patient characteristics and setting	
Index tests	AFP cut-off level 20 ng/mL; US no specification
Target condition and reference standard(s)	HCC: HCC was diagnosed based on standard criteria on CT or MRI scan. Of 163 patients receiving US, 72 received follow-up CT/MRI scan within 6 months of US examination.
Flow and timing	Use of and findings from CT or MRI scan within 6 months of receiving US was recorded for patients who had US as the initial imaging. No information for AFP. The AFP analysis included 356 patients. 10 patients excluded without explanation

Raff 2014 *(Continued)*

Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Raff 2014 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Reichl 2015

Study characteristics	
Patient Sampling	This is a case-control study which included serum samples from HCC patients (311) as well as healthy (125) and cirrhotic (30) controls from Shanghai, Vienna, Brno, and Hong Kong. The analysis included HCC patients and cirrhotic controls. Exclusion criteria were alterations in liver serology, viral or nonviral liver disease, as well as other malignancies. Age range not reported. Males 82%
Patient characteristics and setting	
Index tests	AFP: for AFP, the clinically well-established cut-off value of 20 ng/mL was used.
Target condition and reference standard(s)	HCC: all patients were diagnosed by ultrasound, computed tomography or magnetic resonance imaging, AFP and liver enzyme serology, and histopathologically confirmed by two individual board certified pathologists after surgical resection. Controls: people with liver cirrhosis (controls) were histopathologically confirmed and screened for tumour formation by ultrasound, computed tomography, or magnetic resonance imaging.
Flow and timing	No information on interval between index test and reference standard. Of 311 HCC patients included, 309 had available AFP values.
Comparative	
Notes	Conflicts of interest: K.S. received travel grants from Roche, MSD and Novartis as well as speaker honorarium from Roche and Biotest.
Methodological quality	

Reichl 2015 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Ricco 2018

Study characteristics	
Patient Sampling	"We studied retrospectively 388 consecutive patients with liver cirrhosis, enrolled in three Italian Hepatology centres. 258 cirrhotic patients with HCC (diagnosis performed from 2010 to 2015), for whom a serum sample at the time of diagnosis was available; 2) 130 cirrhotic patients on ultrasound (US) surveillance for at least 12 months, without evidence of HCC and with a serum sample available at the beginning of their follow-up." Age range 33-88. Males 79%
Patient characteristics and setting	
Index tests	"Quantitative measurements of AFP were performed on sera stored at -20° C since they were obtained at the time of diagnosis in 258 HCC patients or at a single point evaluation in 130 cirrhotic patients without HCC during their surveillance follow-up. AFP serum levels were measured using fully automated chemiluminescent enzyme immunoassays (CLEIA) on Lumipulse G1200 (Fujirebio Inc, Tokyo, Japan). Several fixed cut-off values were used, both overall and according to the aetiology of CLD. The thresholds used as cut-offs were arbitrarily chosen in accordance to previous studies [23–26]: 10–20–100 and 400 ng/mL for AFP."
Target condition and reference standard(s)	Liver cirrhosis was diagnosed by clinical, biochemical, and imaging data (presence of US signs and transient elastography > 13 kPa) or liver biopsy: all the patients underwent every 6-month US surveillance during a median follow-up of 25.2 months. HCC diagnosis and staging were performed according to European Association for the Study of the Liver (EASL) guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Conflict of interest: none

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Ricco 2018 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Saada 1997

Study characteristics	
Patient Sampling	From July 1993 to December 1994, 50 patients underwent a first elective orthotopic liver transplantation (OLT) at our centre for end-stage chronic liver disease and were included in this study. Preoperative imaging included hepatic ultrasonography (ATL, Ultramark 9) and CT (Somatom DR2, Siemens, Erlangen, Germany)

Saada 1997 *(Continued)*

	consisting of 1 cm contiguous sections through the entire liver, prior to and after 75 mL of intravenous contrast agent. Of the 50 cases considered for inclusion, there were protocol violations in 11 due to a donor liver becoming available after Lipiodol administration but prior to iodised oil computed tomography (IOCT) examination. Complete pre-OLT imaging was available in the remaining 39 patients (14 women, 25 men). Age range not reported. Males 64%
Patient characteristics and setting	
Index tests	Abdomnal US (ATL, Ultramark 9), no definition of positivity criteria
Target condition and reference standard(s)	Following transplantation the explant liver was cut into 10 mm slices. Each slice was closely inspected for atypical nodules according to a standard protocol.
Flow and timing	The median time between iodised oil computed tomography (IOCT) and orthotopic liver transplantation (OLT) was 20 days (range 10 days – 80 days).
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Saada 1997 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Sadeghi 2015

Study characteristics	
Patient Sampling	The following parameters were measured in 139 cirrhotic patients (aged 52.0 ± 11.2 years, 32 female, 61 cirrhotic HCC positive and 78 cirrhotic HCC negative) who underwent deceased donor liver transplantation between January 2008 and April 2011. Age range: 39-62. Males 77%
Patient characteristics and setting	
Index tests	AFP: the most sensitive cut-off values were calculated by receiver operating curve (ROC) analysis.
Target condition and reference standard(s)	HCC: HCC diagnosis was confirmed by pathological reports.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared no conflict of interest.
Methodological quality	

Sadeghi 2015 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Sadik 2019

Study characteristics	
Patient Sampling	We performed this case-control study from April 2017 to June 2018 on 81 participants divided into 3 groups. Group (1) included 30 patients having HCC, group (2) included 31 patients having liver cirrhosis (LC) secondary to hepatitis C virus (HCV) infection (HCV-related LC) from preliminary 80 patients with liver cirrhosis, selected from the inpatients of Kasr Al Ainy University Hospital, Internal Medicine Department and Group (3) included 20 healthy age-matched control participants. We excluded patients with liver cirrhosis secondary to hepatitis B virus (HBV) infection, autoimmune, metabolic liver diseases, and on hepatotoxic drugs. Age range not reported. Males 67%
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. No predefinition of a cut-off value
Target condition and reference standard(s)	HCC diagnosed clinically and radiologically by triphasic abdominal CT as recommended by the European Association for the Study of the Liver (EASL) guidelines. Control: no definition
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare no conflict of interest, financial, or otherwise."

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Sadik 2019 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Sadik 2019 *(Continued)*

Could the patient flow have introduced bias?	High risk

Saitta 2017

Study characteristics	
Patient Sampling	90 cirrhotic patients who had evidence of liver nodule(s) at US examination for the first time and who consecutively attended the liver unit of the University Hospital of Messina from November 2011 to October 2013 were enrolled. All of them underwent blood sampling within 1 week before or after the US identification of liver nodules, and the corresponding serum samples were aliquoted and stored at 80°C until testing. Age range: 52-79. Males 72%
Patient characteristics and setting	
Index tests	AFP: AFP serum levels were measured on a Lumipulse G1200 (Fujirebio Inc.), using the LUMIPULSE G AFP-N kit (Fujirebio Tokyo, Japan), respectively, according to the manufacturer's instructions. All tests were performed in duplicate. To determine the optimal cut-off value for PIVKA-II and AFP in the diagnosis of HCC, receiver operating characteristic curves were constructed using all possible cut-offs for each assay. Receiver-operating characteristic curves were plotted to identify PIVKA-II and AFP cut-off values that would best distinguish cirrhotic patients with HCC nodules from patients with regenerative/dysplastic nodules. The optimal cut-off was 60 mAU/mL for PIVKA-II and 6.5ng/mL for AFP.
Target condition and reference standard(s)	All patients were followed up for at least 18 months after US nodule (s) detection through imaging techniques – contrast-enhanced computed tomography and/or magnetic resonance imaging – and/or nodule needle biopsy performed according to the American Association for the Study of Liver Disease guidelines for HCC management.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors have no funding and conflicts of interest to disclose."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		

Saitta 2017 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Salem 2013

Study characteristics	
Patient Sampling	This study was conducted on 60 patients (after approval of the ethical committee); they were selected from the Tropical Medicine Depart-

Salem 2013 *(Continued)*

	ment, Cairo University. Patients (aged 40 to 70 years old) were divided as follows: Group I: 30 patients with hepatocellular carcinomas (proved by histopathology or combined triphasic CT and elevated alpha-fetoprotein) on top of hepatitis C virus infection as diagnosed by seropositivity for HCV antibodies Group II: 30 patients with HCV infection as diagnosed by seropositivity for HCV antibodies. Patients with other chronic liver diseases (for example, hepatitis B virus (HBV)), patients with bony lesions or inflammatory diseases, and patients with poorly controlled diabetes mellitus or systemic hypertension were excluded. Age range: 40-70. Males 77%
Patient characteristics and setting	
Index tests	Serum alpha fetoprotein assayed using enzymatic immunochemiluminesent using IMMULITE (Semeins). No cut-off value predefined
Target condition and reference standard(s)	- hepatocellular carcinoma (proved by histopathology or combined triphasic CT and elevated alpha-fetoprotein) on top of hepatitis C virus (HCV) - patients with HCV infection as diagnosed by seropositivity for HCV antibodies
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"All authors disclose that there are not any financial arrangement(s) they may have with any company related to the submitted manuscript or with a company making a competing product."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Salem 2013 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sanai 2010

Study characteristics	
Patient Sampling	"We conducted a case-control study of patients with HCC and cirrhosis whose clinical records were available for retrospective reviews at Riyadh Military Hospital (RMH) and King Khalid University Hospital (KKUH). Patients with HCC were identified by screening individual hospitals' computer-based databases and retrieving the results of all serum AFP performed from January 2006 to March 2008. In total, 210 treatment-naive patients. A total of 199 unselected, consecutive, control patients with cirrhosis were identified. As a control group, another 197 biopsy-proven, noncirrhotic chronic hepatitis patients with a serum AFP level available within 6 months of the liver biopsy. Four patients were labelled as HCC, however they did not fulfil the diagnostic criteria described above, and therefore were excluded from the analysis. We did not utilise serum AFP as one of the diagnostic criteria of HCC for the 206 patients included in the analysis in order to exclude incorporation bias." Age range: 13-93. Males 61%

Sanai 2010 *(Continued)*

Patient characteristics and setting	
Index tests	AFP: AFP was measured by a conventional immunoassay (Elecsys 2010, Roche Diagnostics GmbH, Mannheim, Germany). All AFP measurements in HCC cases were recorded prior to any therapy for HCC, cirrhosis or chronic hepatitis. Characteristics of test procedure (sensitivity, specificity, PPV, NPV, likelihood ratios, receiver operating characteristic (ROC) curve, and area under the curve) were used to evaluate the optimal cut-off value for AFP.
Target condition and reference standard(s)	HCC: "the diagnosis of HCC was established on the presence of hepatic lesions with typical arterial hypervascularisation and washout in the early or delayed venous phase on liver CT and/or MRI. All imaging studies were read by radiologists with extensive expertise in liver radiology. All patients underwent either CT liver and/or MRI. Needle aspiration or histological sampling was obtained only in conditions when non-invasive parameters were not diagnostic. We did not utilize serum AFP as one of the diagnostic criteria of HCC for the 206 patients included in the analysis in order to exclude incorporation bias. Control liver cirrhosis group: HCC was excluded by imaging studies [US, CT, and/or MRI], one of which must have been performed at least 6 months following the measurement of AFP."
Flow and timing	The median time between AFP and diagnostic imaging study was 50 days (range 1–364 days).
Comparative	
Notes	No conflicts of interest reported

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		

Sanai 2010 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sarwar 2014

Study characteristics	
Patient Sampling	This study was conducted at the Department of Medicine, The King Edward Medical University, Lahore, from November 2007 to August 2011. Consecutive patients with HCC presenting the study centre were enrolled (173 cases). People included as controls were 102 consecutive patients with cirrhosis without evidence of HCC. Patients with suspicion of ovarian or testicular malignancy on examination or diagnostic workup were excluded. Age range: 45-68. Males 65%

Sarwar 2014 *(Continued)*

Patient characteristics and setting	
Index tests	AFP: "ROC was used to determine area under curve (AUC) and cut-off value of AFP with best possible sensitivity and specificity. We used cut-off values of 20, 50, 100, 200, and 400 ng/mL for diagnosis of HCC, as mentioned in previous studies. AFP level with best possible sensitivity and specificity for diagnosing HCC was determined using ROC curve and it was 20.85 ng/mL with sensitivity of 72% and specificity of 86.3%."
Target condition and reference standard(s)	HCC: "diagnosis of HCC was made in accordance with AASLD guidelines. Control group: all patients had serum alpha-fetoprotein and abdominal US to exclude HCC. Patients with an elevated AFP (> 20 ng/mlL at enrolment were required to have a CT or MRI showing no lesion suggestive of HCC. Cirrhotic patients with nodules larger than 1 cm on US underwent biphasic CT abdomen or dynamic contrast enhanced MRI. If the appearance was typical of HCC i.e. hypervascular in arterial phase with washout in the portal venous phase, lesion was regarded as hepatocellular carcinoma. But if the findings were not characteristic or the vascular profile was not typical, a second contrast enhanced study with other imaging technique was performed or the lesion was biopsied. Those with lesion less than 1 cm were not included and were advised follow-up with repeat ultrasonography after 6 months."
Flow and timing	Interval between index test and reference standard not mentioned
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Sarwar 2014 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sassa 1999

Study characteristics	
Patient Sampling	61 patients with small (< cm^2) HCC, and 134 controls (59 with chronic hepatitis and 75 with cirrhosis) Age range not reported. Males 71%
Patient characteristics and setting	
Index tests	Serum AFP measurement by conventional radioimmunoassay; cut-off value 200ng/mL
Target condition and reference standard(s)	US CT histology, follow-up
Flow and timing	No information on interval between index test and reference standard
Comparative	

Sassa 1999 *(Continued)*

Notes	No information on funding or conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Sassa 1999 *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sato 1993

Study characteristics			
Patient Sampling	The study included 361 cirrhotic patients who were admitted to the hospital between 1980 and 1990 and were followed with measurements of AFP and US or CT of the liver every three months. 33 patients were found to have HCC. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	Serum AFP concentrations were measured in duplicate by radioimmunoassay with kits obtained from Dainabot Radioisotope (Tokyo, Japan). AFP cut-off prespecified at 30 ng/mL		
Target condition and reference standard(s)	HCC: the diagnosis of HCC was based on histological findings in tissue obtained at the time of surgery or US-guided tumour biopsy and on US, CT, and angiography.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Sato 1993 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Seo 2015

Study characteristics	
Patient Sampling	"A total of 1255 patients with CHB were retrospectively included at Hallym University Medical Center, Seoul, Korea, from January 2005 to December 2012. All patients who enrolled in this study demonstrated positivity for hepatitis B surface antigen for at least 6 months. A total of 1255 patients were divided into three subgroups: (1) non-cirrhotic CHB (G1, n = 879); (2) cirrhosis without HCC (G2, n = 219); and (3) HCC (G3, n =157). The exclusion criteria were as follows: the patients who (1) were positive for other markers of hepatitis such as hepatitis C virus or human immunodeficiency virus; (2) were heavy alcoholics (more than 80 g of ethanol daily); and (3) were taking warfarin or antibiotics that might influence the metabolism of vitamin K." Age range: 17-97. Males 66%
Patient characteristics and setting	

Seo 2015 *(Continued)*

Index tests	AFP: the serum AFP concentrations were determined with a commercially available electrochemiluminescence immunoassay kit (Elecsys AFP immunoassay, Roche, Mannheim, Germany). To find the optimal cut-off value of AFP and PIVKA-Ⅱ in the diagnosis of HCC, the receiver operating characteristic (ROC) curves were plotted using all possible cut-off values for each assay. The areas under the ROC (AUROC) curves of PIVKA-Ⅱ, AFP and the combination of the two were calculated and compared. Youden's index was calculated as an index of sensitivity and specificity. A P value < 0.05 was considered significant. The best cut-off value for AFP was 10 ng/mL.
Target condition and reference standard(s)	HCC: all patients with HCC were newly diagnosed, and the diagnosis of HCC was based on liver histology or appropriate imaging characteristics as defined by accepted guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Seo 2015 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shaheen 2015

Study characteristics	
Patient Sampling	This study was conducted on 100 individuals who were divided into 3 groups; group 1 included 40 patients with newly diagnosed HCC, group 2 included 30 patients with liver cirrhosis (LC), and group 3 included age- and sex-matched apparently healthy participants serving as a control group. Patients with previous HCC treatment and liver tumours other than HCC and those with Barcelona Clinic Liver Cancer (BCLC) stage D were excluded from the study. Age range not reported. Males 69%.
Patient characteristics and setting	
Index tests	AFP: ROC curve was performed for the best cut-off point to differentiate between HCC group and LC group using MDK and AFP. The best cut-off value was determined at 88.5 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was confirmed according to the 2011 American Association for the Study of Liver Diseases (AASLD) practice guidelines.
Flow and timing	No information on interval between index test and ref. standard

Shaheen 2015 *(Continued)*

Comparative	
Notes	The authors declare that there is no conflict of interests regarding the publication of this paper.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Shaheen 2015 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shaheen 2018

Study characteristics			
Patient Sampling	This study was conducted on 120 Egyptian adults who were divided into three groups. Group I: 40 patients with HCC, post HCV infection Group II: 40 patients with HCV infection who were further subdivided into 2 groups according to presence of cirrhosis: 20 patients with cirrhotic liver and 20 patients with non-cirrhotic liver. Group III: 40 age- and sex-matched healthy individuals as a control group. Patients with chronic HBV infection, patients who received anti-viral therapy for HCV infection or any loco-regional therapy for HCC were excluded. Age range 43-67. Males 51%		
Patient characteristics and setting			
Index tests	AFP: cut-off value of 400 ng/mL prespecified		
Target condition and reference standard(s)	HCC: diagnosis was done according to European Association for the study of the liver (EASL) guidelines. Group I and II patients were subjected to ultrasound to document the presence of cirrhosis and hepatic focal lesion(s). Only patients with hepatic focal lesion(s) underwent Triphasic abdominal CT for the diagnosis of HCC.		
Flow and timing	No data on interval between index test and reference standard		
Comparative			
Notes	All authors have no conflict of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Shaheen 2018 *(Continued)*

Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shang 2012a

Study characteristics	
Patient Sampling	Plasma samples were collected following informed consent from patients enrolled at the University of Michigan (Ann Arbor, MI) (Co-

Shang 2012a *(Continued)*

	hort 1). The cohort included 40 HCC patients, 73 cirrhosis patients, 32 with CHC, and 28 healthy controls. Age range 22-77. Males 67%
Patient characteristics and setting	
Index tests	AFP: cut-off prespecified at 20 ng/mL
Target condition and reference standard(s)	HCC was diagnosed according to the American Association for the Study of Liver Diseases (AASLD) practice guidelines.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Potential conflict of interest: nothing to report

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Shang 2012a *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shang 2012b

Study characteristics			
Patient Sampling	The second cohort (cohort 2) included patients at the National Cancer Institute (NCI) of Thailand: 91 HCC patients, 23 with cirrhosis or CHB, and 25 healthy controls. Age range: 32-81. Males 76%		
Patient characteristics and setting			
Index tests	AFP: cut-off prespecified at 20 ng/mL		
Target condition and reference standard(s)	HCC: HCC diagnosis was based on a clinical algorithm, including imaging (i.e. ultrasonography [US] and computerised tomography) and biochemistry (i.e. AFP and liver-function enzyme testing).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"No conflicts of interest to report"		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Shang 2012b *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shariff 2010

Study characteristics			
Patient Sampling	A total of 43 people from Nigeria, in three cohorts, were recruited for study: 18 patients with radiologically-proven (ultrasound or computed tomography) HCC; 10 patients with clinically-confirmed cirrhosis with features of portal hypertension, but no HCC; and 15 healthy people from Nigeria as controls. Age range: 23-85. Males 71%		
Patient characteristics and setting			
Index tests	AFP: serum AFP was measured using an automated Siemens Immulite 2500 Analyzer (Deerfield). Cut-off prespecified: 20 IU/L (24.2 ng/mL)		
Target condition and reference standard(s)	HCC: US or CT		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Shariff 2010 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shariff 2016

Study characteristics	
Patient Sampling	Patients were recruited at six hospital sites around the UK: London, Manchester, Newcastle, Nottingham, Plymouth and Southampton. 13 patients with HCC and 25 with cirrhosis were included. Exclusion criteria included those patients not meeting the diagnostic criteria for HCC and cirrhosis, those patients with HCC who had undergone curative resection or transplant, patients co-infected with HIV virus and those samples identified as outliers on principal component analysis. Age range: 28-82. Males 66%
Patient characteristics and setting	
Index tests	AFP: cut-off values predefined at 20, 200 and 400 IU/mL (24,2, 242, and 484 ng/mL)
Target condition and reference standard(s)	HCC was diagnosed with two confirmatory imaging modalities and cirrhosis with histological and/or radiological confirmation.
Flow and timing	No information on interval between index test and reference standard.
Comparative	

Shariff 2016 *(Continued)*

Notes	"No conflicts of interest to declare"		

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Shariff 2016 *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sharma 2010

Study characteristics			
Patient Sampling	A total of 138 consecutive patients with liver disease (70 HCC; 38 cirrhosis; 30 chronic hepatitis) who attended the Liver Clinic from June 2006 to March 2009 at Post Graduate Institute of Medical Education and Research, Chandigarh, India and 30 healthy volunteers were included in the study. All patients were naive to treatment and did not receive any antiviral therapy for hepatitis B, or C, or HCC-directed therapy like TACE, RFA, PEI, or resection prior to inclusion. Age range: 26-70. Males 83%		
Patient characteristics and setting			
Index tests	AFP: PIVKA-II and AFP levels were measured in all the patients and healthy volunteers using commercially available kits according to the manufacturer's instructions. Kits for the plasma PIVKA-II and serum AFP levels were purchased from Diagnostica Stago, France, and Smart Diagnostics, Israel, respectively. Receiver operating characteristics (ROC) curves were constructed to compare the performance and also to set the optimal cutoff value of AFP and PIVKA-II.		
Target condition and reference standard(s)	HCC: histological confirmation or two concordant imaging studies with typical findings of HCC, which includes a high-density mass in the arterial phase and a low-density mass in the portal phase on dynamic computed tomography or magnetic resonance imaging (MRI).		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Sharma 2010 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shen 2012a

Study characteristics	
Patient Sampling	We recruited consecutive patients with HCC to a test cohort, from the Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, China, from December, 2008, to June, 2009. We also recruited consecutive patients with chronic hepatitis B virus (HBV) or liver cirrhosis and healthy controls from the Department of Infectious Disease, First Affiliated Hospital of Soochow University, Suzhou, China, from April to July, 2009. The test

Shen 2012a *(Continued)*

	cohort included 831 patients (424 HCC patients, 98 with chronic hepatitis B (CHB), 96 patients with liver cirrhosis, and 213 healthy controls). Patients who had a history of other solid tumours were excluded from the study. Age range: 42-68. Males 80%
Patient characteristics and setting	
Index tests	AFP: AFP concentrations were measured with commercially available ELISA (R&D Systems), according to the manufacturer's recommendations. Cut-off value was prespecified at 20 ng/mL.
Target condition and reference standard(s)	HCC: HCC was defined on the basis of ultrasound, CT, or MRI characteristics and biochemistry (AFP serology and liver function enzymes), and was confirmed by histopathology, according to the American Association for the Study of Liver Diseases guidelines. Control group: patients with cirrhosis who had raised AFP concentrations were required to have undergone imaging by multiple methods (ultrasonography, CT, or MRI) and to have had no evidence of a hepatic mass for at least 3 months before enrolment.
Flow and timing	No information on interval between index test and reference standard; 98 CHB patients were included in the control group out of which 41 had AFP values available.
Comparative	
Notes	"The authors declare no conflicts of interest"

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		

Shen 2012a *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Shen 2012b

Study characteristics	
Patient Sampling	A validation cohort comprising patients with HCC, chronic HBV infection, and cirrhosis and healthy controls was recruited from Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China, from July 2010, to June, 2011. The validation cohort included 453 patients (209 HCC, 73 chronic hepatitis B (CHB), 72 liver cirrhosis, and 99 healthy patients). Patients who had a history of other solid tumours were excluded from the study. Age range: 45-69. Males 66%
Patient characteristics and setting	

Shen 2012b *(Continued)*

Index tests	AFP: AFP concentrations were measured with commercially available ELISA (R&D Systems), according to the manufacturer's recommendations. Cut-off value was prespecified at 20 ng/mL
Target condition and reference standard(s)	HCC: HCC was defined on the basis of ultrasound, CT, or MRI characteristics and biochemistry (AFP serology and liver function enzymes), and was confirmed by histopathology, according to the American Association for the Study of Liver Diseases guidelines. Control group: patients with cirrhosis who had raised AFP concentrations were required to have undergone imaging by multiple methods (ultrasonography, CT, or MRI) and to have had no evidence of a hepatic mass for at least 3 months before enrolment.
Flow and timing	No information on interval between index test and reference standard 73 CHB patients were included in the control group, out of which 55 had AFP values available.
Comparative	
Notes	"The authors declare no conflicts of interest."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Shen 2012b *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Sherman 1995

Study characteristics	
Patient Sampling	Authors have carried out a prospective study of HBV carriers in the greater Toronto area, using serum AFP and US as the screening tests for HCC. Individuals who tested positively for hepatitis B surface antigen for more than 6 months and who were over the age of 18 years were eligible. Between February 1989 and March 1994, 1069 chronic hepatitis B (CHB) carriers were referred to the Liver Cancer Screening Program. A total of 13 participants with HCC were identified. 538 participants were randomised to be screened with US and AFP (data for accuracy of US only is provided in this cohort). Age range: 27-51. Males 65%
Patient characteristics and setting	
Index tests	AFP: AFP assay (normal value < 5 ng/mL) was also performed by commercial kit (Abbott Laboratories). Cut-off prespecified at 20 ng/mL. US: patients who were randomised to US had high-resolution real-time US examination of the upper abdomen. US criteria for further evaluation: liver mass
Target condition and reference standard(s)	HCC: the diagnosis of HCC was confirmed by histological examination of tissue obtained from liver biopsy or surgical resection, or the combination of diagnostically increased AFP plus typical features on ultrasonography or computed tomography.
Flow and timing	No information on interval between index test and reference standard. In 11 women (10%), the increase in serum AFP levels was

Sherman 1995 *(Continued)*

	caused by pregnancy. These were excluded from specificity and sensitivity calculations because there was no uncertainty about the cause in these cases.
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Sherman 1995 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Shimizu 2002

Study characteristics	
Patient Sampling	This is a case-control study which included 56 liver cirrhosis patients with HCC and 34 liver cirrhosis patients without HCC. In the liver cirrhosis only group: three of the 39 patients who had developed HCC within one year and 2 patients with warfarin therapy were excluded. Age range: 35-84. Males 79%
Patient characteristics and setting	
Index tests	AFP: serum AFP levels were measured by EIA (TOSOH, Yamaguchi, Japan). Cut-off values predefined at 20, 100, and 200 ng/mL
Target condition and reference standard(s)	HCC: the diagnosis of HCC was based on histological findings in tissue obtained at the time of surgery (n = 6) or ultrasonography guided tumour biopsy (n = 25) in 31 patients. For the remaining 25 patients, the diagnosis was made by imaging modalities, such as ultrasonography, computed tomography, magnetic resonance imaging, and angiography, or was based on elevated serum concentrations of AFP or des-gamma-carboxyprothrombin (DCP). Liver cirrhosis control group: all of the patients were regularly checked at 1- or 2-month intervals, at 3-month intervals for ultrasonography and every 12 months for computed tomography in order to detect HCC.
Flow and timing	No information on interval between index test and reference standard

Shimizu 2002 *(Continued)*

Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Shimizu 2002 *(Continued)*

Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Shu 2010

Study characteristics			
Patient Sampling	This case-control study included 162 patients with HCC and 130 patients with LC and no HCC. Patients were recruited from First Affiliated Hospital of Guangxi Medical University (Nanning, China) between February 2007 and March 2009. Age range not reported. Males 78%		
Patient characteristics and setting			
Index tests	AFP: the clinically acceptable normal serum AFP was defined as < 20 ng/mL.		
Target condition and reference standard(s)	Liver cirrhosis and HCC diagnosis was confirmed by ultrasound imaging and biopsy.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		

Shu 2010 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Simão 2015

Study characteristics	
Patient Sampling	Patients followed between Jun 2013 and May 2014 at the Liver Disease Unit–Internal Medicine Department and Hepatic Transplantation Unit at Coimbra Hospital and University Centre were included. A total of 90 consecutively-observed patients with alcoholic cirrhosis (AC) were included and divided into two groups. Group I: 45 patients with AC Goup II: 45 patients with AC and HCC All patients had a history of alcohol intake > 60 g/day for more than 10 years. Other causes of liver disease (HBV, HCV, autoimmune and metabolic diseases) were excluded. Age range: 48-72. Males 99%

Simão 2015 *(Continued)*

Patient characteristics and setting	
Index tests	AFP: serum AFP levels were measured with the same sample by the chemiluminescence method using IMMULITE® 2000 AFP kit (Siemens Healthcare Diagnostics, Tarrytown, New York) according to the manufacturer's instructions. Receiver operating characteristics (ROC) analysis was used to evaluate the diagnostic value of AFP, and to identify the optimal threshold values. The sensitivity and specificity of AFP levels in HCC relative to AC group were 57.8 % and 93.3 %, respectively, at a cut-off value of 8.2 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was based according to the non invasive criteria of EASL–EORTC (European Organization for Research and Treatment of Cancer) Clinical Practice Guidelines on Management of HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare that they have no competing interests."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Simão 2015 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Singal 2012

Study characteristics	
Patient Sampling	Between January 2004 and September 2006, consecutive patients with cirrhosis were prospectively identified and entered into a surveillance program using ultrasound and AFP. Patients were enrolled from the University of Michigan (Ann Arbor, MI) General Hepatology or Liver Transplant outpatient clinics if they had Child-Pugh class A or B cirrhosis and absence of known HCC at the time of initial evaluation. Exclusion criteria included clinical evidence of significant hepatic decompensation (refractory ascites, grade III–IV encephalopathy, active variceal bleeding, or hepatorenal syndrome), co-morbid medical conditions with a life expectancy of less than 1 year, prior solid organ transplant, and a known extrahepatic primary tumour. HCC cases diagnosed within the first 6 months of enrolment (prevalent cases) were excluded. Age age: 24-82. Males 59%
Patient characteristics and setting	
Index tests	AFP and US: patients with an AFP level greater than 20 ng/mL or mass lesion on ultrasound underwent further evaluation.
Target condition and reference standard(s)	HCC: HCC was diagnosed using AASLD guidelines, and the Barcelona Clinic Liver Cancer (BCLC) system was used for tumour staging. For tumours greater than 2 cm in size, the diagnosis was made by the presence of a typical vascular pattern on dynamic imaging (arterial enhancement and washout on delayed images) or an AFP level greater than 200 ng/mL. For tumours with a maximum diameter of 1 cm to 2 cm, the diagnosis was made by the presence

Singal 2012 *(Continued)*

	of a typical vascular pattern on 2 dynamic imaging studies or histology. Absence of HCC was determined by imaging lacking any suspicious appearing masses within 6 months of enrolment. Patients with an AFP level greater than 20 ng/mL at enrolment were only included if computed tomography (CT) or MRI confirmed the absence of any suspicious masses within 3 months of enrolment.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No potential conflicts of interests were disclosed.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Singal 2012 *(Continued)*

DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Snowberger 2007

Study characteristics	
Patient Sampling	This was a retrospective study conducted at the Baylor Regional Transplant Center, including Baylor University Medical Center in Dallas, TX and Baylor All Saints Hospital in Fort Worth, TX, USA. The study group consisted of patients with cirrhosis who were discovered to have HCC, either before or at the time of orthotopic liver transplantation. Participants without HCC who were transplanted during the same time period served as controls. 2372 patients were approved for listing and underwent transplant at Baylor between January 1, 1988 and December 31, 2004. HCC was present in 239 (10.1%) patients who underwent transplantation. Age range: 17-32. Males 73%

Snowberger 2007 *(Continued)*

Patient characteristics and setting	
Index tests	"AFP: a normal AFP in our laboratory is less than 8.9 ng/mL."
Target condition and reference standard(s)	HCC: all cases of HCC identified by imaging before transplant were confirmed by pathologic examination of the explanted liver and these cases were defined as known. Cases only identified in the explant were labelled as incidental. All patients underwent liver transplantation.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Conflicts of interest: declaration of personal interests: Drs. Snowberger, Chinnakotla, Lepe, and Goldstein have no interests to declare. Ms. Peattie has no outside interests to declare. Dr. Klintmalm receives clinical research funding from AbSorber, Astellas, Genxyme, Isotechnika, Novartis, Pfizer, Roche, and Y's Therapeutics. Dr. Davis receives clinical research funding from Roche, Schering-Plough, Human Genome Science, and Vertex. Declaration of funding interests: This study was funded entirely by the Baylor Regional Transplant Institute"

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Snowberger 2007 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Son 2019

Study characteristics			
Patient Sampling	Prospective cohort of patients with cirrhosis Age range: 52-62. Male 57%		
Patient characteristics and setting			
Index tests	Ultrasound; positivity criteria according to US LI-RAD category		
Target condition and reference standard(s)	CT MRI pathological examination		
Flow and timing	MRI within 7 days, additional CT within 3 months		
Comparative			
Notes	Conflicts of interest present and reported		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Son 2019 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Son 2019 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Song 2002

Study characteristics			
Patient Sampling	Among 234 HCC patients diagnosed at the liver clinic of our institution, the Asan Medical Center, for a year (from May 1998 to April 1999), 42 patients (17.9%) had small HCC. Age range: 32-75. Males 88%		
Patient characteristics and setting			
Index tests	AFP: serum AFP levels were determined using a commercially-available radioimmunoassay kit (Abbott Laboratories, North Chicago, IL). Cut-off values pre-specified at 20, 100, 200, 400 ng/mL		
Target condition and reference standard(s)	HCC: HCCs were diagnosed clinically in patients with hypervascular mass in the liver and serum AFP levels exceeding 400 ng/mL (n = 8) or through histological means (n = 30). Control group: ultrasonography was performed at 3-6 month intervals for a follow-up period of 12 months or more to determine the presence or absence of intrahepatic masses, which were not found in liver cirrhosis control group.		
Flow and timing	No interval between index test and reference standard		
Comparative			
Notes	No data on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Song 2002 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Song 2011

Study characteristics			
Patient Sampling	This is a prospective cohort study which included consecutive patients with HBV-associated liver cirrhosis from 2003 to 2007. All patients underwent AFP and US as screening modalities. 561 patients were included, out of which 87 patients developed HCC. Exclusion criteria: other malignancies, detection of HCC within 6 months of enrolment. Age range: not reported. Males 71%		
Patient characteristics and setting			
Index tests	AFP: cut-off predefined at 15.5 IU/mL (18,76 ng/mL)		
Target condition and reference standard(s)	HCC was confirmed by CT or MRI.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No potential conflicts of interest reported		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Song 2011 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Song 2014

Study characteristics	
Patient Sampling	This is a case-control study which included consecutive patients divided into five groups. The groups of interest are the following. 1) HCC group, which involved HCC patients proved by pathology after hepatic resection (550 cases) 4) Chronic liver disease group (85 cases), which involved patients with hepatitis or liver cirrhosis Age range: 15-82. Males 87%
Patient characteristics and setting	
Index tests	AFP: serum AFP levels were tested using a commercial ELISA kit in accordance with instructions from the manufacturer (Biocell Biotech, Zhengzhou, China). Youden's index was calculated as an index of sensitivity and specificity. To determine the optimal cut-off values for DCP and AFP to diagnose HCC, receiver operating characteristic (ROC) curves were created using all possible cutoffs for each assay. The optimal cut-off value for AFP was 21 ng/mL.
Target condition and reference standard(s)	HCC: HCC patients proved by pathology after hepatic resection
Flow and timing	No information on interval between index test and reference standard

Song 2014 *(Continued)*

Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Song 2014 *(Continued)*

Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Song 2020a

Study characteristics			
Patient Sampling	Patients with liver cirrhosis or HCC, admitted to Qingdao Sixth People's Hospital from July 2014 to November 2017, were enrolled in this study. Patients with hepatitis C virus infection, alcoholic liver disease, and biliary cirrhosis were excluded. Age range not reported. Males 72%		
Patient characteristics and setting			
Index tests	Serum levels of AFP were detected using a fully automated chemiluminescent enzyme immunoassay. The predefined cut-off value was 10 ng/mL.		
Target condition and reference standard(s)	The diagnosis of HCC was made based on the imaging or histopathology findings. Patients with liver cirrhosis (LC) were diagnosed by liver biopsy and also underwent magnetic resonance imaging or computed tomography screening to exclude the possibility of HCC. Further follow-up for at least 12 months was performed to ensure that no LC patients developed HCC.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare no conflict of interest."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Song 2020a *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		

Song 2020a *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Song 2020b

Study characteristics			
Patient Sampling	Serum samples from 80 patients with hepatitis B virus (HBV) infection and HCC (HCC group), 80 patients with HBV-related liver cirrhosis (LC group), 80 patients with chronic hepatitis B virus infection (HBV group) or 80 healthy controls (HC group). Age range: 32-76. Males 88%		
Patient characteristics and setting			
Index tests	AFP was detected using electrochemiluminescence immunoassay (Elecsys and cobas e analyzers, Roche Diagnostics Mannheim, Germany) No predefinition of a cut-off value.		
Target condition and reference standard(s)	The diagnosis of primary HCC was based on guidelines of the American Association for the Study of Liver Diseases [22]. All patients were diagnosed either by histopathological results after surgical resection or by imaging findings (ultrasound, computed tomography or magnetic resonance) combined with AFP serum levels. Controls: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Conflicts of interest relevant to this article were not reported.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Song 2020b *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Song 2020b *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Soroida 2012

Study characteristics			
Patient Sampling	Consecutive HCC patients with cirrhosis caused by hepatitis B virus (HBV) or hepatitis C virus (HCV), who were treated at the Department of Gastroenterology, of the University of Tokyo Hospital, Tokyo, Japan, between January and April 2010, were enrolled (n = 147). Patients with cirrhosis caused by hepatitis B virus (HBV) or hepatitis C virus (HCV), but who did not have HCC (n = 92), were also enrolled. Age range not reported. Males 63%		
Patient characteristics and setting			
Index tests	Methods for AFP determination were not explained. The cut-off value was 20 ng/mL.		
Target condition and reference standard(s)	Diagnosis of cirrhosis was based on the presence of clinical and laboratory features indicating portal hypertension (the presence of oesophageal varices and/or collateral circulation as observed using an endoscopy, ultrasonography, CT, or MRI). The diagnosis of HCC was made by a dynamic CT or MRI, with hyperattenuation during the arterial phase and washout during the late phase regarded as definite signs of HCC.		
Flow and timing	Blood samples were drawn within one month after the diagnosis and prior to the initiation of treatment in HCC patients. In non-HCC patients, blood samples were obtained within one month since the last surveillance imaging, and the absence of HCC was confirmed at least 6 months after the analysis of blood samples.		
Comparative			
Notes	No conflicts of interest declared		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	

Soroida 2012 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Sterling 2009

Study characteristics	
Patient Sampling	Between June 1 2000, and June 30 2004, there were 372 patients who met eligibility criteria and were enrolled and followed up prospectively by the 7 participating hospitals. Patient inclusion criteria were age 40 to 70 years and a clinical history of cirrhosis diagnosed by histology or a combination of clinical, biochemical, and imaging findings or newly diagnosed HCC. All patients were positive for hepati-

Sterling 2009 *(Continued)*

	tis C virus (HCV) RNA or HCV antibody on a commercial assay. Patients were excluded if they had cancers other than HCC, recurrent HCC, HCC larger than 5 cm on imaging, were pregnant, or currently were undergoing interferon therapy. Age range: 45-64. Males 74%
Patient characteristics and setting	
Index tests	AFP: AFP, AFP-L 3%, and des-gamma-carboxy prothrombin (DCP) levels in serum were measured using the LiBASys automated immunologic analyser (Wako Pure Chemical Industries, Ltd., Osaka, Japan). A generally accepted cut-off value of 20 ng/mL for AFP was used for analysis.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made based either on histology or by the European Association for the Study of the Liver criteria as follows: presence of cirrhosis and a newly diagnosed focal lesion 2 cm or larger with arterial enhancement on 2 imaging studies (including ultrasound, contrast-enhanced computed tomography, contrast magnetic resonance imaging, or angiography) or a focal lesion 2 cm or larger with arterial enhancement on one imaging study associated with a total AFP level greater than 400 ng/mL. Patients without HCC at study entry were followed up every 3 to 6 months for up to 24 months for development of HCC. At each study visit, total AFP, AFP-L3%, and DCP were obtained (Wako Diagnostics, Richmond, VA). Liver imaging was performed at baseline and every 6 to 12 months as per standard protocol at each centre.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors disclose no conflicts of interest."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		

Sterling 2009 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sterling 2012

Study characteristics	
Patient Sampling	Patients in the HALT-C Trial were tested every 3 months for 42 months. Screening ultrasound was performed every 12 months. Age range and % males not reported
Patient characteristics and setting	
Index tests	The absolute cut-off values were: AFP = 20, = 50, or = 200 ng/mL.
Target condition and reference standard(s)	Definite HCC was defined by histological confirmation or by the appearance of a new mass lesion on imaging with AFP levels increasing to ≥1,000 ng/mL.

Sterling 2012 *(Continued)*

	All patients were required to have an ultrasound, computed tomography (CT) or magnetic resonance imaging (MRI) with no evidence of hepatic mass lesions suspicious for HCC and a serum AFP < 200 ng/mL prior to enrolment (protocol exceptions were allowed for three patients who had AFP values of 206, 212, and 315 ng/mL, respectively, and negative imaging).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Financial relationships of the authors with Hoffmann-La Roche, Inc. (now Genentech), are as follows: R.K. Sterling is a consultant and receives research support; T.R. Morgan receives research support; J.C. Hoefs is on the Speaker's Bureau; A.M. Di Bisceglie is a consultant and receives research support; and A.S. Lok is a consultant and receives research support. Financial relationships of the authors with Wako Diagnostics (a division of Wako Chemicals USA, Inc.) are as follows: R.K. Sterling is a consultant; and T.R. Morgan receives research support. Authors with no financial relationships related to this project are: E.C. Wright, L.B. Seeff, and J.L. Dienstag."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Sterling 2012 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sultanik 2017

Study characteristics	
Patient Sampling	From the 4434 patients with at least one hospital record, we selected all patients with virus-related cirrhosis. For patients with a HCC, we selected those with hospital serum samples available at time of HCC diagnosis. Patients with cirrhosis and without HCC were selected for the control group (two controls: one HCC). Patients who were treated with vitamin K antagonists were excluded. Overall, 162 patients with virus-related cirrhosis were retrospectively recruited from 2011 to 2015. Overall, 162 patients with cirrhosis were selected: 46 (28%) patients had HCC and 116 (72%) patients were control patients. Age range: 47-64. Males 62%
Patient characteristics and setting	
Index tests	"AFP: we used serum samples designated for AFP determination. After collection of blood, the tubes were centrifuged at +18°C for 15 minutes at 3000 g, aliquoted and kept frozen at −30°C until analysis. A 20 ng/mL threshold was used for AFP."
Target condition and reference standard(s)	HCC: diagnosis of HCC was ascertained following the recommended guidelines based on imaging criteria (ultrasonography, computed tomography scanning and magnetic resonance imaging) with or without elevated serum AFP concentration. Control patients: all control patients had viral-related cirrhosis and were enrolled during the same period as HCC patients. The absence of HCC was confirmed 1 year after the time of tumour biomarker measurement.

Sultanik 2017 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare no potential conflicts of interest for this study."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			

Sultanik 2017 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sun 2010

Study characteristics			
Patient Sampling	This case-control study included Hong Kong Chinese patients with chronic HBV infection. The cohorts of interest were HCC patients (88) and non-neoplastic control patients (64). Age range: 30-67. Males 69%		
Patient characteristics and setting			
Index tests	AFP: pre-defined cut-off values were: 20, 100, 400 ng/mL.		
Target condition and reference standard(s)	HCC: CT, MR, histology		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		

Sun 2010 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sun 2020

Study characteristics	
Patient Sampling	The study included 146 HCV-infected patients; 40 patients with early-stage HCC and 106 non-malignant HCV-associated chronic liver disease. Age range not reported. Males 63%
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. No definition of a cut-off value
Target condition and reference standard(s)	All HCC patients were on top of HCV cirrhosis and were confirmed by histological examination. Diagnosis of HCV-related chronic liver disease was based on standard clinical, biochemical, serological,

Sun 2020 *(Continued)*

	and ultrasonographic criteria, as well as the histopathological data obtained at liver biopsy.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Conflicts of interest: the authors declare that they have no competing interests."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			

Sun 2020 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Sutherland 2017

Study characteristics	
Patient Sampling	Recruitment criteria included patients aged over 18 years who were referred by the gastroenterology department with chronic liver disease for hepatocellular carcinoma screening liver ultrasound. Exclusion criteria included the presence of a known mass as indicated on the ultrasound request form, non-English speaking (due to inability to gain informed consent), or contraindications to MRI such as pacemaker or contraindicated metallic implant. Age range: 27-80. Males 72%.
Patient characteristics and setting	

Sutherland 2017 *(Continued)*

Index tests	US: the US studies were reviewed by a single abdominal radiologist. Ultrasound lesions were considered suspicious if they were solid and were not clearly focal fat infiltration or focal fat sparing.
Target condition and reference standard(s)	HCC: gold standard for the diagnosis of HCC was by the American Association for the Study of Liver Diseases (AASLD) practice guidelines of arterial phase hyperenhancement followed by washout on either CT or MRI, or by histology (biopsy or resection).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No authors have conflicts of interest to declare.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		

Sutherland 2017 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tahon 2019

Study characteristics	
Patient Sampling	Group 1: 40 cirrhosis patients with primary HCC Group 2: 30 cirrhosis patients without HCC, proved by history, clinical examination, laboratory, and US findings Group 3: 15 healthy control individuals Inclusion criteria: a confirmed clinical picture of cirrhosis, with positive US and routine laboratory tests for cirrhosis, age > 50 years Age range not reported. Males 80%
Patient characteristics and setting	
Index tests	Serum AFP is quantified using chemiluminescence immunonoassay kit manufactured by SIEMENS Health Care Diagnostic Products, LTD No predefinition of a cut-off value
Target condition and reference standard(s)	HCC: a positive US and triphasic CT for malignant focal lesion. Cirrhosis: confirmed clinical picture of cirrhosis, with positive US and routine laboratory tests for cirrhosis
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare that they have no conflicts of interest."
Methodological quality	

Tahon 2019 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			

Tahon 2019 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Takaya 2019

Study characteristics			
Patient Sampling	A retrospective review of medical records was performed including 61 consecutive patients aged ≥ 20 years with cirrhosis, of whom 41 (67.2%) developed HCC and visited the Nara Medical University, Kashihara, Nara, Japan between April and November 2016. Age range: 67-79. Males 69%		
Patient characteristics and setting			
Index tests	The serum AFP level was determined by enzyme-linked immunosorbent assay using a commercially available kit. The predefined cut-off value: 10 ng/mL		
Target condition and reference standard(s)	Diagnosed using dynamic contrast-enhanced CT (DCE-CT), DCE-MRI, or DCE ultrasound (DCE-US)		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare that they have no conflicts of interest."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Takaya 2019 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Takaya 2019 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Takikawa 1992

Study characteristics	
Patient Sampling	628 patients who were admitted to Iwate Medical University Hospital or its affiliated hospitals served as the study population. It included 116 patients with HCC (104 with and 12 without liver cirrhosis), 9 with cholangiocellular carcinoma, 18 with metastatic liver cancer, 29 with acute hepatitis, 128 with chronic hepatitis, 253 with liver cirrhosis without HCC, 6 with primary biliary cirrhosis, 2 with focal nodular hyperplasia of the liver, 6 with hepatic haemangioma, 1 with liver abscess, 20 with fatty liver, 22 with extrahepatic malignancies, 13 with disseminated intravascular coagulation syndrome, and 5 asymptomatic carriers of hepatitis B virus. Age range and % males not reported
Patient characteristics and setting	
Index tests	AFP: serum levels of AFP were measured by a latex immuno-agglutination assay kit (LA-AFP'Eiken', Eiken chemical Co., Tokyo). Cut-off values were prespecified at 20, 100, 200, and 400 ng/mL. Alpha-foetoprotein had the highest validity, at the cut-off value of 100 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made histologically in 53 patients, and in others by typical findings of imaging methods including ultrasonography, computerized tomography, and angiography. Control group: patients with cirrhosis were followed for at least 6 months from the study in order to exclude coexistent HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Takikawa 1992 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Takikawa 1992 *(Continued)*

Could the patient flow have introduced bias?	High risk

Talkahn 2018

Study characteristics			
Patient Sampling	This case-control study included 90 patients with liver cirrhosis divided into three groups. Group I: 40 patients with HCC and cirrhosis Group II: 30 patients with liver cirrhosis and without HCC Group III: 20 healthy people. Age range and % of males not reported		
Patient characteristics and setting			
Index tests	AFP: cut-off value of 220 ng/mL		
Target condition and reference standard(s)	HCC: US and CT were performed.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Authors state: "nothing to disclose"		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Talkahn 2018 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tan 2012

Study characteristics	
Patient Sampling	In this case-control study, serum specimens were collected from 262 patients with HCC, 76 patients with cirrhosis, and 74 patients with hepatitis B. Age range: 22-79. Males 78%
Patient characteristics and setting	
Index tests	AFP: cut-off value pre-defined at 20 ng/mL
Target condition and reference standard(s)	HCC: HCC was histopathologically diagnosed after the tumour excision.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Tan 2012 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		

Tan 2012 *(Continued)*

Could the patient flow have introduced bias?	Unclear risk

Tan 2014

Study characteristics			
Patient Sampling	A multi-stage, case-control study was designed to identify a serum miRNA profile as a surrogate marker for HCC. A total of 261 HCC patients, 233 patients with cirrhosis and 173 healthy controls were enrolled in our study. Validation set (cohort of interest) included 103 HCC patients, 78 cirrhosis patients and 60 healthy controls serum samples (from The Third Hospital of Zhenjiang Affiliated Jiangsu University). Age range: 32-63. Males 67%		
Patient characteristics and setting			
Index tests	AFP: cut-off values not predefined or mentioned		
Target condition and reference standard(s)	HCC: the diagnosis of HCC and cirrhosis was histopathologically confirmed.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"Competing Interests: the authors have declared that no competing interests exist."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		

Tan 2014 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Tanaka 1986

Study characteristics	
Patient Sampling	5339 patients who resided in Osaka prefecture were selected as the subjects of this study. Age range and % of males not reported
Patient characteristics and setting	
Index tests	US was routinely conducted by several well-trained physicians specializing in digestive diseases. The testing took about 15 minutes for each patient.
Target condition and reference standard(s)	HCC: histological, 23.9% by angiographic, and 44.2% by clinical diagnosis
Flow and timing	No information on interval between index test and reference standard

Tanaka 1986 *(Continued)*

Comparative	
Notes	No information on funding or conflicts of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Tanaka 1986 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Tang 2017a

Study characteristics	
Patient Sampling	A total of 366 patients with CHB were retrospectively enrolled at West China Hospital of Sichuan University from November 2015 to August 2016. All patients enrolled in this study were hepatitis B surface antigen positive for at least 6 months. The participants were divided into three groups: HCC; liver cirrhosis (LC) without HCC; and noncirrhotic chronic hepatitis B (CHB). A total of 366 patients were included in this study and divided into three groups: HCC group (n = 176); LC group (n = 98); and CHB group (n = 92). Exclusion criteria: participants who were: heavy alcoholics (more than 80 g of ethanol daily); suffered from cholestatic autoimmune diseases; taking vitamin K or warfarin before PIVKA-II measurement; had evidence of other malignancies; were positive for other virus markers such as HCV, human acquired immunodeficiency virus, cytomegalovirus and Epstein–Barr virus infection. Age range: 33-64. Males 79%
Patient characteristics and setting	
Index tests	AFP: AFP level was measured using electrochemiluminescence immunoassay kit (ECLIA) on E170 analyzer (Roche, Tokyo, Japan). To determine the cut-off values that would best distinguish HCC from non-HCC, ROC analysis was performed for PIVKA-II and AFP, respectively. The optimal cut-off values for PIVKA-II and AFP were 40.5 mAU/mL and 12.3 ng/mL, respectively.
Target condition and reference standard(s)	HCC: HCC was diagnosed on the basis of either histological confirmation or two concordant imaging studies with typical findings of HCC, including abdominal contrast-enhanced ultrasonography (CEUS), dynamic contrast-enhanced CT, or MRI. Control group: the diagnosis of LC was based on the histopathology of a liver biopsy and/or ultrasonic/CT imaging features and was supplemented by clinically-related portal hypertension (e.g. oesophageal and/or gastric varices, ascites, splenomegaly with a platelet count of < 100,000 mm3). CEUS, contrast-enhanced CT or MRI was used to exclude HCC when there was a nodule in the liver.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Conflicts of interest: this work was supported by Science and Technology Support Program of Sichuan Province, China (No. 2015SZ0049). The authors have no other relevant affiliations or financial involvement with any organisation or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Tang 2017a *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Tang 2017a *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tanglijvanich 2010

Study characteristics			
Patient Sampling	Six groups were studied which included 40 healthy individuals, 50 patients with chronic hepatitis (CH), 50 patients with liver cirrhosis (LC), 100 patients with HCC, 50 patients with intrahepatic cholangiocarcinoma (ICC) and 50 patients with metastatic carcinoma (MCA). Age range and % of males not reported		
Patient characteristics and setting			
Index tests			
Target condition and reference standard(s)			
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No conflicts of interest disclosure		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Tanglijvanich 2010 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Tayob 2016a

Study characteristics			

Tayob 2016a *(Continued)*

Patient Sampling	"De-identified data from the Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) Trial, which enrolled 1050 patients with hepatitis C and advanced fibrosis or cirrhosis prospectively followed every 3-6 months, were analysed. During a median follow-up of 80 months, 88 patients (48/427 with cirrhosis and 40/621 with advanced fibrosis) were diagnosed with HCC. The HALT-C Trial enrolled patients with chronic hepatitis C in a randomised controlled trial in which the patients had to have at least stage 3 fibrosis (bridging fibrosis or cirrhosis) by the Ishak scoring system (range 0-6) and a history of failure to respond to previous interferon-based therapy. All patients had radiological imaging to exclude HCC prior to enrolment. All participants were had HCV infection and treatable with interferon." Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP: the corresponding AF thresholds for the ST method were 22.3, 29.0 and 42.6 ng/mL, respectively in the cirrhosis subgroup, and 14.0, 16.6 and 22.9 ng/mL, respectively in the advanced fibrosis subgroup.
Target condition and reference standard(s)	HCC: diagnosis of HCC was based on histology and in the absence of histology, by imaging with or without AFP. All patients had radiological imaging to exclude HCC prior to enrolment. Patients with elevated AFP or new lesions on ultrasound were further evaluated with CT or MRI.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Conflicts of interest declared: Anna Lok has no conflicts to declare other than that she was one of the HALT-C investigators. The other authors have no conflicts of interest to declare."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Tayob 2016a *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tayob 2016b

Study characteristics	
Patient Sampling	"De-identified data from the Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) Trial, which enrolled 1050 patients with hepatitis C and advanced fibrosis or cirrhosis prospectively followed every 3-6 months, were analysed. During a median follow-up of 80 months, 88 patients (48/427 with cirrhosis and 40/621 with advanced fibrosis) were diagnosed with HCC. The HALT-C Trial enrolled patients with chronic hepatitis C in a randomised controlled trial in which the patients had to have at least stage 3 fibrosis (bridging fibrosis or cirrhosis) by the Ishak scoring system (range 0-6) and a history of failure to respond to previous interferon-based ther-

Tayob 2016b *(Continued)*

	apy. All patients had radiological imaging to exclude HCC prior to enrolment." All participants were had HCV infection and treatable with interferon. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP: the corresponding AF thresholds for the ST method were 22.3, 29.0 and 42.6 ng/mL, respectively in the cirrhosis subgroup, and 14.0, 16.6, and 22.9 ng/mL, respectively in the advanced fibrosis subgroup.
Target condition and reference standard(s)	HCC: diagnosis of HCC was based on histology and in the absence of histology, by imaging with or without AFP. All patients had radiological imaging to exclude HCC prior to enrolment. Patients with elevated AFP or new lesions on ultrasound were further evaluated with CT or MRI.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Conflicts of interest declared: Anna Lok has no conflicts to declare other than that she was one of the HALT-C investigators. The other authors have no conflicts of interest to declare."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Tayob 2016b *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tayob 2019

Study characteristics	
Patient Sampling	"The study cohort included patients with cirrhosis of any aetiology identified in the VA corporate data warehouse (CDW), a national repository of VA clinical and administrative data from a network of 153 VA hospital facilities. Patients were eligible if they had a diagnosis of cirrhosis, evidenced by the presence of International Classification of Diseases, 9th Revision (ICD-9) codes 571.2 or 571.5, between October 1, 1996 and May 30, 2015. In addition, the analysis cohort was restricted to include (1) HCC cases with at least 1 pre-diagnosis AFP test and (2) controls with at least 1 AFP test and a minimum of 12 months of follow-up to confirm no HCC. For both cases and controls, we only included AFP tests with ALT and platelet laboratory tests performed within 6 months before the AFP test." Age range not reported. Males 97%
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. Predefined cut-off value: 400 ng/mL

Tayob 2019 *(Continued)*

Target condition and reference standard(s)	"We determined HCC diagnosis in the cirrhosis cohort by using a sequential procedure. First, we identified patients with probable HCC via ICD-9 codes, which were defined as at least 1 inpatient or 2 outpatient 155.0 codes (but without 155.1). Next, we verified these HCC diagnoses by incorporating information from the VA Central Cancer Registry (VACCR) and the VA CDW oncology raw data files."
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors disclose no conflicts of interest."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			

Tayob 2019 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Teefey 2003

Study characteristics	
Patient Sampling	"Between August 1996 and December 1998, we examined 37 patients. In an effort to recruit patients who were at increased risk for malignancy, only patients with an elevated serum-fetoprotein level (30 ng/mL) or with primary sclerosing cholangitis were eligible.Ten of the patients either died prior to liver transplantation (without autopsy or biopsy being performed) or their names were removed from the transplant list. Two patients whose names had been on the transplant list for more than 2 years were not included in the study because of an inability to obtain follow-up images. The remaining 25 patients form the study population."

Teefey 2003 *(Continued)*

	Age range: 19-63. % of males not reported
Patient characteristics and setting	
Index tests	Results of each imaging test (CT, MR imaging, US, and PET) were interpreted independently by two radiologists experienced (10 years of experience for all radiologists). For each patient, the reviewer was asked to indicate his or her degree of confidence that a malignancy was present on the basis of a six-point confidence scale: 1, definitely present; 2, probably present; 3, possibly present; 4, possibly not present; 5, probably not present; and 6, definitely absent.
Target condition and reference standard(s)	Gross and histologic analyses of all explanted livers were performed by an experienced hepatobiliary pathologist. If a lesion identified at an imaging test could not be demonstrated in the explant, representative histologic sections were obtained from the region of the liver that best corresponded.
Flow and timing	The interval between the last imaging study and the liver transplantation in the 21 patients who had a liver transplant ranged from 1 to 15 months (mean, 5.3 months).
Comparative	
Notes	No information on conflicts of interest or funding

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Teefey 2003 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Teng 2016

Study characteristics	
Patient Sampling	"A total of 205 subjects were retrospectively collected in this study, including 111 patients with HCC, 66 patients with CHB, and 28 healthy controls (HCs), from March 2013 to June 2015 at the Department of Hepatology, Qilu Hospital of Shandong University. Exclusion criteria included other tumours, co-infection with hepatitis C virus or human immunodeficiency virus, autoimmune liver diseases, non-alcoholic fatty liver diseases, alcoholic liver diseases and other causes of chronic liver diseases." Age range: 42-64. Males 80%
Patient characteristics and setting	
Index tests	AFP: AFP cut-off value of 20 ng/mL
Target condition and reference standard(s)	HCC patients were diagnosed according to the 2010 update of the American Association for the Study of Liver Diseases (AASLD) Practice Guidelines for Management of HCC

Teng 2016 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Teng 2016 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tian 2017

Study characteristics			
Patient Sampling	"The study included 120 patients with HCC associated with hepatitis B, 146 patients with chronic hepatitis B (CHB) and 27 healthy controls (HCs). Exclusion criteria included coinfection with hepatitis C virus (HCV) or human immunodeficiency virus (HIV), autoimmune liver disease, alcoholic liver diseases, non-alcoholic fatty liver diseases (NAFLD) and other causes of chronic liver diseases." Age range: 40-64. Males 74%		
Patient characteristics and setting			
Index tests	AFP: serum AFP was assayed by an electro-chemiluminescence immunoassay using an automatic analyser (COBAS e 601, Roche Diagnostics, Mannhein, Germany). Serum AFP level > 20 ng/mL was regarded as abnormal.		
Target condition and reference standard(s)	HCC: HCC was diagnosed following the 2010 update of the American Association for the Study of Liver Diseases (AASLD) Practice Guidelines for Management of Hepatocellular Carcinoma.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"Authors declare that they have no competing interest."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Tian 2017 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tong 2001

Study characteristics	
Patient Sampling	"From 1991 to 1998, 602 patients who were referred to the Liver Center at the Huntington Memorial Hospital in Pasadena, California, USA were enrolled in the surveillance for HCC. All patients were positive for either hepatitis C virus antibodies (anti-HCV; Ortho HCV EIA; Ortho Di-

Tong 2001 *(Continued)*

	agnostics, Raritan, NJ, USA), or for hepatitis B virus surface antibodies. To be included, all patients had to have at least 1 year of follow-up in our clinic (Liver Center)." Age range not reported. Males 59%
Patient characteristics and setting	
Index tests	"AFP levels were measured at two commercial laboratories. One laboratory at Huntington Memorial Hospital in Pasadena, California used the Axsym EIA (Abbott Laboratories; the upper limit of normal was 10.9 ng/mL. The second laboratory, Nichols Laboratories in Los Angeles, California, used an in-house chemoluminescent method from 1990 to 1995 (the upper limit of normal was 18 ng/mL). For serum samples sent to Nichols Laboratories after 1995, the ACS-180 chemoluminescent test (Chiron Diagnostics, Emeryville, CA, USA) was used (the upper limit of normal was 8.1 ng/mL.For final data presentation, ratios were reconverted to AFP values by multiplication of the AFP ratio by the upper limit of normal of the currently available test (8.1 ng/mL)."
Target condition and reference standard(s)	US, CT, and histology
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on funding or conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Tong 2001 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Toraih 2018

Study characteristics	
Patient Sampling	70 individuals were enrolled in the study (20 controls, 20 patients with liver cirrhosis (LC) caused by HCV infection, and 30 patients with hepatocellular carcinoma on top of HCV). Patients with non-HCV induced HCC, other autoimmune or metabolic liver diseases were excluded. Age range: 23-80. Males 62%
Patient characteristics and setting	
Index tests	AFP: serum AFP concentration was measured by the chemiluminescent immunometric assay on Siemens IMMULITE® 2000 (Siemens Healthcare Diagnostics, USA). ROC analysis revealed the diagnostic performance of AFP that differentiates cancer patients from normal and cirrhotic individuals at the cut-off values of 131 ng/mL and 205 ng/L.

Toraih 2018 *(Continued)*

Target condition and reference standard(s)	Liver cancer had typical imaging findings and elevated serum alpha foetoprotein (AFP).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors have declared that no competing interests exist."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Toraih 2018 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tremolada 1989

Study characteristics			
Patient Sampling	During the whole 1987, 247 patients with cirrhosis were enrolled. Age range: 24-81. Male 69%		
Patient characteristics and setting			
Index tests	Serum AFP measurement by RIA (Abbot); predefined cut-off value 20 ng/mL. US real time Ansaldo, no definition of positivity criteria. AFP +US - one positive (AFP cut-off 20 ng/mL))		
Target condition and reference standard(s)	Histology, US and follow-up		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Tremolada 1989 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Tremolada 1989 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Trevisani 2001

Study characteristics			
Patient Sampling	This case-control study aimed to identify the best cut-off value of serum AFP to discriminate chronic liver disease (CLD) patients with and without HCC. Two hundred and ten cases fulfilled these criteria. Among them, we were able to match 170 cases (135 males and 35 females) with 170 controls with CLD seen during the same period according to the following criteria: age (within 6 years), sex, underlying CLD (cirrhosis/chronic hepatitis), HBsAg and HCV status. Patients with liver disease due to genetic and autoimmune disorders, primary biliary cirrhosis and sclerosing cholangitis were excluded. Age range: 50-70. Male 79%		
Patient characteristics and setting			
Index tests	AFP: AFP was measured by conventional assays (radioimmunoassay, Eiken Chemical Co., Tokyo, Japan; LA-AFP test, Poli, Milan, Italy; immunoenzymatic assay, Abbott Laboratories, Rome, Italy). The analysis was performed using these cut-off values: the best discriminating value provided by the receiver-operating characteristic (ROC) curve, the value of 20 ng/mL, and 100, 200 ng/mL and 400 ng/mL.		
Target condition and reference standard(s)	HCC: the diagnosis of HCC was based on histological or cytological findings in 128 patients, while it was confirmed by clinical and imaging data or necropsy in the remainder. Control group: In control patients, the presence of HCC was ruled out by ultrasonography and also by excluding patients who developed HCC during the following 6 months.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Trevisani 2001 *(Continued)*

Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tsai 1995

Study characteristics			

Tsai 1995 *(Continued)*

Patient Sampling	The study population comprised 101 consecutive cirrhotic HCC patients and 101 sex-matched and age-matched (± 5 years) patients with cirrhosis alone. Age range 26-87. Males 91%.
Patient characteristics and setting	
Index tests	AFP: serum was tested for HBsAg and AFP (Ausria II and a-Feto Riabead, Abbott Laboratories, Chicago, IL, USA). Receiver operating characteristic (ROC) curves were constructed by calculating the sensitivities and specificities of AFP or CIC assays at several cut-off points: 3, 4, 5, 8, 22, 40, 83, 120, 400 ng/mL.
Target condition and reference standard(s)	HCC: HCC was diagnosed by liver biopsy or aspiration cytology. LC was clinicopathologically proven. There was no space-occupying lesion in LC patients and healthy controls as evidenced by normal abdominal sonography.
Flow and timing	No information on interval between index test and reference standard.
Comparative	
Notes	No information on conflicts of interest.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Tsai 1995 *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tsai 1997

Study characteristics	
Patient Sampling	The study population comprised 94 non-alcoholic consecutive cirrhotic HCC patients and 94 sex-matched and age-matched (± 5 years) patients with cirrhosis alone.
Patient characteristics and setting	
Index tests	AFP: HBsAg, anti-HCV and AFP were tested with Ausria-I1, second generation HCV enzyme immunoassay (EIA) and a-feto RIABEAD (Abbott Laboratories, Chicago, IL, USA) ROC curves were constructed by calculating the sensitivities and specificities of AFP or TGF-P1 assays at several cut-off points (3, 4, 7, 12, 16, 28, 100, 400 ng/mL). The cut-off value with the highest accuracy was selected as diagnostic cut-off point.
Target condition and reference standard(s)	HCC was diagnosed by liver biopsy or aspiration cytology. There was no space-occupying lesion in patients with cirrhosis alone and healthy controls as evidenced by normal abdominal sonography.

Tsai 1997 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Tsai 1997 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Tsai 2017

Study characteristics			
Patient Sampling	We retrospectively reviewed the medical records of pathology-proven patients with cirrhotic HCC who had pre-biopsied AFP level and those with cirrhosis alone. A total of 986 patients with HCC or cirrhosis were enrolled. Age range: 29-72. Males 40%		
Patient characteristics and setting			
Index tests	AFP: authors used pre-specified cut-off values at 20, 100, 200, and 400 ng/mL.		
Target condition and reference standard(s)	"We retrospectively reviewed the medical records of pathology-proven patients with cirrhotic HCC who had pre-biopsied AFP level and those with cirrhosis alone."		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Tsai 2017 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Tsuda 2004

Study characteristics	
Patient Sampling	A total of 56 consecutive patients with HCC accompanied with LC (40 men and 16 women; mean age, 69 years old), who had been followed-up at Liver Unit in the First Department of Internal Medicine of Osaka Medical College Hospital during the period from December 1999 to November 2000 were enrolled in this study. Thirty-two patients with liver cirrhosis without HCC d (23 men and 9 women; mean age, 65 years) who had been followed at our hospital during the same period were also studied as a control group. None of the patients had bacterial or other viral infection, chronic renal damage, insulin-dependent diabetes mellitus (IDDM), other malignant disease, hepatic encephalopathy, and obvious flare-up of hepatitis. The pa-

Tsuda 2004 *(Continued)*

	tients undergoing Interferon administration or immunosuppressive therapy were also excluded from this study. Age range: 57-76. Males 72%
Patient characteristics and setting	
Index tests	AFP: serum AFP level were measured by using available commercial radioimmunoassay (a-FETO RIABEAD, Dinabot, Tokyo). Cut-off values prespecified at 20 an 100 ng/mL.
Target condition and reference standard(s)	The diagnosis of LC was performed by biochemical data, histological findings of liver biopsy, typical findings by US and abdominal computed tomographic Scan (CT) such as nodular surface, dull edge, course parenchyma and splenomegaly. The existence of liver tumour was detected by using US and/or CT, and the diagnosis of HCC was made by the typical findings of tumour staining in hepatic angiography and/or by the histology of needle biopsy under ultrasonography from liver tumour.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Tsuda 2004 *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ungtrakul 2016

Study characteristics	
Patient Sampling	In the present study, we undertook a screening and surveillance program involving treatment-naïve chronic hepatitis B (CHB) patients using abdominal US and serum AFP assay. We enrolled male and female Thai patients, aged 20-65 years, who were serologically positive for hepatitis B surface antigen (s-Ag). The exclusion criteria included: decompensated cirrhosis (Child-Pugh class C or Model for End-stage Liver Disease score > 15); a history of any cancer in the last 5 years; previous antiviral treatment for CHB; concurrent infection with hepatitis C virus infection or human immunodeficiency virus infection; a Karnofsky Performance Status score < 60%; or any medical condition preventing eligibility to complete the protocol (e.g., poor renal function, a serum creatinine level > 1.5 mg/dL, or creatinine clearance < 50 mL/minute. Age range: 20-65. Males 47%
Patient characteristics and setting	

Ungtrakul 2016 *(Continued)*

Index tests	AFP: AFP assays were performed with COBAS 6000/e601 Roche Diagnostics, Mannheim, Germany. Cut-off value prespecified at 20 ng/mL. US: US examinations were performed by experienced radiologists at the initial screening and every 6 months thereafter. Diagnostic criteria for further diagnostic evaluation: focal solid liver nodule.
Target condition and reference standard(s)	HCC was diagnosed using the American Association for the Study of Liver Diseases (AASLD) practice guidelines. If the serum AFP was ≥ 20 mg/L or a focal solid liver nodule was detected on US, further diagnostic studies were performed including computerized tomography, magnetic resonance imaging, or biopsy of the liver lesion. AUS examinations were performed by experienced radiologists at the initial screening and every 6 months.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors of this manuscript declare that they have no conflicts of interest to disclose."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Ungtrakul 2016 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Unic 2013

Study characteristics

Unic 2013 *(Continued)*

Patient Sampling	This case-control study included 32 patients with HCC on the basis of alcoholic liver disease and 28 patients with alcohol-related liver cirrhosis as a control group. Age range and % of males not reported
Patient characteristics and setting	
Index tests	AFP: AFP concentrations were determined by Cobas e411 analyser (Hitachi High Technologies Corporation, Tokyo, Japan) with no predefined cut-off value.
Target condition and reference standard(s)	No information on reference standard
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			

Unic 2013 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Van Thiel 2004

Study characteristics	
Patient Sampling	This study included people with end-stage liver disease due to any cause who were evaluated and found to be free of an identifiable HCC and who met United Network for Organ Sharing (UNOS) criteria for listing for liver transplantation. Specifically, from October 1998 through July 2003, a total of 300 individuals were evaluated and presented to the liver transplant review board at Loyola University Medical Center. Of these, 282 were listed for transplantation. Fifteen of these cases were identified as having an HCC at the time of listing and five were listed because of fulminant hepatic failure. These cases were eliminated from the subsequent analysis leaving a total of 262 listed liver transplant candidates. Of these, 105 (41%) were transplanted with four individuals receiving two and one individual receiving three transplants. These later cases receiving multiple transplants were eliminated leaving 100 cases for analysis. Age range: 48-61. Males 68%
Patient characteristics and setting	
Index tests	US: the US criteria used to identify a new hepatic lesion consisted of the finding of either a hypoechoic lesion 1 cm in diameter, or a target lesion consisting of a hypoechoic lesion 1.5 cm in diameter with a central hyperechoic area or a mass adjacent to a thrombosed intrahepatic portal vein radicle.
Target condition and reference standard(s)	All included patients: "the liver transplant evaluation procedures consist of a complete virological, serological, and biochemical evaluation for the recognised causes of end-stage liver disease. In addition, imaging procedures consisting of a CT of the head, triphase CT of the abdomen, and an US examination of the liver and its vessels as well as the biliary tree are obtained. Finally, AFP, CT, and US studies are obtained to screen for the presence of hepatic cancer. Each listed candidate underwent continuous surveillance for the presence of hepatic cancer utilising a quarterly determination of the serum FP level and an abdominal US examination and a semi-annual triphasic CT scan of the abdomen.

Van Thiel 2004 *(Continued)*

	The explant liver and hepatic vessels were examined grossly for the presence of any tumour. In addition, any lesion recognised by the pathologist but not recognised by the CT or US surveillance studies were examined histologically."
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Van Thiel 2004 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Villacastin Ruiz 2016

Study characteristics	
Patient Sampling	"From November 2001 to December 2011, 323 orthotopic LTs were performed on 313 patients at our centre. Our study is based on the retrospective analysis of data from 273 patients (213 men and 60 women), of an average age of 55 years (31–79), who underwent scheduled transplants because of cirrhosis. Exclusion criteria were as follows: having undergone urgent non elective transplants; having undergone retransplantation; and absence of cirrhosis. Ultrasonography was carried out in 270 patients." Age range: 31-79. Males 78%
Patient characteristics and setting	
Index tests	Abdominal ultrasonography was performed using a Toshiba SSA-340 (Toshiba Corporation, Tokyo, Japan) (November 2001 to May 2009) and a Toshiba Aplio XG (Toshiba Corporation) (June 2009 to December 2011) equipped with a 3.5 MHz curved array transducer. We retrospectively revised all of the pretransplant reports carried out by experienced radiologists for each imaging study. A negative result (no HCC) was recorded when no lesion was detected or when the lesions were benign. Studies registered as positive were those in which a lesion suggesting HCC was observed.
Target condition and reference standard(s)	"The pathological analysis of the explant livers provided our reference standard. The reports were reviewed retrospectively, and the presence, size and location of HCC nodules were recorded.

Villacastin Ruiz 2016 *(Continued)*

	Correlation of nodules between the image and pathological results was based primarily on location and secondarily on size."
Flow and timing	The average waiting time between imaging tests and transplant was 105 days.
Comparative	
Notes	No conflicts of interest declared

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Villacastin Ruiz 2016 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Volk 2007

Study characteristics	
Patient Sampling	All patients were consecutively enrolled from the liver clinics at the University of Michigan Medical Center between February 2003 and December 2004. Two groups of consecutive patients were enrolled: patients with a diagnosis of HCC, and control individuals with cirrhosis without HCC. A total of 84 patients with HCC and 169 patients with cirrhosis and no HCC were enrolled. Age range: 45-71. Males 67%
Patient characteristics and setting	
Index tests	AFP: total AFP was tested using commercially available immunometric assays with enhanced chemiluminescence at the University of Michigan Hospital Clinical Diagnostic Laboratory and at Wako Diagnostics (Richmond, Virginia). Receiver-operating characteristic (ROC) curves were constructed to determine the optimal cutoff for each marker in differentiating between HCC and cirrhosis without HCC. The optimal cut-off that maximized the sensitivity and specificity for AFP was a total AFP > 23 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was based on the European Association for the Study of the Liver (EASL) criteria. Control group: to ensure that cirrhosis controls did not have HCC, an ultrasound showing no mass was required if the total AFP was < 20 ng/mL, and a triple phase CT or dynamic MRI was required if the AFP was > 20 ng/mL. Additionally, these patients were followed for a median of 14 months (range: 8–37 months) and had at least one follow-up imaging to assure that none had developed HCC. Computed tomography (CT) and magnetic resonance imaging (MRI) studies of patients with HCC were reviewed by one radiologist who was not aware of the serum marker results.
Flow and timing	No information on interval between index test and reference standard

Volk 2007 *(Continued)*

Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Volk 2007 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Vongsuvanh 2016

Study characteristics	
Patient Sampling	This case-control study involved four independent groups comprising a total of 344 participants recruited from a single tertiary liver clinic in Sydney, Australia. The HCC group comprised 86 patients with tumours diagnosed by characteristic radiological appearances on 4-phase CT or MRI according to the European Association for the Study of the liver (EASL) guidelines 2012, or by histology. The HCC cases were age and sex-matched (+/- 10 years) to three additional cohorts comprising patients with cirrhosis, chronic liver disease without cirrhosis, and healthy controls. Age range not reported. Males 87%
Patient characteristics and setting	
Index tests	Serum AFP was measured using a chemiluminsecent microparticle immunoassay (Abbott Diagnostics, Illinois US. Cut-off value 20 ng/mL
Target condition and reference standard(s)	The HCC group comprised 86 patients with tumours diagnosed by characteristic radiological appearances on 4-phase CT or MRI according to the European Association for the Study of the liver (EASL) guidelines 2012, or by histology. Patients in the cirrhosis and chronic liver disease groups were undergoing 6-monthly HCC surveillance with no evidence of HCC at the time blood was collected for the study and for a minimum follow-up of 6 months thereafter.
Flow and timing	No information on interval between index test and reference standard.
Comparative	
Notes	The authors have declared that no competing interests exist

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Vongsuvanh 2016 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2005

Study characteristics

Wang 2005 *(Continued)*

Patient Sampling	A total of 127 patients who were regularly followed up at Ren-Ai Branch, Taipei City Hospital were consecutively enrolled. Among the 127 patients, 32 had chronic hepatitis for at least 6 months before enrolment, 34 had compensated cirrhosis and 61 had HCC. Patients who had a history of alcohol consumption in excess of 80 g/ethanol per day for more than 5 years, serum total bilirubin level of more than 20 mg/L or under vitamin K medication were excluded. Age range: 42-76. Males 74%
Patient characteristics and setting	
Index tests	AFP: AFP was tested by using commercially-available immunometric assay (Architect AFP assay, Abbott Laboratories, North Chicago, IL, USA). The cut-off value of AFP for HCC was set at 20 ng/mL, the most commonly set value.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made on 47 (77%) histologically-confirmed patients. The remaining 14 (23%) patients, who had advanced HCC with tumour size larger than 3 cm or patients with portal vein invasion, were confirmed by various combination of imaging studies, such as ultrasonography, enhanced CT, magnetic resonance imaging and/or angiography. Control group: among all patients with chronic hepatitis and cirrhosis, HCC must be ruled out on the basis of imaging examinations including sonography and/or computed tomography (CT) performed on a regular examination. Also, cirrhotic patients who developed HCC within 6 months after getting serum were excluded.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			

Wang 2005 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2009

Study characteristics	
Patient Sampling	Consecutive patients with HCC and patients with cirrhosis that were age, gender, and race/ethnicity matched to the HCC patients were enrolled from the Liver Clinic from Saint Louis University School of Medicine or the University of Michigan. The study included 113 patients with cirrhosis, 108 patients with stage I or II HCC, and 56 patients with stage III or IV HCC. Age range: 42-71. Males 71%
Patient characteristics and setting	

Wang 2009 *(Continued)*

Index tests	AFP: AFP was tested using commercially-available immunoassays using enhanced chemiluminescence at the University of Michigan Hospital Clinical Diagnostic Laboratory. The upper limit of normal was 8 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made by histopathology, including all T1 lesions, and, if histopathology was not available, by two imaging modalities [ultrasound (US), magnetic resonance imaging (MRI), or computed tomography (CT)] showing a vascular enhancing mass of > 2 cm. Control group: each of the patients with cirrhosis had a normal US and, if serum AFP was elevated, a MRI of the liver within 3 months before enrolment and another one 6 months after enrolment that showed no liver mass. The cirrhotic controls have been followed for a median of 12 months (range, 7-18 months) after enrolment, and no one has developed HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No potential conflicts of interest exist

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Wang 2009 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2013a

Study characteristics	
Patient Sampling	Blood samples were collected under informed consent from 55 HCC patients in the infectious department of our hospital from 2009 to 2010 (Union Hospital, Wuhan, China). For comparison, 40 patients with liver cirrhosis we encountered during the same period were also included. Age range: 39-65. Males 79%
Patient characteristics and setting	
Index tests	AFP: taking both sensitivity and specificity into account, the cut-off point was selected according to maximum number of sensitivity and specificity. AFP showed 85.71% specificity and 40.00% sensitivity at the cut-off value of 20 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was based on typical findings in three-phase dynamic CT or MRI, and the diagnosis was confirmed by histopathology.
Flow and timing	No information on interval between index test and reference standard
Comparative	

Wang 2013a *(Continued)*

Notes	"The authors declare that there is no conflict of interest."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Wang 2013a *(Continued)*

Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2013b

Study characteristics			
Patient Sampling	From 2000 to 2011, serum from 126 patients with completely resected HCC and 115 non-HCC chronic HBV carriers (NC group) were collected from the Department of General Surgery and the Department of Gastroenterology and Hepatology. Age range: 15-86. Males 69%		
Patient characteristics and setting			
Index tests	AFP: the area under the Receiver Operating Characteristic (ROC) Curve was reported to evaluate the ability of the potential serum markers in discriminating HCC patients from the controls. The analysis was performed based on optimal cut-off value of 4 ng/mL and commonly used 20 ng/mL.		
Target condition and reference standard(s)	HCC: the histology of the resected specimens confirmed the diagnosis of HCC. Control group: a follow-up visit six months after serum collection confirmed these HBV carriers' non-HCC status.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors have declared that no competing interests exist."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			

Wang 2013b *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2014a

Study characteristics	
Patient Sampling	Serum samples of 80 patients with HCC and 67patients with liver cirrhosis were analysed by MALDI-TOF-MS for peptide expression. Independent training and test sets were created, with similar representation of age and gender in each set. Diagnostic accuracy analysis was performed in the test set which comprised of 40 patients with HCC and 34 patients with liver cirrhosis alone. Age range: 42-65. Males 73%
Patient characteristics and setting	

Wang 2014a *(Continued)*

Index tests	AFP: cut-off value of 20 ng/mL prespecified		
Target condition and reference standard(s)	Blood biochemistry, AFP assay, computed tomography, and liver biopsy were performed on all patients.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Conflicts of interest: none		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Wang 2014a *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Wang 2014b

Study characteristics			
Patient Sampling	A total of 257 individuals were enrolled from November 2007 to May 2010, including healthy control (n = 61), HBV carrier (n = 32), patients with cirrhosis (n = 80), and patients with HCC (n = 84). Age: 38-86. Males 79%		
Patient characteristics and setting			
Index tests	AFP: AFP levels were examined using commercially available immunoassays at the PUMC Hospital Clinical Diagnostic Laboratories. No information on AFP cut-off level.		
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made either by histopathology or by two different imaging tests (ultrasound, computed tomography [CT], magnetic resonance imaging [MRI] or angiography) showing an arterial enhancing lesion with HBV infection. Liver biopsy was obtained to confirm the diagnosis in some cases.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare that they have no competing interests."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		

Wang 2014b *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2016a

Study characteristics	
Patient Sampling	This was a nested case–control study from the University of Michigan (UM). Patients with cirrhosis were enrolled from UM Liver Clinics between September 2001 and August 2004.

Wang 2016a *(Continued)*

	Age range: 45-71. Males 63%
Patient characteristics and setting	
Index tests	AFP was tested using commercially available immunoassays utilising enhanced chemiluminescence at the UM Hospital Clinical Diagnostic Laboratory with no predefined cut-off value.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was made by histopathology, including all T1 lesions, or by two imaging modalities MRI or CT, showing a vascular enhancing mass > 2 cm with delayed washout. Control: cirrhosis controls were followed for a median of 12 months (range, 7–18 months) after enrolment to confirm absence of HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Conflicts of interest: T. Block reported receiving commercial research grant from Arbutus BioPharma, has ownership interest (including patents) in Glycotest, and was consultant/advisory board member for Glycotest. No potential conflicts of interest were disclosed by the other authors."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Wang 2016a *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2016b

Study characteristics	
Patient Sampling	"This is a randomised clinical trial which included HCV-positive patients with bridging fibrosis or cirrhosis that did not respond to peg-interferon and ribavirin which were randomised to groups that were given maintenance peg-interferon for 3.5 years or no treatment. Patients with detectable HCV RNA at 10 clinical centres had to meet the following criteria for enrolment: failure to have achieved a sustained virologic response (SVR) after previous interferon treatment with or without ribavirin, the presence of advanced hepatic fibrosis on liver biopsy (Ishak fibrosis score ≥ 3), no history of hepatic decompensation or HCC, and the absence of defined exclusion criteria (e.g., liver disease other than hepatitis C, uncontrolled medical or psychiatric conditions, or contraindications to use of interferon or ribavirin. For this study, 151 individuals (49 HCC cases and 102 HCV non-HCC controls) were examined." Age range: 45-57. Males 69.5%
Patient characteristics and setting	
Index tests	AFP measurement with no predefined cut-off value

Wang 2016b *(Continued)*

Target condition and reference standard(s)	"HCC: two definitions of HCC were adopted, one for “definite” HCC and one for “presumed” HCC. Definite HCC was defined by histologic confirmation or a new mass lesion on imaging with AFP levels increasing to >1000 ng/mL. Presumed HCC was defined as a new mass lesion on ultrasound in the absence of histology and AFP was <1000 ng/mL in conjunction with one of the following characteristics: a) 2 liver imaging studies showing a mass lesion with characteristics of HCC (vascular enhancement, wash out), b) progressively enlarging lesion on ultrasound leading to death, or c) 1 additional imaging study showing a mass lesion with characteristics of HCC that either increased in size over time or was accompanied by increasing AFP levels."
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Block T. reported receiving commercial research grant from Arbutus Bio-pharma, had ownership interest (including patents) in Glycotest, and was consultant/advisory board member for Glycotest. No potential conflicts of interest were disclosed by the other authors.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			

Wang 2016b *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2016c

Study characteristics	
Patient Sampling	The cohort consisted of 870 patients (432 HCC cases and 438 non-HCC cirrhosis controls). Cases included consecutive adult patients with HCC seen between February 2005 and August 2007 at seven medical centres in the USA. Patients with HCC were excluded if they were younger than 18 years of age, had prior treatment of their tumour, or history of other solid tumours. Age range: 46-71. Males 74.5%
Patient characteristics and setting	
Index tests	AFP measurement with no predefined cut-off value
Target condition and reference standard(s)	HCC: HCC was defined by histological examination or by the appropriate imaging characteristics as defined by accepted guidelines. To assure that controls did not have HCC, all controls were assessed by AFP and an imaging test (US, CT, or MRI) 6 months after enrolment.

Wang 2016c *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	T. Block reported receiving commercial research grant from Arbutus Bio-pharma, had ownership interest (including patents) in Glycotest, and was consultant/advisory board member for Glycotest. No potential conflicts of interest were disclosed by the other authors.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Wang 2016c *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2016d

Study characteristics			
Patient Sampling	This study included 699 patients (113 HBV-related HCC and 586 HBV-positive controls). Patients included Asian Americans who had HCC induced by chronic HBV infection (excluding all other aetiologies) or HBV-infected patients without HCC (excluding coinfection with HCV). Age range: 31-66. Males 69%		
Patient characteristics and setting			
Index tests	AFP measurement with no predefined cut-off value		
Target condition and reference standard(s)	Liver cirrhosis and HCC were determined through liver biopsy supplemented by imaging examinations, mainly MRI.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	T. Block reported receiving commercial research grant from Arbutus Bio-pharma, had ownership interest (including patents) in Glycotest, and is consultant/advisory board member for Glycotest. No potential conflicts of interest were disclosed by the other authors.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		

Wang 2016d *(Continued)*

Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2016e

Study characteristics

Wang 2016e *(Continued)*

Patient Sampling	This study included data from University of Texas Southwestern and the Parkland Health and Hospital System, consisting of 1,229 patients (425 HCC cases and 804 cirrhosis controls). Age range: 45-70. Males 70%
Patient characteristics and setting	
Index tests	Serum AFP was determined using commercially available immunoassays with no predefined cut-off value.
Target condition and reference standard(s)	HCC: diagnosis based on AASLD criteria Control group: all control patients were required to have 6 months of follow-up to confirm absence of HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Block reported receiving commercial research grant from Arbutus Bio-pharma, had ownership interest (including patents) in Glycotest, and is consultant/advisory board member for Glycotest. No potential conflicts of interest were disclosed by the other authors.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High

Wang 2016e *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2017

Study characteristics	
Patient Sampling	Patients were enrolled from January 2014 to March 2015 in Yantai Yu Huang top Hospital and Infectious Disease Hospital of Yantai City. Patients were allocated into two different categories: HBV-related HCC patients (HCC group - 113 patients) and chronic HBV infected non-HCC patients (CHB group - 161 patients). Age range: 21-75. Males 93%
Patient characteristics and setting	
Index tests	AFP: AFP level was detected with an AFP reagent kit (Abott, IL USA) on an Abott Architect Plus automatic biochemical analyzer (Abott, IL USA) according to the manufacturer's manual. Cut-off values and area under curve (AUC) were calculated. The optimal cut-off value of AFP was 17.56 ng/mL.
Target condition and reference standard(s)	The diagnosis of liver cancer was made in accordance with the standards of diagnosis and treatment of primary liver cancer (2011 Edition) issued by the Ministry of Public Health of the People's Republic of China.

Wang 2017 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare that they have no competing interests."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Wang 2017 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2019a

Study characteristics			
Patient Sampling	A total of 535 patients with chronic hepatitis B (CHB), including 176 HCC patients and 359 CHB patients with other liver diseases, were retrospectively enrolled at the Affiliated Hospital of Northern Sichuan Medical College from January 2017 to March 2019. A total of 359 CHB patients, including 186 with cirrhosis, 53 with cholecystitis, 37 with bile duct stones, 21 with drug-induced hepatitis, 51 with alcoholic hepatitis, 8 with hepatitis E. infection, and 3 with hepatitis C infection. Age range: 39-62. Males 81%		
Patient characteristics and setting			
Index tests	Serum levels of AFP were measured by electrochemiluminescence immunoassay. No predefinition of a cut-off value		
Target condition and reference standard(s)	The diagnosis of HCC was made in accordance with the standards of the guidelines for the diagnosis and treatment of primary HCC issued by the Chinese Society of Clinical Oncology. Controls: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors have no conflicts of interest to declare."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Wang 2019a *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Wang 2019a *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wang 2019b

Study characteristics			
Patient Sampling	The study population consisted of 90 patients with HBV-associated HCC, 90 patients with HBV-associated liver cirrhosis (LC), 90 patients with CHB, and 90 healthy people. HCC patients and LC patients were admitted at the Second Affiliated Hospital of Harbin Medical University between January 2017 and December 2017. CHB patients and healthy participants were recruited from the Second Affiliated Hospital of Harbin Medical University. They were matched for age, gender, and body mass index. Age range not reported. Males 61%		
Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification. No predefinition of a cut-off value		
Target condition and reference standard(s)	The diagnosis of HCC was confirmed by histology, and none of HCC patients received any form of treatment before enrolment. Cirrhosis was diagnosed based on a biopsy or on a combination of clinical, endoscopic, and radiological evidence of portal hypertension or cirrhosis.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare that they have no conflicts of interest."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Wang 2019b *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Wang 2019b *(Continued)*

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Weiss 2019

Study characteristics			
Patient Sampling	60 patients with chronic hepatitis C were subdivided into 3 cohorts: mild disease (fibrosis stage F0-2; n = 20); cirrhosis (n = 20); and cirrhosis with HCC (n = 20). Age range not reported. Males 60%		
Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification. Predefined cut-off value 20 ng/mL		
Target condition and reference standard(s)	The diagnosis of HCC was established according to currently accepted professional guidelines. Controls: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare no conflicts of interest."		

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Weiss 2019 *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Weiss 2019 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Wong 2008

Study characteristics			
Patient Sampling	"All patients referred from different hospitals and clinics to the Joint Hepatoma Clinic of the Chinese University of Hong Kong, which is a tertiary centre for the management of suspected or confirmed liver tumours, between January 2003 and June 2005 were retrospectively studied. Nine hundred and eighty-two patients visited the Joint Hepatoma Clinic during the study period. Five hundred and seventy-nine patients were included in the analysis after excluding patients who had non viral hepatitis related disease (n = 168), had missed records (n = 79), had pre-existing HCC (n = 30), had no USG (n = 119) and no AFP test (n = 7). Patients having chronic liver diseases caused by other aetiologies including alcoholic liver disease, autoimmune liver disease and primary biliary cirrhosis were excluded from the analysis." Age range: 46-70. Males 83%		
Patient characteristics and setting			
Index tests	Patients with elevated AFP levels (defined as > 20 ng/mL) and/or with suspicious lesions (any space occupying lesion) on USG were referred to the Joint Hepatoma Clinic for further investigation.		
Target condition and reference standard(s)	All patients were assessed with same-day US and AFP testing in the clinic within 2 weeks of referral. HCC was confirmed by histologic evidence of HCC from liver biopsy, typical appearance of tumour in a triphasic computerised tomography (CT) scan, characteristic lipoidal uptake in CT and/or neovascularisation and arterio-venous shunting in hepatic arteriography.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Conflict of interest: no conflicts of interest exist.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Wong 2008 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		

Wong 2008 *(Continued)*

Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Wong 2009

Study characteristics	
Patient Sampling	"This was a nested case-control study. All patients who subsequently developed HCC were identified. Control participants were age- (62 years) and gender-matched chronic hepatitis B patients without HCC. By February 2008, 37 patients developed HCC, and 37 age- and gender-matched control subjects were identified. Consecutive patients with chronic hepatitis B were recruited from the liver clinic of the Prince of Wales Hospital, Hong Kong, from December 1997 to July 2000.9 We included patients aged 18 years or above who had positive hepatitis B surface antigen for at least 6 months. We excluded patients who were co-infected by hepatitis C virus or HIV. Patients with other concomitant chronic liver diseases (e.g. haemochromatosis, Wilson's disease, primary biliary cirrhosis, autoimmune hepatitis, drug-induced liver injury) were excluded. We excluded patients who consumed more than 20 g of alcohol per week." Age range: 46-61. Males 89%
Patient characteristics and setting	
Index tests	AFP: conventional cut-off value was set at 20 ng/mL.
Target condition and reference standard(s)	Ultrasound of the abdomen, computerised tomography, hepatic angiogram and/or liver biopsy were performed if AFP levels were higher than 50 lg/L or demonstrated a rising trend over 20 lg/L to confirm the diagnosis of HCC. For patients with normal AFP levels, ultrasound scan was performed every 1–2 years.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Conflicts of Interest Statement: Prof. Francis Chan has served as a speaker, a consultant and an advisory board member for Pfizer, and a speaker for TAP Pharmaceuticals, AstraZeneca and Takeda. Prof. Joseph Sung received consulting fees from the National Health Research Institutes of Taipei, The Hong Kong Police Force, Lippincott Williams & Wilkins and the Hong Kong College of Physicians, and received lecture fees from AstraZeneca Hong Kong Limited, GSK Pharmaceuticals International and the American Society for Gastrointestinal Endoscopy. Prof. Henry Chan is a member of the advisory board of Novartis, Schering-Plough and Pharmasset.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Wong 2009 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wong 2014a

Study characteristics	
Patient Sampling	The entecavir cohort was composed of consecutive CHB patients who had received entecavir (0.5 mg) daily for at least 12 months in the hepatitis clinics, Prince of Wales Hospital, from December 2005 to March 2013. Patients who received entecavir before October 2009 were retrospectively identified from the HBV DNA record, and were recruited into the prospective follow-up study. All patients newly started on entecavir after October 2009 were recruited into the longitudinal study in a prospective manner. Patients suffering from chronic hepatitis C (CHC), pre-existing HCC, or HCC diagnosed within the first year of entecavir treatment were excluded. Age range: 40-60. Males 72%
Patient characteristics and setting	
Index tests	AFP: two cut-off values were used: 6 ng/mL was chosen because the sum sensitivity and specificity was the highest at this cut-off value and conventional 20 ng/mL.
Target condition and reference standard(s)	HCC: the diagnosis of HCC was established based on histopathological confirmation detection of a positive lesion with at least two imaging techniques (trans-abdominal USG, triphasic CT, magnetic resonance imaging, or hepatic angiogram), or detection with one imaging technique coupled with an AFP concentration greater than 400 lg/L.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	Potential conflicts of interest: Grace L.H. Wong has served as an advisory committee member for Otsuka and Gilead; she is also on the speakers' bureau for Echosens, Furui, and Otsuka. Henry L.Y. Chan is a consultant for Abbott, Bristol-Myers Squibb, Furui, Gilead, Merck, Novartis, and Roche, has received honoraria for lecturing for Abbott, Bristol-Myers Squibb, Echosens, Gilead, GlaxoSmithKline, Merck, Novartis, and Roche, and has received an unrestricted grant from Roche for hepatitis B research. He is on the speakers' bureau for Abbott, Bristol-Myers Squibb, Echosens, Gilead, GlaxoSmithKline, Merck, Novartis, and Roche. Vincent W.S. Wong has served as an advisory committee member for Roche, Novartis, Gilead, and Otsuka; he is also on the speakers' bureau for Bristol-Myers Squibb, Roche, Novartis, Abbott Diagnostics, and Echosens.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		

Wong 2014a *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wong 2014b

Study characteristics	
Patient Sampling	A historical control cohort of 424 treatment-naive patients was recruited from December 1997 to July 2000 from the hepatitis clinic at the Prince of Wales Hospital, Hong Kong. These patients underwent routine clinical care until the mid to late 2000s, when antiviral treatments were not readily available or re-imbursable. Patients suffering from chronic hepatitis C (CHC), preexisting HCC, or HCC diagnosed within the first year of entecavir treatment were excluded. Age range: 28-54. Males 65%
Patient characteristics and setting	
Index tests	AFP: cut-off predefined at 6 ng/mL
Target condition and reference standard(s)	HCC: the diagnosis of HCC was established based on histopathological confirmation detection of a positive lesion with at least two imaging techniques (trans-abdominal USG, triphasic CT, magnetic resonance imaging, or hepatic angiogram), or detection with one imaging technique coupled with an AFP concentration greater than 400 lg/L.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Potential conflicts of interest: Grace L.H. Wong has served as an advisory committee member for Otsuka and Gilead; she is also on the speakers' bureau for Echosens, Furui, and Otsuka. Henry L.Y. Chan is a consultant for Abbott, Bristol-Myers Squibb, Furui, Gilead, Merck, Novartis, and Roche, has received honoraria for lecturing for Abbott, Bristol-Myers Squibb, Echosens, Gilead, GlaxoSmithKline, Merck, Novartis, and Roche, and has received an unrestricted grant from Roche for hepatitis B research. He is on the speakers' bureau for Abbott, Bristol-Myers Squibb, Echosens, Gilead, GlaxoSmithKline, Merck, Novartis, and Roche. Vincent W.S. Wong has served as an advisory committee member for Roche, Novartis, Gilead, and Otsuka; he is also on the speakers' bureau for Bristol-Myers Squibb, Roche, Novartis, Abbott Diagnostics, and Echosens."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Wong 2014b *(Continued)*

DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wu 2009

Study characteristics	
Patient Sampling	Two groups of consecutive participants were enrolled into the present study. One group included 29 HBV-related HCC patients and the other group included 30 HBV-related liver cirrhosis (LC). Eligible criteria for HCC group included pathologically proven HBV-related HCC and LC; hepatitis B e antigen (HBeAg) positive or a quantity of HBV-DNA > 103/unit; no risk factors for HCV and he-

Wu 2009 *(Continued)*

	patitis D virus (HDV); alcohol consumption < 40 g/week; and no previous history of surgery, chemotherapy and radiotherapy Eligible criteria for LC group included pathologically proven HBV-related LC; HBeAg(+) or a quantity of HBV-DNA > 103/unit; no risk factors for HCV and HDV; and alcohol consumption < 40 g/week. Age range: 34-80. Males 88%
Patient characteristics and setting	
Index tests	AFP: AFP was detected by chemiluminescence assay at the Clinical Diagnostic Laboratory of Changzheng Hospital. The cut-off value was generated from the ROC curve.
Target condition and reference standard(s)	HCC and LC were pathologically proven.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Wu 2009 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Wu 2017

Study characteristics			
Patient Sampling	According to eligibility criteria listed in Table 1, we collected 1028 serum samples from the five groups of participants (Table 2). For each group, the age, sex, race, and the time and location sample collection were well matched. Age range: 37-61. Males 69%		
Patient characteristics and setting			
Index tests	Serum AFP measurement with a cut-off value of 20 ng/mL		
Target condition and reference standard(s)	HCC: histopathology Control: US or CT		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors declare no conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Wu 2017 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Wu 2017 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wu 2018

Study characteristics	
Patient Sampling	The study was performed in the second Xiangya Hospital of Central South University from November 2016 to March 2017. The groups were the following: 143 HCC patients, controls were 37 patients with chronic hepatitis, 43 patients with cirrhosis and enrolled during the same period as HCC cases. Healthy controls included 51 healthy volunteers. Age range: 31-65. Males 80%
Patient characteristics and setting	
Index tests	AFP: AFP was measured by the electrochemiluminescence immunoassay. The optimal cut-off values for PIVKA-II and AFP in differentiating HCC cases from non-cirrhotic chronic hepatitis and cirrhosis without HCC controls were 104 mAU/mL and 209.2 ng/mL, respectively.
Target condition and reference standard(s)	HCC: the diameter of the tumour was measured by ultrasound and/or CT.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare no competing financial interests."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Wu 2018 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Wu 2020

Study characteristics			
Patient Sampling	A total of 374 participants from Beijing YouAn Hospital were included in this study and divided into seven groups: the healthy control(HC), chronic hepatitis B (CHB), liver cirrhosis (LC), very early stage HCC, early stage HCC, advanced stage HCC and late stage HCC groups. Exclusiion criteria: combined hepatocellular and cholangiocarcinoma, intrahepatic cholangiocarcinoma, mixed HCC, HCC without HBV infection and HCC with HCV infection. Age range not reported. Males 63.5%		
Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification. No predefinition of a cut-off value		
Target condition and reference standard(s)	The diagnosis of HCC was confirmed by histological examination. LC and CHB groups underwent magnetic resonance imaging and were followed up for six months to exclude potential HCC		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors report no conflicts of interest in this work		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Wu 2020 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Xing 2019

Study characteristics	
Patient Sampling	Data were collected on consecutive patients referred to the Department of Hepatobiliary Surgery at Eastern Hepatobiliary Surgery Hospital in Shanghai between April 2016 and June 2017 Four groups: HCC, cirrhosis, chronic hepatitis, benign liver disease Exclusion criteria: unavailability of baseline AFP; presence of malignant tumours other than HCC (including intrahepatic cholangiocarcinoma, colorectal liver metastases, gallbladder cancer); absence of important variables regarding clinical and tumour characteristics Age range: 23-80. Males 80%
Patient characteristics and setting	
Index tests	Serum concentrations of AFP were measured with the commercially available ARCHITECT immunoassay. The predefined cut-off value was 20 ng/mL
Target condition and reference standard(s)	The diagnosis of HCC was made based on the imaging tests, such as angiography, computed tomography and magnetic resonance imaging, according to accepted guidelines. Diagnosis of liver cirrhosis was established by ultrasound test, CT scan or magnetic resonance imaging and supplemented by portal hypertension symptom (anorectal varices, splenomegaly, thrombocytopenia).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"This work was supported in part by the National Natural Science Foundation of China (no. 81472284 and 81672699). The authors have no other relevant affiliations or financial involvement with any organisation or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Xing 2019 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Xing 2019 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Xu 2018

Study characteristics			
Patient Sampling	291 participants divided into four age- and gender-matched groups, including a HCC group (n = 88), a liver cirrhosis (LC) group (n = 67), a chronic hepatitis B (CHB) group (n = 68) and a healthy control group (n = 68), were enrolled. The participants with autoimmune hepatitis, alcoholic liver disease, Wilson's disease, other types of viral hepatitis and other major diseases were excluded.		
Patient characteristics and setting			
Index tests	AFP cut-off value pre-specified at 20 ng/mL		
Target condition and reference standard(s)	HCC was diagnosed based on ultrasound, computed tomography (CT), serum AFP, and histopathological examination.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"Competing interests: the funding organisation(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Xu 2018 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yan 2018

Study characteristics	
Patient Sampling	A total of 86 patients were enrolled in this study, including 62 patients with HBV-related liver fibrosis and 24 patients with HCC. The 62 liver fibrosis patients were selected randomly from the China HepB Related Fibrosis Assessment Research cohort supported by the China Mega-project for Infectious Diseases. The 24 HCC patients were included from those who had been diagnosed with

Yan 2018 *(Continued)*

	HCC by imaging and pathological evaluations at the Peking Union Medical College Hospital and Henan Cancer Hospital. Exclusion criteria were: hepatitis C virus or human immunodeficiency virus coinfection; presence of other causes of chronic liver diseases such as alcoholic, autoimmune, genetic, drug-induced, and nonalcoholic fatty; and pregnancy. Age range: 28-66. Males 76%
Patient characteristics and setting	
Index tests	AFP: optimal cut-off level was 80.5 ng/mL
Target condition and reference standard(s)	The 24 HCC patients were included from those who had been diagnosed with HCC by imaging and pathological evaluations. Controls: All of the selected liver fibrosis patients underwent a liver biopsy; the degree of inflammation and the fibrosis stage were assessed according to the Ishak criteria.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflicts of interest are declared by any of the authors

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Yan 2018 *(Continued)*

DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yang 2013a

Study characteristics	
Patient Sampling	Plasma was collected from 179 HCC patients (35 females and 144 males, with a mean age of 54.0 years) before hepatectomy at the Cancer Hospital in China. All HCC patients were chronically infected with HBV. Cirrhosis plasma was obtained from 80 liver cirrhosis patients (24 females and 56 males, with a mean age of 53.5 years) with chronic HBV infection at Beijing You'an Hospital, Capital Medical University. Age range not reported. Males 77%
Patient characteristics and setting	
Index tests	AFP levels were tested using a commercial immunoassay with enhanced chemiluminescence at the Clinical Diagnostic Laboratories of the Cancer Hospital, Chinese Academy of Medical Sciences. Cut-off value 20 ng/mL
Target condition and reference standard(s)	The histological diagnosis of the tissue samples was confirmed by experienced pathologists.
Flow and timing	No information on interval between index test and reference standard

Yang 2013a *(Continued)*

Comparative	
Notes	"This work was supported by the National Natural Science Foundation of China (81172035, 30973388), the National Excellent Doctoral Dissertation of China (2007B68), the National High Technology Research and Development Program of China (2012AA020206), and the Basic Research Program of the Cancer Hospital, PUMC & CAMS (JK2009B08, LC2009B45). The authors declare no competing financial interest."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Yang 2013a *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yang 2014

Study characteristics			
Patient Sampling	A total of 200 individuals visiting the Qilu Hospital of Shandong University from July 2011 to December 2012 were consecutively enrolled, including 123 patients with HBV-related HCC, 28 patients with liver cirrhosis, 29 patients with chronic hepatitis B, and 20 healthy controls. Exclusion criteria: secondary liver cancer from other primary origins, history of other solid tumour. Age range and % males not reported		
Patient characteristics and setting			
Index tests	AFP: cut-off predefined at 20, 200, and 400 ng/mL		
Target condition and reference standard(s)	Patients with HCC were diagnosed based on the guidelines of the American Association for the Study of Liver Disease (2005). HCC was defined on the basis of at least two dynamic imaging modalities including angiography, computed tomography (CT), and magnetic resonance imaging (MRI), or by tumour biopsy. Control group: all patients with CHB or LC were confirmed not having HCC using ultrasonography or CT; no patients had newly developed HCC for at least 3 months before enrolment.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	The authors declare no conflict of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Yang 2014 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yang 2017

Study characteristics	
Patient Sampling	The study included 233 consecutive early-stage HCC patients and 412 cirrhotic patients without HCC seen between February 2005 and August 2007 at seven tertiary referral centres in the USA. Age range: 45-64. Males 70%
Patient characteristics and setting	
Index tests	AFP: optimal cut-off value was calculated to be 9.9 ng/mL. Serum AFP was measured by automated systems (Wako) at the time of enrolment prior to HCC-specific treatment.
Target condition and reference standard(s)	HCC was defined by histopathological examination or by the specific radiologic characteristics endorsed by AASLD. All controls were assessed by AFP and imaging 6 months after enrolment to ensure that they did not have HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"A.G. Singal has received speakers bureau honoraria from Bayer and is a consultant/advisory board member for Bayer and Wako Diagnostics. No potential conflicts of interest were disclosed by the other authors."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Yang 2017 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yang 2019

Study characteristics	
Patient Sampling	The control group consisted of patients who were candidates for HCC surveillance, namely those with cirrhosis or chronic hepatitis B without HCC seen at Mayo Clinic between October 2013 and October 2016, (1) who were tested for AFP, as part of their regular clinical care or (2) had provided stored serum with research consent authorisation for the measurement of AFP. The case group consisted of patients with newly diagnosed HCC in the setting of cirrhosis or chronic hepatitis B during the same study period, (1) who were tested for AFP, as part of their regular clinical care or (2) had provided stored serum with research consent authorization for the measurement of AFP at the time of tumour diagnosis. Age range not reported. Males 62%
Patient characteristics and setting	
Index tests	Serum AFP measurement with 20 ng/mL as cut-off value. Serum biomarkers were measured using the WAKO mTASWako i30 Immunoanalyzer. US no specification

Yang 2019 *(Continued)*

Target condition and reference standard(s)	The diagnosis of HCC was made by dynamic contrast CT or MRI of the liver or biopsy based on the guidelines of the AASLD. Control patients were required to have at least 6 months of follow-up after GALAD score assessment to confirm the absence ofHCC or have a negative contrast-enhanced multiphasic CT, MRI, or liver biopsy at the time of GALAD score assessment.
Flow and timing	Demographic and clinical characteristics of patients were abstracted closest to the time of blood collection within a maximum time window of 3 months.
Comparative	
Notes	M.H. Nguyen reports receiving commercial research support from Bristol-Myers Squibb, Janssen Pharmaceuticals, Gilead Sciences, Laboratory for Advanced Medicine, and Exact Science; received honoraria from the speakers bureau of Bristol-Myers Squibb, Janssen Pharmaceuticals, Intercept Pharmaceuticals; Roche Laboratories, Dynavax Laboratory, Alnylam Pharmaceuticals, Novartis, Laboratory for Advanced Medicine, and Eisai Science. L.R. Roberts reports receiving commercial research funding from Ariad, Wako, Gilead, BTG, and Bayer, and Redhill; received honoraria from the speakers' bureau of Wako, Medscape, NACCME, and OncLive; and is a consultant/advisory board member for Bayer, Exact Sciences, Tavec, and Grail. No other potential conflicts of interest were disclosed by the other authors.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Yang 2019 *(Continued)*

DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yao 2016

Study characteristics	
Patient Sampling	A total of 1845 patients diagnosed either with chronic hepatitis, cirrhosis, or HCC with different backgrounds were enrolled between December 2008 and December 2013 at Henan Cancer Hospital in Zhengzhou, and Beijing Hospital. The study included 318 cases of hepatitis, 731 cases of cirrhosis and 796 HCC cases. Age range: 31-65. Males 79.5%

Yao 2016 *(Continued)*

Patient characteristics and setting	
Index tests	AFP: Receiver operating characteristic (ROC) curves were plotted to identify a cut-off value that would best distinguish HCC patients from the other two groups of participants. The optimal cut-off value for AFP was 11.62 ng/mL. The measurement of AFP in the two hospitals were achieved by using same electrochemiluminescence immunoassay system Modular E170 (Roche, Mannheim, Germany). The normal range is 0 ng/mL to 20 ng/mL.
Target condition and reference standard(s)	The diagnosis of HCC was confirmed by pathologic examination of the resected liver specimens. Patients were examined for HCC by abdominal ultrasonography, dynamic CT, and/or MRI every 3-6 months.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No conflict of interest was disclosed in this study.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Yao 2016 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ye 2019a

Study characteristics	
Patient Sampling	This study enrolled a total of 1244 participants, including HCC, healthy controls , benign liver tumours , chronic hepatitis B, and liver cirrhosis Age range and % males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement with no predefined cut-off value
Target condition and reference standard(s)	Diagnosis of LC was based on a history of CHB infection, confirmed by biopsy or two imaging technologies, i.e. hepatic ultrasound with CT or MRI. To limit the possible presence of early-stage HCC clinically unrecognised in cirrhosis Patients with cirrhosis with < 20 years of chronic hepatitis history and in compensated phase of the disease were preferred. HCC was diagnosed based on ultrasound, CT, or MRI and AFP serology and confirmed by histopathology according to guidelines of the American Association for the Study of Liver Disease (AASLD).
Flow and timing	No information on interval between index test and reference standard
Comparative	

Ye 2019a *(Continued)*

Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			

Ye 2019a *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ye 2019b

Study characteristics	
Patient Sampling	This study enrolled a total of 1244 participants, including HCC, healthy controls, benign liver tumours, chronic hepatitis B, and liver cirrhosis Age range and % males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. No predefinition of a cut-off value
Target condition and reference standard(s)	Diagnosis of LC was based on a history of CHB infection, confirmed by biopsy or two imaging technologies, i.e., hepatic ultrasound with CT or MRI. To limit the possible presence of early-stage HCC clinically unrecognised in cirrhosis Patients with cirrhosis with < 20 years of chronic hepatitis history and in compensated phase of the disease were preferred. HCC was diagnosed based on ultrasound, CT, or MRI and AFP serolology and confirmed by histopathology according to guidelines of the American Association for the Study of Liver Disease (AASLD).

Ye 2019b *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Ye 2019b *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yoon 2009

Study characteristics	
Patient Sampling	Between April 2001 and December 2006, 3147 patients with chronic HBV infections were under surveillance for HCC detection at Severance Hospital, Seoul, Korea. The surveillance program included ultrasonography (US) and AFP every 3, 6 or 12 months. During surveillance, 100 randomly selected non-HCC test results were included in the control group. A total of 113 patients with chronic HBV infections were found to have HCC while under surveillance. Of these patients, 7 whose serum PIVKA-II was not measured at the time of HCC diagnosis were excluded from the analysis; 106 patients were included in the case group. Age: 48-63. Males 82%
Patient characteristics and setting	
Index tests	AFP: serum AFP was measured by electrochemiluminescence assay using a cut-off value of 20 ng/mL (Roche Diagnostics, Mannheim, Germany).

Yoon 2009 *(Continued)*

Target condition and reference standard(s)	HCC was diagnosed either histologically or by typical HCC imaging patterns using angiography, computed tomography (CT) and/or magnetic resonance imaging (MRI).
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors reported no conflicts of interest. The authors wer responsible for the content and writing of the paper.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Yoon 2009 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Youns 2013

Study characteristics			
Patient Sampling	120 patients were included in the study and divided into three groups: 40 patients with HCC, 40 patients with liver cirrhosis and 40 healthy individuals. Age range not reported. Males 56%		
Patient characteristics and setting			
Index tests	AFP: optimal cut-off value calculated at 9 ng/mL		
Target condition and reference standard(s)	HCC: diagnosis of HCC was based on elevated AFP values, the presence of focal hepatic lesion detected by liver ultrasound, and confirmed by CT or MRI.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Youns 2013 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yu 2011

Study characteristics	
Patient Sampling	This retrospective study provides a broad survey of the accuracy of US, CT, and MRI for HCC detection in a large population of cirrhotic patients undergoing liver transplantation in a single major USA transplantation centre Query of our database yielded 1097 adults receiving orthotopic liver transplantation at our institution from January 1999 to November 2006. Of these, 638 consecu-

Yu 2011 *(Continued)*

	tive patients (407 men, 231 women; age 18–75 years, mean 53.2) with chronic liver disease who underwent unenhanced US, contrast-enhanced single or multidetector helical CT, and/or dynamic contrast-enhanced MRI at our institution within 6 months of the transplantation comprised the study population. HCC was confirmed in 638 patients. Age range: 18-75. Males 64%
Patient characteristics and setting	
Index tests	AFP: cut-off values prespecified at 10 ng/mL and 20 ng/mL; US no definition of positivity criteria
Target condition and reference standard(s)	Reference standard: pathology of the explanted liver
Flow and timing	Patients who had imaging modalities performed within 6 months of the transplantation comprised the study population.
Comparative	
Notes	The author discloses the following: "Dr Lu has received educational and research support from Philips and GE Healthcare. The other authors disclose no conflicts of interest."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Yu 2011 *(Continued)*

DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Yu 2016

Study characteristics	
Patient Sampling	In the Hepatitis Biobank of Southwest Hospital (HBS) cohort at Southwest Hospital, we did a two-stage nested case-control study. Totally, 51 HCC cases versus 138 matched controls were enrolled to compare levels of α-fetoprotein (AFP) and PIVKA-II in sequential sera at −12, −9, −6, −3 and 0 months before imaging diagnosis. At-risk controls were randomly selected and matched according to age, gender and liver cirrhosis status. Patients receiving warfarin or vitamin K before haemospasia were screened out for the influence on PIVKA-II level. Age range: 39-67. Males 74%

Yu 2016 *(Continued)*

Patient characteristics and setting			
Index tests	AFP: serum levels of AFP were measured by AFP. Reagent kit via chemiluminescent microparticle immunoassay (CMIA) (ARTHITECT i2000, Abbott Laboratories, America). Cut-off values prespecified at 5, 20, 400, 200 ng/mL		
Target condition and reference standard(s)	HCC: any participants diagnosed as HCC should met two imaging criteria (hepatic ultrasound plus CT or MRI), and then all cases were confirmed by biopsy.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare no competing financial interests."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		

Yu 2016 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yu 2020a

Study characteristics			
Patient Sampling	The included patients were subdivided in: - chronic hepatitis B patients (CHB group) - HBV-related liver cirrhosis patients (liver cirrhosis group) - HBV-related HCC patients (HCC group) Age range and % males not reported		
Patient characteristics and setting			
Index tests	AFP measurement with a cut-off value of 20 ng/mL		
Target condition and reference standard(s)	HCC diagnosed by two experienced pathologists. If no tissue available, diagnosis must be supported by two image reports (ultrasound B, CT, or MRI). Cirrhosis diagnosed by two experienced pathologists. If no tissue available, diagnosis must be supported by two image reports (ultrasound B, CT, or MRI). CHB no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare no potential conflicts of interest."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Yu 2020a *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		

Yu 2020a *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yu 2020b

Study characteristics	
Patient Sampling	The include patients were subdivided in: - chronic hepatitis B patients (CHB group) - HBV-related liver cirrhosis patients(liver cirrhosis group) - HBV-related HCC patients (HCC group) Age range and % males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. Predefined cut-off value 20 ng/mL
Target condition and reference standard(s)	HCC diagnosed by two experienced pathologists. If no tissue available, diagnosis must be supported by two image reports (ultrasound B, CT, or MRI). Cirrhosis diagnosed by two experienced pathologists. If no tissue available, diagnosis must be supported by two image reports (ultrasound B, CT or MRI). CHB no definition.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declare no potential conflict of interest.
Methodological quality	

Yu 2020b *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			

Yu 2020b *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Yu 2020c

Study characteristics	
Patient Sampling	The included patients were subdivided in : - chronic hepatitis B patients (CHB group) - HBV-related liver cirrhosis patients(liver cirrhosis group) - HBV-related HCC patients (HCC group) Age range and % males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement: no specification. Predefined cut-off value 20 ng/mL
Target condition and reference standard(s)	HCC diagnosed by two experienced pathologists. If no tissue available, diagnosis must be supported by two image reports (ultrasound B, CT, or MRI). Cirrhosis diagnosed by two experienced pathologists. If no tissue available, diagnosis must be supported by two image reports (ultrasound B, CT, or MRI). CHB no definition
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare no potential conflict of interest."

Yu 2020c *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			

Yu 2020c *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zekri 2013

Study characteristics	
Patient Sampling	"The study included 130 patients with chronic HCV genotype 4 (HCVG4)-associated liver disease and HCC. Patients were classified into four groups. Group I: [chronic, asymptomatic carriers group (AC)], which included 30 patients with chronic HCVG4 and persistent normal liver profile. Group II: [chronic active hepatitis (CAH) non-cirrhotic patients], which included 30 patients with elevated liver enzymes and no cirrhosis in liver biopsy. Group III: [cirrhotic hepatitis C patients], which included 30 patients with cirrhosis on top of CAH as confirmed by liver biopsy (F5-6/6 by Ishak score) and Group IV: [HCC patients], which included 40 HCC patients. Patients with any cause of liver disease other than HCV, other malignancies, a family history of malignancy and those with any contraindication to liver biopsy were excluded from the study." Age range: 25-58. Males 75%
Patient characteristics and setting	
Index tests	"AFP: all patients were subjected to complete clinical assessment and laboratory investigations including Quantitative Real Time PCR (Stratagene, USA) for HCV, CBC, liver profile, INR, alfa fetoprotein (AFP), ANA and HCV antibody (using Axyam-Abbot). The best cutoff for AFP was 10.35 ng/mL."
Target condition and reference standard(s)	"HCC: HCC patients diagnosed according to BCLC guidelines and by histopathological examination of ultrasound-guided liver biopsies taken from the focal lesions.

Zekri 2013 *(Continued)*

	Controls: Group II: [chronic active hepatitis (CAH) non-cirrhotic patients], which included 30 patients with elevated liver enzymes and no cirrhosis in liver biopsy. Group III: [cirrhotic hepatitis C patients], which included 30 patients with cirrhosis on top of CAH as confirmed by liver biopsy."
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Zekri 2013 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Zhan 2020

Study characteristics	
Patient Sampling	The included patients were subdivided in three groups: CHB, n = 10), liver cirrhosis (LC, n = 10), hepatocellular carcinoma (HCC, n = 10) and healthy control participants (HC, n = 10) Age range and % males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement with commercial kit. No pre-definition of a cut-off value
Target condition and reference standard(s)	Diagnoses of CHB, LC and HCC were made following the clinical practice guidelines set forth by the American Association for the Study of Liver Diseases (AASLD), the European Association for the Study of the Liver (EASL) and/or the Asian Pacific Association for the Study of the Liver (APASL)
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"The authors declare no potential conflicts of interest."

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Zhan 2020 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Zhan 2020 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zhang 2020

Study characteristics			
Patient Sampling	"From August 2015 to August 2017, healthy individuals and patients with HBV-related diseases who were older than 18 years were recruited from the Department of Hepatobiliary Surgery, Xijing Hospital, and the Center of Infectious Diseases, Tangdu Hospital, Xi'an City, Shaanxi Province, China. All patients were positive for HBsAg, and none of the patients had any other type of liver disease, such as chronic hepatitis C infection, alcoholic liver disease, autoimmune liver disease, or metabolic liver disease." Age range and % males not reported		
Patient characteristics and setting			
Index tests	Serum AFP measurement: no specification; predefined cut-off value 29 ng/mL		
Target condition and reference standard(s)	The diagnosis of HCC and cirrhosis was histopathologically confirmed.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare no conflicts of interest for this article."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Zhang 2020 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Zhang 2020 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zheng 2017

Study characteristics			
Patient Sampling	333 cases of HCC and 164 cases of cirrhosis were recruited in China. Age range and % males not reported		
Patient characteristics and setting			
Index tests	AFP: optimal cut-off determined at 30.5 ng/mL		
Target condition and reference standard(s)	Not reported		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No information on conflicts of interest		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	

Zheng 2017 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Zhou 2012

Study characteristics	
Patient Sampling	The present study was performed between September 2009 and August 2010. After informed consents were obtained, 118, 94 and 47 serum samples from the inpatients with HCC, chronic liver disease, and liver cirrhosis were collected from the Department of Hepatitis, The Second Hospital of Nanjing, China.

Zhou 2012 *(Continued)*

	Age range: 33-65. Males 77%
Patient characteristics and setting	
Index tests	AFP:tThe qualitative measurement of serum AFP was performed by using enzyme immunoassay method with commercial kit (Abbott Laboratories, USA). According to the instruction of manufacture, the normal range of AFP was 0 ng/mL to 10.9 ng/mL.
Target condition and reference standard(s)	Diagnosis of HCC relied on the presence of a malignant liver nodule (> 1.0 cm), as established on imaging techniques and by pathological analysis of liver biopsies. Patients with LC were determined by CT, ultrasonography, and pathological analysis.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	The authors declared no potential conflicts of interest.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			

Zhou 2012 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zhou 2019

Study characteristics	
Patient Sampling	361 participants were enrolled including healthy controls and patients with chronic hepatitis B, liver cirrhosis, and hepatocellular carcinoma (HCC). Exclusion criteria for this study: patients with liver cirrhosis or HCC with overlapping etiologies for hepatitis including HCV, human immunodeficiency virus (HIV), alcohol abuse, autoimmune, genetic, drug-induced and nonalcoholic fatty liver disease Age range and % males not reported
Patient characteristics and setting	
Index tests	Serum AFP measurement. No pre-definition of a cut-off value
Target condition and reference standard(s)	The diagnosis of LC was based on a history of CHB infection and confirmed by liver biopsy or imaging techniques, i. e. computed tomography (CT), or magnetic resonance imaging (MRI). Patients with HCC were diagnosed based on ultrasound, CT or MRI and AFP serology, and the diagnosis was ultimately confirmed by a liver biopsy, according to expert consensus.
Flow and timing	No information on interval between index test and reference standard
Comparative	

Zhou 2019 *(Continued)*

Notes	The authors report no conflicts of interest in this work.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			

Zhou 2019 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zhu 2013

Study characteristics			
Patient Sampling	The study included a cohort of 86 hepatocellular carcinomas with early stage (BCLC-0/A) and 40 patients with liver cirrhosis. Age range: 25-78. Males 75%		
Patient characteristics and setting			
Index tests	AFP: cut-off value predefined at 20 ng/mL		
Target condition and reference standard(s)	HCC: pathology or biopsy		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	No potential conflicts of interest were disclosed.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Zhu 2013 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zhu 2020

Study characteristics			
Patient Sampling	322 patients (105 chronic hepatitis, 116 liver cirrhosis, and 101 HCC Age range not reported. Males 64%		
Patient characteristics and setting			
Index tests	Serum AFP measurement. Predefined cut-off values 20 ng/mL and 100 ng/mL		
Target condition and reference standard(s)	HCC patients were diagnosed by histological findings or typical imaging characteristics Controls: no definition		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	"The authors declare that there are no conflict of interests."		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Zhu 2020 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 2: Index Test (US)			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Could the conduct or interpretation of the index test have introduced bias?			
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ziada 2016

Study characteristics	
Patient Sampling	This is a cross-sectional study which prospectively included all adult patients with chronic HCV from the outpatient clinics and

Ziada 2016 *(Continued)*

	inpatient wards of Tropical Medicine and Infectious Diseases Department at Tanta University Hospital, Tanta, Egypt. 103 patients were confirmed to have HCC, and 411 had no HCC. Exclusion criteria: patients with diagnosed HCC, hepatic metastasis, and prior HCC treated lesions Age range: 29-78. Males 81%
Patient characteristics and setting	
Index tests	AFP: cut-off values were predefined at 100 ng/mL and 200 ng/mL
Target condition and reference standard(s)	All cases were screened for HCC using ultrasonography and AFP. Individuals with solid focal lesion in ultrasound examination, and/or serum AFP level > 200 ng/mL were examined by tri-phasic CT, and/or MRI to confirm or roll out the diagnosis of HCC.
Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	"Authors declare no conflicts of interest."

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			

Ziada 2016 *(Continued)*

DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zinkin 2008

Study characteristics	
Patient Sampling	Two groups of patients were studied. The first group consisted of 41 patients with clinical or biopsy-confirmed cirrhosis complicated by HCC. The second group included 51 patients with compensated HCV cirrhosis and no HCC. Age range: 23-82. Males 83%
Patient characteristics and setting	
Index tests	AFP: total AFP (AFP-L1 + AFP-L3) and the percentage of AFP-L3 contents of serum samples were determined using the LiBASys clinical autoanalyzer (Wako Diagnostics). For AFP, the prespecified cut-off value was 20 ng/mL .
Target condition and reference standard(s)	HCC: the diagnosis of HCC was confirmed by at least one of the following: (a) histology, (b) new hepatic lesion with an AFP of > 1000 ng/mL, and/or (c) new hepatic lesion with arterial phase enhancement on computed tomography or magnetic resonance imaging. Control group: the cirrhosis group had at least 2 years of follow-up from the time serum was obtained for these studies. The follow-up included ultrasound and AFP every 6 months for at least 2 years with no evidence of development of HCC.

Zinkin 2008 *(Continued)*

Flow and timing	No information on interval between index test and reference standard
Comparative	
Notes	No information on conflicts of interest

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			

Zinkin 2008 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Zuo 2016

Study characteristics			
Patient Sampling	A total of 90 HBV-related HCC patients, 30 healthy controls, and 30 chronic hepatitis B (CHB) patients were included in this study. The exclusion criteria for healthy controls were as follows: age < 18 years, any treatment prior to surgery, positive markers of hepatitis viruses, and history of malignant disease in the preceding 5 years. Inclusion criteria were as follows: patients with chronic hepatitis B (CHB) > 18 years of age, no history of any cancer, positive for HBsAg for at least 6 months prior to the start of the study, and no infection with other hepatitis viruses. Age range: 31-65. Males 67.5%.		
Patient characteristics and setting			
Index tests	AFP: serum AFP level was analysed according to the manufacture's instruction by enzyme-linked immunosorbent assay (ELISA) kit (Cusabio, China and eBioscience, San Diego, CA). AFP prespecified at 20 ng/mL.		
Target condition and reference standard(s)	The diagnosis of HCC was confirmed by histopathology.		
Flow and timing	No information on interval between index test and reference standard		
Comparative			
Notes	Conflicts of interest: none		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Zuo 2016 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index Test (AFP)			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index Test (US+AFP)			
DOMAIN 2: Index Test (US)			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

AASLD: American Association for the Study of Liver Diseases;**AFP:** alpha-foetoprotein;**AIH:** autoimmune hepatitis; **ALT:** aminotransferase; **AST:** and aspartate aminotransferase; **AUC:** area under the curve; **BCLC:** Barcelona Clinic Liver Cancer; **CECT: c**ontrast-enhanced computed tomography; **CEMRI;** contrast-enhanced magnetic resonance imaging**; CEUS:** contrast-enhanced ultrasound; **CT:** computed tomography; **CHB:** chronic hepatitis B;**CHC:** chronic hepatitis C;**CT:** computer tomography; **DCP:** des-gamma-carboxy prothrombin; **ELISA:** enzyme-linked immunosorbent assay); **HBsAg:** serum hepatitis B surface antigen; **HCC:** hepatocellular carcinoma; chronic hepatitis B (CHB); **CHC:** chronic hepatitis C;**HCV:** hepatitis C virus; **IOCT:** iodised oil computed tomography; **LC:** liver cirrhosis; **MELD:** model for end-stage liver disease; **MRI**: magnetic resonance imaging; **NASH:** nonalcoholic steatohepatitis; **NPV:** negative predictive values; **OLT:** orthotopic liver transplantation; **OPN:** osteopontin; **PEI**: percutaneous ethanol injection; **RFA**: radiofrequency ablation; **RNA:** ribonucleic acid; **PPV:** positive predictive values; **SD:** standard deviation; **SE:** standard error; **TACE:** transarterial chemoembolisation; **US:** ultrasound.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abdelgawad 2013	Participants with HCC (40) were compared to participants with cirrhosis (10) and healthy participants (10). The results of the comparison of participants with HCC and participants with liver cirrhosis are not available No definition of the reference standard
Abd El Gawad 2014	Participants with HCC (40) were compared with 10 participants with liver cirrhosis and 10 normal healthy participants. The results of comparison of participants with HCC and participants with liver cirrhosis are not available.
Abdel-Hafiz 2018	80 participants with HCC were compared with 20 healthy volunteers who were collected from the staff of the National Cancer Institute, Cairo University and Theodore Bilharz Research Institute (TBRI), Giza.
Abelev 1971	Participants with HCC compared with participants with other cancer and healthy volunteers
Abouzied 2017	25 participants with HCC compared with 50 healthy controls
Aburano 1979	The 2 by 2 table was not reported directly in the study, and could not be calculated/extracted based on the data that were available.
Asim 2010	No data on index tests (AFP, US, AFP+US)
Åström 2017	> 5 % included patients with recurrent HCC (5/32)
Bago 1993	Review. No original data
Baig 2009	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Banales 2019	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Bao 2013	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Baumgarten 2001	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Beaugrand 2000	Review; no original data
Ben Hassine 2007	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Bialecki 2006	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Bird 2016	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Biwole Sida 1992	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Bolondi 1990	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.

Study	Reason for exclusion
Bottelli 1998	Non-systematic review. No original data
Bowry 1980	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Bröker 2014	Only participants without chronic liver disease and with focal liver lesions i.e. hepatic adenoma, focal nodular hyperplasia and HCC were included.
Buell 2001	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Cai 2019	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Carriere 1993	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were reported.
Chen 1995	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were reported.
Chen 2002	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were reported.
Chen 2010	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Chen 2011	41 participants with HCC compared to 38 healthy controls
Chen 2013	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Cheng 2009	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were reported.
Choi 2013	9.6% of patients with recurrent HCC
Chun 2015	Study conducted in general population undergoing "routine health check ": only less than 5 % (2286/49381) participants with chronic viral hepatitis.
Colombo 1991	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were reported.
Cui 2013	Comparison of 175 HCC with 80 cirrhosis patients and 105 healthy volunteers. Data on comparison with 80 cirrhotics not available
Del Vecchio-Blanco 1977	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Dengler 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Deshpande 1981	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Di Martino 2013	Reported only analyses per lesion and not per patient

Study	Reason for exclusion
Ding 1995	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Di Poto 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Di Poto 2018	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Divella 2012	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Donato 1995	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Dong 2008	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Dou 2016	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
El Abd 2016	50 participants with HCC compared to 30 participants with cirrhosis and 20 healthy participants. The results of comparison of participants with HCC and participants with liver cirrhosis are not available.
El-Ahwany 2019	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
El-Attar 2010	30 participants with HCC compared to 20 participants with cirrhosis and 20 healthy participants. The results of comparison of participants with HCC and participants with liver cirrhosis are not available.
El Azm 2013	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
El-Emshaty 2014	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
El-Emshaty 2015	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
El Gedawy 2017	30 participants with HCC compared to 20 participants with chronic liver disease and 20 healthy participants. The results of compariso0n of participants with HCCc and participants with chronic liver disease are not available.
Elghoroury 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
El-Mazny 2014	35 participants with HCC compared to 15 with liver cirrhosis and 10 healthy volunteers. No separate analysis available.
El-Saadany 2018	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
El-Serag 2005	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.

Study	Reason for exclusion
Elsewify 2020	Control group included healthy participants. No separate analysis was available.
Elshimi 2018	The results of comparison of participants with HCC and participants with liver cirrhosis are not available.
Eltabbakh 2015	The results of comparison of participants with HCC and participants with liver cirrhosis are not available.
El-Zefzafy 2015	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Esfeh 2020	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Fan 2017	52 participants with HCC compared with 30 participants with cirrhosis and 32 healthy participants. No separated analysis available.
Farag 2018	Control group included healthy participants. No separate analysis was available.
Fouad 2014	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Fouad 2015	Control group included healthy participants. No separate analysis was available.
Frey 2015	The results of comparison of participants with HCC and participants with liver cirrhosis are not available.
Fujiyama 1991	The results of comparison of participants with HCC and participants with liver cirrhosis are not available.
Gandolfi 1987	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Ganne-Carrié 1996	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Gao 2012	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Geramizadeh 2013	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Geramizadeh 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Gheorghe 1986	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available
Gheorghe 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Giangregorio 2009	Index test - contrast enhanced ultrasound, no data on standard US
Giannini 2012	Not pertinent: a prognostic study in a cohort of participants with HCC
Giannini 2013	Review; no original data

Study	Reason for exclusion
Giannini 2014	Comment to an included study Gopal 2014, no original data
Giusti 2005	Imaging findings for hepatocellular adenoma
Goldaracena 2019	Definition of a prognostic score for the development of HCC recurrence following liver transplantation
Gomez Rodriguez 2012	Assessment of the AFP measurement as prognostic factor for patients with HCC
Gomez Rubio 2005	Index test is laparoscopic US, no data on abdominal ultrasound
Gorbatenko 1974	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Gounder 2016	Analysis of the cost of screening using AFP/US
Grąt 2016	AFP for the prediction of HCC recurrence after OLT
Ha 2012	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Hacque 2016	Comparison of AFP in patients with HCC vs normal individuals; only mean values
Hagag 2020	Control group included healthy participants. No separate analysis was available.
Hajiani 2005	Not pertinent. Definition of risk factors for HCC, no data on diagnosis
Han 2018b	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Hanaoka 2011	No data available on diagnostic accuracy of AFP: only an assessment of the highly sensitive fucosylated fraction of α-fetoprotein in patients with AFP < 10 ng/mL.
Hashemi 2008	Assessment of the role of US in the differential diagnosis of liver masses. Only patients with known focal lesions in the liver are included.
Hass 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Hemken 2019	Control group including healthy participants. No separate analyses
Hernandez 2011	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Heyward 1985	Duplicate: reporting preliminary data fully reported in an included study McMahon 2000.
Hiraoka 2016	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available. Only patients with diabetes and without liver disease are included.
Hu 2010	HCC vs all others (including healthy individuals). No separate analysis available
Hussein 2008	HCC vs all the others (including healthy controls). No separate analysis available
Hwang A 2018	The study was conducted on animals.

Study	Reason for exclusion
Hwang H 2018	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Imberti 1993	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Izuno 1995	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Izzo 1998	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Jirun 2011	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Johnson 1997	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Jun 2019	Assessment of the prediction of future HCC
Junna 2017	Patients with HCC compared with healthy controls. No separated analysis available
Kim 2011a	Only patients with hepatic mass(es) >2 cm who underwent biopsy or surgical resection were included, no participants with chronic liver disease without HCC.
Kim 2011b	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Kim 2013	Assessment of tumour recurrence
Kim 2015	HCC vs all other groups (including healthy controls). No possibility to separate data
Kim 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
King 1989	HCC vs all other groups (including healthy controls). No separated analysis available
Kiyokawa 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Ko 2017	The study does not meet the inclusion criteria: the target condition is primary liver cancer including cholangiocarcinoma (30 % of cases) and participants were enrolled in a routine health examination.
Kobayashi 1985	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Kobeisy 2012	AFP to diagnose severe fibrosis
Kuromatzo 1993	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Larcos 1998	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.

Study	Reason for exclusion
Lee 2016	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Lee 2017	Surgical series including other tumours
Leerapun 2007	Index test: AFP-L3%. No data on AFP
Li 2016	HCC compared to healthy controls. No separated analysis available
Li 2018	The index test was a combination of AFP + CENP-F: it is not possible to separate data.
Li 2019b	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Li 2019c	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Liaw 1986	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Lim 2006	Only participants with HCC included. No per patient analysis
Liu 2003	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Liu 2008	Healthy participants in control group. No separated analysis available
Liu 2010b	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Liu 2013	90 participants with HCC compared to 30 normal healthy participants.
Liu 2014	Healthy participants in the control group. No separated analysis available
Liu 2017a	Assessment of the prediction of future HCC
Liu 2017b	Control group included healthy participants. No separate analysis was available.
Lu 2008	Participants without chronic liver disease were included. No separate analysis available
Luning 1991	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Lv 2013	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Mac Kinnon 1985	58 patients with suspected liver malignancy, 27 metastases only 5 HCC.
Maeda 2019	Control group included healthy participants. No separate analysis was available.
Mao 2010	HCC vs all other groups (also healthy participants). No separated analysis was available.
Matboli 2018	Control group included healthy participants. No separate analysis was available.
Matboli 2020	Control group included healthy participants. No separate analysis was available.

Study	Reason for exclusion
Matsubara 2013	Control group included healthy participants. No separate analysis was available.
Maussier 1990	Comparison only with a healthy control group
Mclntire 1972	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Mebazaa 1985	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Mehta 2018	More than 5% of recurrent HCC
Melia 1983	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Merchante 2019	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Mita 1998	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Morimoto 2002	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Morimoto 2012	Assessment of HCC recurrence
Morota 2011	HCC vs all other groups (including healthy controls). No separated analysis available
Nayak 1988	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Oka 1990	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Ola 2006	Patients with HCC compared to normal healthy participants.
Pan 2019	Control group included healthy participants. No separate analysis was available.
Peterson 2000	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Pocha 2013	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Pratedrat 2020	Control group included healthy participants. No separate analysis was available.
Qiao 2011	HCC vs all other groups (including healthy controls). No separated analysis available.
Qu 2011	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Quaia 2004	Only participants with solid focal hepatic lesions were included for differential diagnosis.
Rao 2003	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.

Study	Reason for exclusion
Rickes 2003	No data on US accuracy, only for CEUS
Rizzi 1994	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Rode 2001	Only per lesion analysis
Saber 2017	Study on surgical specimen. No data on the index tests
Sakai 1991	Review, no original data
Salmi 1988	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Sangiovanni 2004	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Santagostino 2003	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Sato 2009	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Seif 2019	Control group included healthy participants. No separate analysis was available.
Sekoguchi 1994	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Shalably 2019	Control group included healthy participants. No separate analysis was available.
Shao 2015	Only participants with suspected liver malignancy who underwent surgery were included for differential diagnosis.
Shapiro 1996	0nly per lesion analysis data provided
Shehab-Eldeen 2019	Control group included healthy participants. No separate analysis was available.
Sherman 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Sheu 1985	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Shinagawa 1984	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Singal 2017	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Tameda 2013	More than 5% of participants (55/96) was treated and were assessed for recurrent HCC.
Thakur 2014	Only participants with suspected liver malignancy were included.
Toyoda 2011	Participants only with AFP < 20 ng/mL

Study	Reason for exclusion
Tradati 1998	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Tung 2012	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Ueda 1995	Only per lesion analysis only
Uenishi 2006	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Wang 2020	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Wei 2012	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Welberry 2020	Control group included healthy participants. No separate analysis was available.
Worland 2018	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Xiao 2019	Control group included healthy participants. No separate analysis was available.
Xu 1990	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Xu 2012	Only per-lesion analysis
Xu 2014	Participants without chronic liver disease were included. No separate analysis available
Xu 2019	Comparison of participants with HCC vs participants with focal nodular hyperplasia and without chronic liver disease
Yamamoto 2009	No control group, only patients with HCC included
Yamamoto 2010	Assessment of treatment response
Yamashiki 2011	More than 5% (65/106) of patients with previously treated HCC
Yamashita 2020	Participants were included regardless of treatment history and with unspecified criteria for diagnosis.
Yang 2013b	More than 5% of HCC were recurrent HCC
Yang 2019	Control group included healthy participants. No separate analysis was available.
Yao 2013	The 2 by 2 table was not reported directly in the study and could not be calculated/extracted based on the data that were available.
Yasmin Anum 2009	The control group included healthy participants. No separate analysis available
Yasuda 2010	The 2 by 2 table was not reported directly in the study, and could not be calculated/extracted based on the data that were available.

Study	Reason for exclusion
Younis 2019	The 2 by 2 table was not reported directly in the study, and could not be calculated/extracted based on the data that were available.
Yvamoto 2015	The control group comprises healthy participants. No separate analysis available
Zhang 1999	The 2 by 2 table was not reported directly in the study, and could not be calculated/extracted based on the data that were available
Zhang 2010	HCC vs all others groups (including healthy controls). No separate analysis available
Zhao 2011	Mixed control group with healthy individuals. No separate analysis available
Zheng 2014	Mixed control group with healthy individuals. No separate analysis available
Zheng 2018	More than 5% patients with recurrent HCC
Zheng 2019	Control group included healthy participants. No separate analysis was available. 46/180 HCC were recurrences.

AFP: alpha-foetoprotein;**CENP-F:** centromere protein F; **CEUS:** contrast-enhanced ultrasound; **HCC:** hepatocellular carcinoma; **OLT:** orthotopic liver transplantation.

DATA

Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 Alpha-foetoprotein	326	144570
2 Ultrasound	39	18792
3 US + AFP	8	5454
4 AFP cut-off around 20 ng/mL	147	52144
5 AFP cut-off around 200 ng/mL	56	20452
6 US + AFP cut-off 20 ng/mL	6	5044
7 US for direct comparison AFP 20 ng/mL	11	6674

Test 1. Alpha-foetoprotein

Alpha-foetoprotein

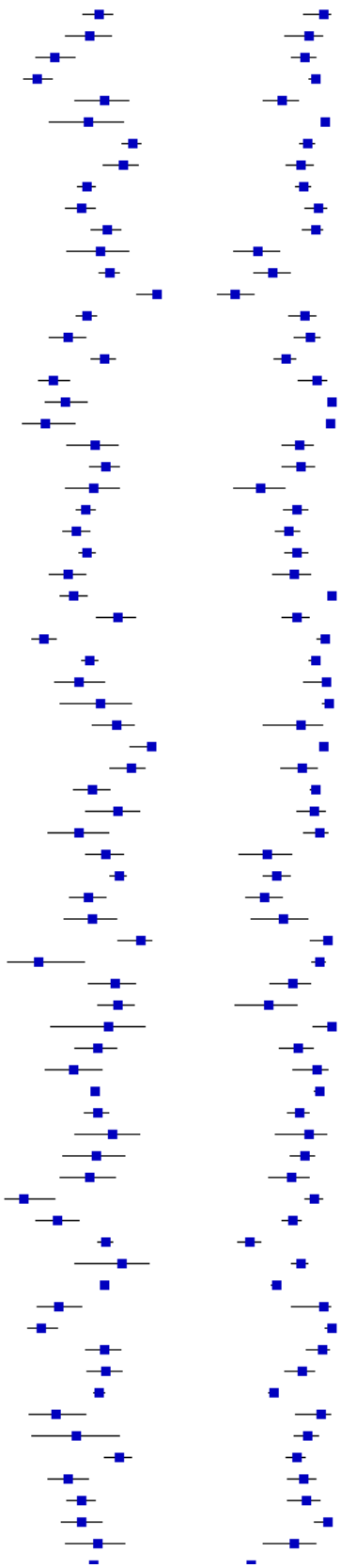
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdel-Aziz 2016	43	8	25	10	0.63 [0.51, 0.75]	0.56 [0.31, 0.78]
Abdelghany 2018	11	4	3	6	0.79 [0.49, 0.95]	0.60 [0.26, 0.88]
Abdel-Hamid 2014	13	1	7	39	0.65 [0.41, 0.85]	0.97 [0.87, 1.00]
Abdel-Razik 2016	68	28	14	52	0.83 [0.73, 0.90]	0.65 [0.54, 0.75]
Aboelfotoh 2018	40	0	0	40	1.00 [0.91, 1.00]	1.00 [0.91, 1.00]
Abu El Makarem 2011	84	0	29	120	0.74 [0.65, 0.82]	1.00 [0.97, 1.00]
Ahmed Mohamed 2016	18	4	12	16	0.60 [0.41, 0.77]	0.80 [0.56, 0.94]
Ahn 2016	225	76	141	290	0.61 [0.56, 0.66]	0.79 [0.75, 0.83]
Alexander 1978	29	20	6	43	0.83 [0.66, 0.93]	0.68 [0.55, 0.79]
Ali 2019	46	11	14	49	0.77 [0.64, 0.87]	0.82 [0.70, 0.90]
Almani 2004	72	11	28	89	0.72 [0.62, 0.81]	0.89 [0.81, 0.94]
Alpert 1971	59	0	58	337	0.50 [0.41, 0.60]	1.00 [0.99, 1.00]
Alsebaey 2016	31	5	12	17	0.72 [0.56, 0.85]	0.77 [0.55, 0.92]
Al-Zoubi 2017	14	0	12	27	0.54 [0.33, 0.73]	1.00 [0.87, 1.00]
Amuro 1988	22	1	30	41	0.42 [0.29, 0.57]	0.98 [0.87, 1.00]
Arrieta 2007	70	0	123	74	0.36 [0.29, 0.43]	1.00 [0.95, 1.00]
Arrigoni 1988	12	9	4	139	0.75 [0.48, 0.93]	0.94 [0.89, 0.97]
Atiq 2017	27	51	51	511	0.35 [0.24, 0.46]	0.91 [0.88, 0.93]
Attallah 2011	59	0	91	100	0.39 [0.31, 0.48]	1.00 [0.96, 1.00]
Attallah 2013	93	3	134	338	0.41 [0.35, 0.48]	0.99 [0.97, 1.00]
Attallah 2017	140	0	178	341	0.44 [0.38, 0.50]	1.00 [0.99, 1.00]
Attallah 2018	43	2	67	70	0.39 [0.30, 0.49]	0.97 [0.90, 1.00]
Attallah 2020	59	0	89	133	0.40 [0.32, 0.48]	1.00 [0.97, 1.00]
Bachtiar 2009	34	5	31	49	0.52 [0.40, 0.65]	0.91 [0.80, 0.97]
Badr 2014	25	2	5	28	0.83 [0.65, 0.94]	0.93 [0.78, 0.99]
Baek 2009	130	23	107	77	0.55 [0.48, 0.61]	0.77 [0.68, 0.85]
Bell 1982	12	10	2	100	0.86 [0.57, 0.98]	0.91 [0.84, 0.96]
Beneduce 2004	26	9	34	41	0.43 [0.31, 0.57]	0.82 [0.69, 0.91]
Beneduce 2008	16	4	17	27	0.48 [0.31, 0.66]	0.87 [0.70, 0.96]
Bessa 2010	18	3	12	27	0.60 [0.41, 0.77]	0.90 [0.73, 0.98]
Best 2016	166	24	119	378	0.58 [0.52, 0.64]	0.94 [0.91, 0.96]
Best 2020	71	3	54	228	0.57 [0.48, 0.66]	0.99 [0.96, 1.00]
Biselli 2015	77	43	39	225	0.66 [0.57, 0.75]	0.84 [0.79, 0.88]
Bolondi 2001	25	46	36	206	0.41 [0.29, 0.54]	0.82 [0.76, 0.86]
Bon 1998	25	13	12	10	0.68 [0.50, 0.82]	0.43 [0.23, 0.66]
Brunello 1993	17	1	22	15	0.44 [0.28, 0.60]	0.94 [0.70, 1.00]
Buffet 1988	13	0	10	184	0.57 [0.34, 0.77]	1.00 [0.98, 1.00]
Cabrera 2012	77	6	66	40	0.54 [0.45, 0.62]	0.87 [0.74, 0.95]
Capurro 2003	20	1	14	19	0.59 [0.41, 0.75]	0.95 [0.75, 1.00]
Caviglia 2016	45	6	10	38	0.82 [0.69, 0.91]	0.86 [0.73, 0.95]
Caviglia 2017	20	4	13	26	0.61 [0.42, 0.77]	0.87 [0.69, 0.96]
Cedrone 2000	39	34	35	242	0.53 [0.41, 0.64]	0.88 [0.83, 0.91]
Chalasani 1999	14	34	8	229	0.64 [0.41, 0.83]	0.87 [0.82, 0.91]
Chan 2013	46	289	11	1185	0.81 [0.68, 0.90]	0.80 [0.78, 0.82]
Chan 2014	460	72	98	171	0.82 [0.79, 0.86]	0.70 [0.64, 0.76]
Chang 1988	26	1	10	26	0.72 [0.55, 0.86]	0.96 [0.81, 1.00]
Chang 2015	250	212	113	1022	0.69 [0.64, 0.74]	0.83 [0.81, 0.85]
Chayvialle 1977	12	3	6	179	0.67 [0.41, 0.87]	0.98 [0.95, 1.00]
Chen 1977	115	25	10	49	0.92 [0.86, 0.96]	0.66 [0.54, 0.77]
Chen 2003	142	466	115	2989	0.55 [0.49, 0.61]	0.87 [0.85, 0.88]
Chen 2015	77	13	26	82	0.75 [0.65, 0.83]	0.86 [0.78, 0.93]
Chen 2018	99	70	103	371	0.49 [0.42, 0.56]	0.84 [0.80, 0.87]
Cheng 2012	71	229	60	1362	0.54 [0.45, 0.63]	0.86 [0.84, 0.87]
Chimparlee 2015	105	4	52	73	0.67 [0.59, 0.74]	0.95 [0.87, 0.99]
Choi 2019	26	22	16	146	0.62 [0.46, 0.76]	0.87 [0.81, 0.92]
Chuaypen 2018	97	14	53	136	0.65 [0.56, 0.72]	0.91 [0.85, 0.95]
Cottone 1988	10	39	5	104	0.67 [0.38, 0.88]	0.73 [0.65, 0.80]
Cui 2002	34	11	26	19	0.57 [0.43, 0.69]	0.63 [0.44, 0.80]
Cui 2003	70	33	50	57	0.58 [0.49, 0.67]	0.63 [0.53, 0.73]
da Costa 2015a	49	0	10	49	0.83 [0.71, 0.92]	1.00 [0.93, 1.00]
da Costa 2015b	38	7	12	43	0.76 [0.62, 0.87]	0.86 [0.73, 0.94]
da Costa 2015c	37	10	38	65	0.49 [0.38, 0.61]	0.87 [0.77, 0.93]
da Costa 2015d	5	18	16	76	0.24 [0.08, 0.47]	0.81 [0.71, 0.88]
Ding 2020	413	131	60	379	0.87 [0.84, 0.90]	0.74 [0.70, 0.78]
Dong 2015	118	62	72	52	0.62 [0.55, 0.69]	0.46 [0.36, 0.55]
Durazo 2008	99	13	45	83	0.69 [0.61, 0.76]	0.86 [0.78, 0.93]
Edis 1998	13	4	26	67	0.33 [0.19, 0.50]	0.94 [0.86, 0.98]
Edoo 2019	680	46	395	191	0.63 [0.60, 0.66]	0.81 [0.75, 0.85]
Fissa 2013	26	0	24	30	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]

Test 1. (Continued)

Edis 1998	13	4	26	67	0.33 [0.19, 0.50]	0.94 [0.86, 0.98]
Edoo 2019	680	46	395	191	0.63 [0.60, 0.66]	0.81 [0.75, 0.85]
Eissa 2013	26	0	24	30	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
El-Abd 2015	28	0	12	40	0.70 [0.53, 0.83]	1.00 [0.91, 1.00]
El-Abd 2016	41	4	9	26	0.82 [0.69, 0.91]	0.87 [0.69, 0.96]
El Gawad 2014	36	4	4	6	0.90 [0.76, 0.97]	0.60 [0.26, 0.88]
el-Houseini 2005	14	5	30	15	0.32 [0.19, 0.48]	0.75 [0.51, 0.91]
El Mahdy 2019	49	25	11	50	0.82 [0.70, 0.90]	0.67 [0.55, 0.77]
El Moety 2011	26	0	24	30	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
Elnemr 2012	45	6	15	54	0.75 [0.62, 0.85]	0.90 [0.79, 0.96]
El-Serag 2017	5	32	16	511	0.24 [0.08, 0.47]	0.94 [0.92, 0.96]
El Shafie 2012	24	14	7	21	0.77 [0.59, 0.90]	0.60 [0.42, 0.76]
El-Shenawy 2012	9	48	5	41	0.64 [0.35, 0.87]	0.46 [0.35, 0.57]
El-Sherif 2012	19	3	11	27	0.63 [0.44, 0.80]	0.90 [0.73, 0.98]
Eltaher 2016	25	6	5	24	0.83 [0.65, 0.94]	0.80 [0.61, 0.92]
El-Tayeh 2012	20	1	17	27	0.54 [0.37, 0.71]	0.96 [0.82, 1.00]
El Zefzafy 2015	14	1	16	29	0.47 [0.28, 0.66]	0.97 [0.83, 1.00]
Erdal 2016	31	11	9	43	0.78 [0.62, 0.89]	0.80 [0.66, 0.89]
Ertle 2013	90	23	74	399	0.55 [0.47, 0.63]	0.95 [0.92, 0.97]
Ette 2015	39	4	23	53	0.63 [0.50, 0.75]	0.93 [0.83, 0.98]
Ezzikouri 2015	85	2	74	72	0.53 [0.45, 0.61]	0.97 [0.91, 1.00]
Fabris 1991	23	38	4	173	0.85 [0.66, 0.96]	0.82 [0.76, 0.87]
Fang 2010	110	16	35	112	0.76 [0.68, 0.83]	0.88 [0.80, 0.93]
Farid 2014	13	1	7	9	0.65 [0.41, 0.85]	0.90 [0.55, 1.00]
Feng 2016	232	26	97	345	0.71 [0.65, 0.75]	0.93 [0.90, 0.95]
Fujii1995	40	6	10	44	0.80 [0.66, 0.90]	0.88 [0.76, 0.95]
Gad 2005	95	52	15	182	0.86 [0.79, 0.92]	0.78 [0.72, 0.83]
Gambarin-Gelwan 2000	11	8	8	79	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]
Gani 2015	43	9	16	38	0.73 [0.60, 0.84]	0.81 [0.67, 0.91]
Ge 2015	64	25	25	76	0.72 [0.61, 0.81]	0.75 [0.66, 0.83]
Gentile 2017	33	24	23	80	0.59 [0.45, 0.72]	0.77 [0.68, 0.85]
Giannelli 2005	54	13	66	77	0.45 [0.36, 0.54]	0.86 [0.77, 0.92]
Giannelli 2007	203	29	296	433	0.41 [0.36, 0.45]	0.94 [0.91, 0.96]
Gopal 2014	317	69	135	607	0.70 [0.66, 0.74]	0.90 [0.87, 0.92]
Grazi 1995	61	3	50	113	0.55 [0.45, 0.64]	0.97 [0.93, 0.99]
Guan 2020	154	12	148	152	0.51 [0.45, 0.57]	0.93 [0.88, 0.96]
Hallager 2018	63	71	41	469	0.61 [0.51, 0.70]	0.87 [0.84, 0.90]
Han 2014	93	46	67	42	0.58 [0.50, 0.66]	0.48 [0.37, 0.59]
Han 2018	58	19	26	84	0.69 [0.58, 0.79]	0.82 [0.73, 0.89]
Hu 2018	213	26	156	150	0.58 [0.53, 0.63]	0.85 [0.79, 0.90]
Hu 2019	422	143	7	434	0.98 [0.97, 0.99]	0.75 [0.71, 0.79]
Huo 2007	146	52	102	835	0.59 [0.52, 0.65]	0.94 [0.92, 0.96]
Ibrahim 2013	56	12	24	23	0.70 [0.59, 0.80]	0.66 [0.48, 0.81]
Iizuka 2010a	62	8	46	48	0.57 [0.48, 0.67]	0.86 [0.74, 0.94]
Iizuka 2010b	68	41	44	105	0.61 [0.51, 0.70]	0.72 [0.64, 0.79]
Ishii 2000	18	153	11	552	0.62 [0.42, 0.79]	0.78 [0.75, 0.81]
Ismail 2017a	45	12	21	87	0.68 [0.56, 0.79]	0.88 [0.80, 0.94]
Ismail 2017b	17	1	13	29	0.57 [0.37, 0.75]	0.97 [0.83, 1.00]
Iyer 2018	165	4	38	115	0.81 [0.75, 0.86]	0.97 [0.92, 0.99]
Izzo 1999	79	19	20	340	0.80 [0.71, 0.87]	0.95 [0.92, 0.97]
Jang 2016	129	19	79	174	0.62 [0.55, 0.69]	0.90 [0.85, 0.94]
Jeon 2016	86	14	71	142	0.55 [0.47, 0.63]	0.91 [0.85, 0.95]
Ji 2016	124	30	76	67	0.62 [0.55, 0.69]	0.69 [0.59, 0.78]
Jiao 2018	92	2	88	158	0.51 [0.44, 0.59]	0.99 [0.96, 1.00]
Johnson 1978	29	1	1	99	0.97 [0.83, 1.00]	0.99 [0.95, 1.00]
Kanmura 2007	12	11	17	22	0.41 [0.24, 0.61]	0.67 [0.48, 0.82]
Khairy 2015	37	9	10	8	0.79 [0.64, 0.89]	0.47 [0.23, 0.72]
Kim 2006a	34	0	28	60	0.55 [0.42, 0.68]	1.00 [0.94, 1.00]
Kim 2006b	48	52	7	10	0.87 [0.76, 0.95]	0.16 [0.08, 0.28]
Kim 2006c	26	26	16	159	0.62 [0.46, 0.76]	0.86 [0.80, 0.91]
Kim 2012	154	57	42	297	0.79 [0.72, 0.84]	0.84 [0.80, 0.88]
Kim 2014	20	7	10	28	0.67 [0.47, 0.83]	0.80 [0.63, 0.92]
Kim 2016	625	392	345	712	0.64 [0.61, 0.67]	0.64 [0.62, 0.67]
Kim 2018	36	4	18	22	0.67 [0.53, 0.79]	0.85 [0.65, 0.96]
Kim 2019	45	18	8	29	0.85 [0.72, 0.93]	0.62 [0.46, 0.75]
Kim 2019a	61	26	21	54	0.74 [0.64, 0.83]	0.68 [0.56, 0.78]
Kim 2019b	32	7	32	321	0.50 [0.37, 0.63]	0.98 [0.96, 0.99]
Krygier 2011	19	26	10	63	0.66 [0.46, 0.82]	0.71 [0.60, 0.80]
Kumada 2014	43	10	61	94	0.41 [0.32, 0.51]	0.90 [0.83, 0.95]
Lee 2004	34	91	20	108	0.63 [0.49, 0.76]	0.54 [0.47, 0.61]
Lee 2014	77	2	43	38	0.64 [0.55, 0.73]	0.95 [0.83, 0.99]
Li 2016a	31	6	22	36	0.58 [0.44, 0.72]	0.86 [0.71, 0.95]
Li 2016b	24	17	41	88	0.37 [0.25, 0.50]	0.84 [0.75, 0.90]

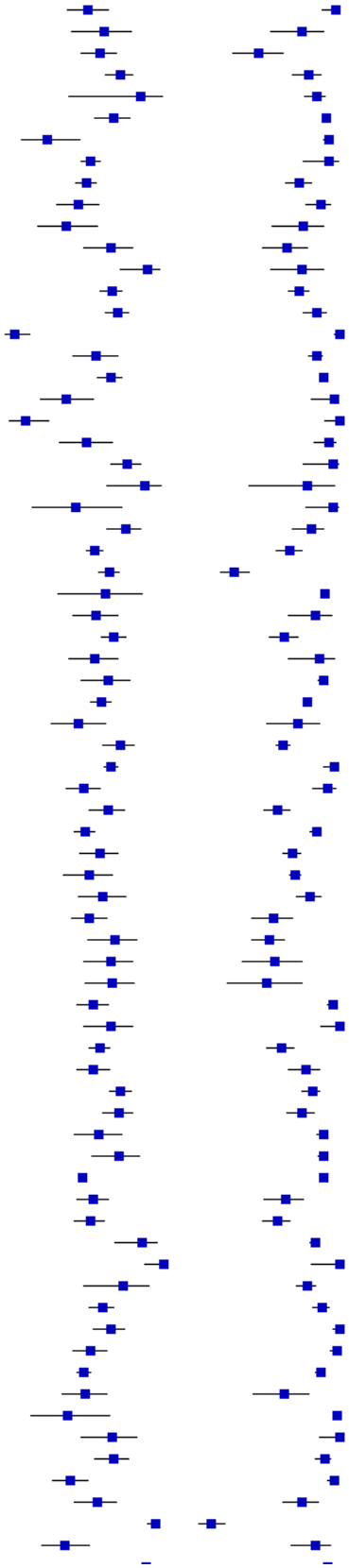
Test 1. (Continued)

Lee 2014	77	2	43	38	0.64 [0.55, 0.73]	0.95 [0.83, 0.99]
Li 2016a	31	6	22	36	0.58 [0.44, 0.72]	0.86 [0.71, 0.95]
Li 2016b	24	17	41	88	0.37 [0.25, 0.50]	0.84 [0.75, 0.90]
Li 2016c	27	30	77	274	0.26 [0.18, 0.35]	0.90 [0.86, 0.93]
Li 2017a	23	23	11	52	0.68 [0.49, 0.83]	0.69 [0.58, 0.79]
Li 2017b	11	10	8	271	0.58 [0.33, 0.80]	0.96 [0.94, 0.98]
Li 2019a	144	35	25	207	0.85 [0.79, 0.90]	0.86 [0.80, 0.90]
Liao 2012	47	18	12	78	0.80 [0.67, 0.89]	0.81 [0.72, 0.88]
Lim 2015	205	48	156	228	0.57 [0.52, 0.62]	0.83 [0.78, 0.87]
Lin 2000	65	6	57	70	0.53 [0.44, 0.62]	0.92 [0.84, 0.97]
Lin 2015	75	10	33	88	0.69 [0.60, 0.78]	0.90 [0.82, 0.95]
Lin 2016	17	23	9	27	0.65 [0.44, 0.83]	0.54 [0.39, 0.68]
Liu 2007	161	29	66	51	0.71 [0.65, 0.77]	0.64 [0.52, 0.74]
Liu 2010a	28	47	0	32	1.00 [0.88, 1.00]	0.41 [0.30, 0.52]
Liu 2017	136	16	104	79	0.57 [0.50, 0.63]	0.83 [0.74, 0.90]
Liu 2018	36	11	44	71	0.45 [0.34, 0.57]	0.87 [0.77, 0.93]
Liu 2019	112	52	54	133	0.67 [0.60, 0.75]	0.72 [0.65, 0.78]
Liu 2020	38	5	67	49	0.36 [0.27, 0.46]	0.91 [0.80, 0.97]
Loglio 2018	28	0	36	148	0.44 [0.31, 0.57]	1.00 [0.98, 1.00]
Loglio 2019	11	1	24	222	0.31 [0.17, 0.49]	1.00 [0.98, 1.00]
Lok 2010	24	15	15	62	0.62 [0.45, 0.77]	0.81 [0.70, 0.89]
Long 2011	76	13	35	56	0.68 [0.59, 0.77]	0.81 [0.70, 0.90]
Luo 2018a	22	18	14	23	0.61 [0.43, 0.77]	0.56 [0.40, 0.72]
Luo 2018b	183	27	142	99	0.56 [0.51, 0.62]	0.79 [0.70, 0.85]
Luo 2018c	78	38	77	105	0.50 [0.42, 0.58]	0.73 [0.65, 0.80]
Ma 2018	210	32	158	120	0.57 [0.52, 0.62]	0.79 [0.72, 0.85]
Mao 2017	37	13	45	44	0.45 [0.34, 0.57]	0.77 [0.64, 0.87]
Maringhini 1988	71	0	75	217	0.49 [0.40, 0.57]	1.00 [0.98, 1.00]
Marrero 2003	42	22	13	82	0.76 [0.63, 0.87]	0.79 [0.70, 0.86]
Marrero 2005	43	6	101	146	0.30 [0.23, 0.38]	0.96 [0.92, 0.99]
Marrero 2009	247	42	172	375	0.59 [0.54, 0.64]	0.90 [0.87, 0.93]
Mashaly 2018	23	1	21	30	0.52 [0.37, 0.68]	0.97 [0.83, 1.00]
Matievskaya 2003	13	2	7	137	0.65 [0.41, 0.85]	0.99 [0.95, 1.00]
Matsuda 2008	36	4	12	17	0.75 [0.60, 0.86]	0.81 [0.58, 0.95]
McMahon 2000	31	70	1	1385	0.97 [0.84, 1.00]	0.95 [0.94, 0.96]
Mehinovic 2018	42	9	8	41	0.84 [0.71, 0.93]	0.82 [0.69, 0.91]
Min 2014	46	40	30	384	0.61 [0.49, 0.72]	0.91 [0.87, 0.93]
Minami 2015a	22	6	7	52	0.76 [0.56, 0.90]	0.90 [0.79, 0.96]
Minami 2015b	15	4	14	54	0.52 [0.33, 0.71]	0.93 [0.83, 0.98]
Miura 2007	44	16	20	24	0.69 [0.56, 0.80]	0.60 [0.43, 0.75]
Miura 2010	232	45	71	89	0.77 [0.71, 0.81]	0.66 [0.58, 0.74]
Mohamed 2020a	46	33	34	47	0.57 [0.46, 0.68]	0.59 [0.47, 0.70]
Mohamed 2020b	24	9	16	21	0.60 [0.43, 0.75]	0.70 [0.51, 0.85]
Montaser 2012	36	1	4	40	0.90 [0.76, 0.97]	0.98 [0.87, 1.00]
Moriya 2013	4	13	11	170	0.27 [0.08, 0.55]	0.93 [0.88, 0.96]
Moriyama 2000	29	12	10	38	0.74 [0.58, 0.87]	0.76 [0.62, 0.87]
Mukozu 2013	45	11	14	17	0.76 [0.63, 0.86]	0.61 [0.41, 0.78]
Mustika 2019	7	0	3	31	0.70 [0.35, 0.93]	1.00 [0.89, 1.00]
Na 2013	36	13	21	51	0.63 [0.49, 0.76]	0.80 [0.68, 0.89]
Nabih 2014	17	3	18	31	0.49 [0.31, 0.66]	0.91 [0.76, 0.98]
Nakamura 2006	844	24	517	324	0.62 [0.59, 0.65]	0.93 [0.90, 0.96]
Nguyen 2002	103	30	60	119	0.63 [0.55, 0.71]	0.80 [0.73, 0.86]
Nomair 2019	16	3	6	19	0.73 [0.50, 0.89]	0.86 [0.65, 0.97]
Nomura 1996	17	17	10	84	0.63 [0.42, 0.81]	0.83 [0.74, 0.90]
Nomura 1999	21	12	15	37	0.58 [0.41, 0.74]	0.76 [0.61, 0.87]
Nomura 2012	5	14	23	123	0.18 [0.06, 0.37]	0.90 [0.83, 0.94]
Oka 1994	21	48	34	150	0.38 [0.25, 0.52]	0.76 [0.69, 0.82]
Oka 2001	266	108	122	104	0.69 [0.64, 0.73]	0.49 [0.42, 0.56]
Okazaki 1984	11	44	3	187	0.79 [0.49, 0.95]	0.81 [0.75, 0.86]
Omar 2017	878	364	413	708	0.68 [0.65, 0.71]	0.66 [0.63, 0.69]
Omran 2016	21	1	32	19	0.40 [0.26, 0.54]	0.95 [0.75, 1.00]
Omran 2020	30	0	74	92	0.29 [0.20, 0.39]	1.00 [0.96, 1.00]
Ozkan 2011	51	3	24	52	0.68 [0.56, 0.78]	0.95 [0.85, 0.99]
Park 2017a	54	14	25	63	0.68 [0.57, 0.78]	0.82 [0.71, 0.90]
Park 2017b	626	392	344	712	0.65 [0.61, 0.68]	0.64 [0.62, 0.67]
Passos-Castilho 2015	12	2	20	28	0.38 [0.21, 0.56]	0.93 [0.78, 0.99]
Pateron 1994	7	15	7	89	0.50 [0.23, 0.77]	0.86 [0.77, 0.92]
Paul 2007	78	42	23	152	0.77 [0.68, 0.85]	0.78 [0.72, 0.84]
Piciocchi 2013	30	13	36	63	0.45 [0.33, 0.58]	0.83 [0.73, 0.91]
Pompili 2003	70	9	61	50	0.53 [0.45, 0.62]	0.85 [0.73, 0.93]
Poon 2001	35	1	30	50	0.54 [0.41, 0.66]	0.98 [0.90, 1.00]
Porta 2008	19	7	11	23	0.63 [0.44, 0.80]	0.77 [0.58, 0.90]



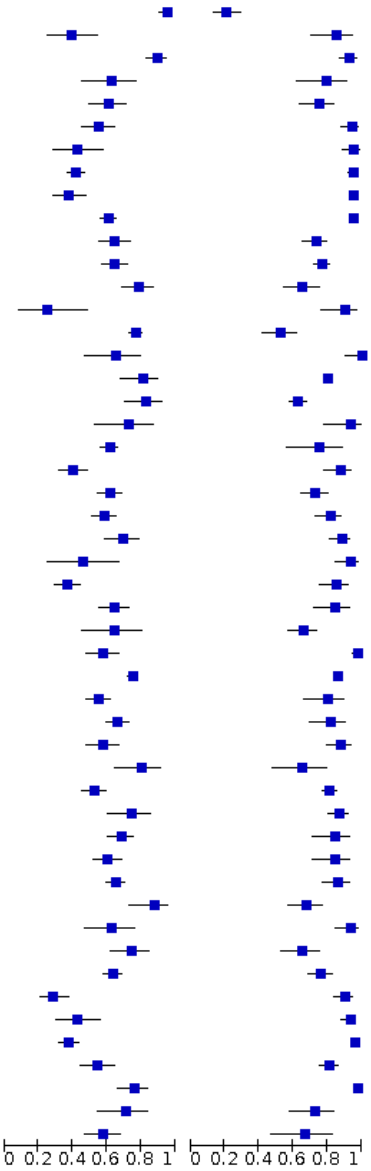
Test 1. (Continued)

Poon 2001	35	1	30	50	0.54 [0.41, 0.66]	0.98 [0.90, 1.00]
Porta 2008	19	7	11	23	0.63 [0.44, 0.80]	0.77 [0.58, 0.90]
Pote 2015	52	22	33	22	0.61 [0.50, 0.72]	0.50 [0.35, 0.65]
Qi 2020	88	17	32	72	0.73 [0.64, 0.81]	0.81 [0.71, 0.88]
Raedle 1995	6	20	1	120	0.86 [0.42, 1.00]	0.86 [0.79, 0.91]
Raedle 1998	52	52	23	584	0.69 [0.58, 0.79]	0.92 [0.89, 0.94]
Raff 2014	8	20	20	308	0.29 [0.13, 0.49]	0.94 [0.91, 0.96]
Reichl 2015	171	2	138	28	0.55 [0.50, 0.61]	0.93 [0.78, 0.99]
Ricco 2018	135	32	123	98	0.52 [0.46, 0.59]	0.75 [0.67, 0.83]
Sadeghi 2015	29	9	32	69	0.48 [0.35, 0.61]	0.88 [0.79, 0.95]
Sadik 2019	12	7	18	24	0.40 [0.23, 0.59]	0.77 [0.59, 0.90]
Saitta 2017	27	16	13	34	0.68 [0.51, 0.81]	0.68 [0.53, 0.80]
Salem 2013	27	7	3	23	0.90 [0.73, 0.98]	0.77 [0.58, 0.90]
Sanai 2010	141	49	65	150	0.68 [0.62, 0.75]	0.75 [0.69, 0.81]
Sarwar 2014	125	14	48	88	0.72 [0.65, 0.79]	0.86 [0.78, 0.92]
Sassa 1999	5	0	56	134	0.08 [0.03, 0.18]	1.00 [0.97, 1.00]
Sato 1993	33	43	23	262	0.59 [0.45, 0.72]	0.86 [0.81, 0.90]
Seo 2015	106	107	51	991	0.68 [0.60, 0.75]	0.90 [0.88, 0.92]
Shaheen 2015	16	1	24	29	0.40 [0.25, 0.57]	0.97 [0.83, 1.00]
Shaheen 2018	6	0	34	40	0.15 [0.06, 0.30]	1.00 [0.91, 1.00]
Shang 2012a	21	5	19	68	0.53 [0.36, 0.68]	0.93 [0.85, 0.98]
Shang 2012b	71	1	20	22	0.78 [0.68, 0.86]	0.96 [0.78, 1.00]
Shariff 2010	16	2	2	8	0.89 [0.65, 0.99]	0.80 [0.44, 0.97]
Shariff 2016	6	1	7	24	0.46 [0.19, 0.75]	0.96 [0.80, 1.00]
Sharma 2010	54	12	16	56	0.77 [0.66, 0.86]	0.82 [0.71, 0.91]
Shen 2012a	245	42	179	95	0.58 [0.53, 0.63]	0.69 [0.61, 0.77]
Shen 2012b	140	82	69	45	0.67 [0.60, 0.73]	0.35 [0.27, 0.44]
Sherman 1995	9	91	5	964	0.64 [0.35, 0.87]	0.91 [0.90, 0.93]
Shimizu 2002	33	5	23	29	0.59 [0.45, 0.72]	0.85 [0.69, 0.95]
Shu 2010	113	44	49	86	0.70 [0.62, 0.77]	0.66 [0.57, 0.74]
Simão 2015	26	3	19	22	0.58 [0.42, 0.72]	0.88 [0.69, 0.97]
Singal 2012	27	38	14	363	0.66 [0.49, 0.80]	0.91 [0.87, 0.93]
Snowberger 2007	148	339	91	1357	0.62 [0.55, 0.68]	0.80 [0.78, 0.82]
Song 2002	18	8	20	23	0.47 [0.31, 0.64]	0.74 [0.55, 0.88]
Song 2011	64	163	23	311	0.74 [0.63, 0.82]	0.66 [0.61, 0.70]
Song 2014	374	3	176	82	0.68 [0.64, 0.72]	0.96 [0.90, 0.99]
Song 2020a	51	5	49	62	0.51 [0.41, 0.61]	0.93 [0.83, 0.98]
Song 2020b	53	61	27	99	0.66 [0.55, 0.76]	0.62 [0.54, 0.69]
Soroida 2012	131	64	121	391	0.52 [0.46, 0.58]	0.86 [0.82, 0.89]
Sterling 2009	45	86	29	212	0.61 [0.49, 0.72]	0.71 [0.66, 0.76]
Sterling 2012	25	219	21	590	0.54 [0.39, 0.69]	0.73 [0.70, 0.76]
Sultanik 2017	29	21	17	95	0.63 [0.48, 0.77]	0.82 [0.74, 0.88]
Sun 2010	48	26	40	38	0.55 [0.44, 0.65]	0.59 [0.46, 0.71]
Sun 2020	28	46	12	60	0.70 [0.53, 0.83]	0.57 [0.47, 0.66]
Tahon 2019	27	12	13	18	0.68 [0.51, 0.81]	0.60 [0.41, 0.77]
Takaya 2019	28	9	13	11	0.68 [0.52, 0.82]	0.55 [0.32, 0.77]
Takikawa 1992	66	10	50	243	0.57 [0.47, 0.66]	0.96 [0.93, 0.98]
Talkahn 2018	27	0	13	30	0.68 [0.51, 0.81]	1.00 [0.88, 1.00]
Tan 2012	160	54	102	96	0.61 [0.55, 0.67]	0.64 [0.56, 0.72]
Tan 2014	59	16	44	62	0.57 [0.47, 0.67]	0.79 [0.69, 0.88]
Tang 2017a	130	32	46	158	0.74 [0.67, 0.80]	0.83 [0.77, 0.88]
Tanglijvanich 2010	73	23	27	77	0.73 [0.63, 0.81]	0.77 [0.68, 0.85]
Tayob 2016a	29	36	19	325	0.60 [0.45, 0.74]	0.90 [0.86, 0.93]
Tayob 2016b	29	58	11	523	0.72 [0.56, 0.85]	0.90 [0.87, 0.92]
Tayob 2019	1744	3363	1712	30264	0.50 [0.49, 0.52]	0.90 [0.90, 0.90]
Teng 2016	63	22	48	44	0.57 [0.47, 0.66]	0.67 [0.54, 0.78]
Tian 2017	66	56	54	90	0.55 [0.46, 0.64]	0.62 [0.53, 0.70]
Tong 2001	27	86	4	485	0.87 [0.70, 0.96]	0.85 [0.82, 0.88]
Toraih 2018	30	0	0	20	1.00 [0.88, 1.00]	1.00 [0.83, 1.00]
Tremolada 1989	15	39	5	155	0.75 [0.51, 0.91]	0.80 [0.74, 0.85]
Trevisani 2001	106	18	64	152	0.62 [0.55, 0.70]	0.89 [0.84, 0.94]
Tsai 1995	68	0	33	101	0.67 [0.57, 0.76]	1.00 [0.96, 1.00]
Tsai 1997	52	1	42	93	0.55 [0.45, 0.66]	0.99 [0.94, 1.00]
Tsai 2017	253	58	240	435	0.51 [0.47, 0.56]	0.88 [0.85, 0.91]
Tsuda 2004	29	11	27	21	0.52 [0.38, 0.65]	0.66 [0.47, 0.81]
Ungtrakul 2016	7	29	10	2247	0.41 [0.18, 0.67]	0.99 [0.98, 0.99]
Unic 2013	22	0	10	28	0.69 [0.50, 0.84]	1.00 [0.88, 1.00]
Volk 2007	58	15	26	154	0.69 [0.58, 0.79]	0.91 [0.86, 0.95]
Vongsuvanh 2016	37	6	49	166	0.43 [0.32, 0.54]	0.97 [0.93, 0.99]
Wang 2005	36	15	25	51	0.59 [0.46, 0.71]	0.77 [0.65, 0.87]
Wang 2009	156	89	8	24	0.95 [0.91, 0.98]	0.21 [0.14, 0.30]
Wang 2013a	19	6	29	34	0.40 [0.26, 0.55]	0.85 [0.70, 0.94]



Test 1. (Continued)

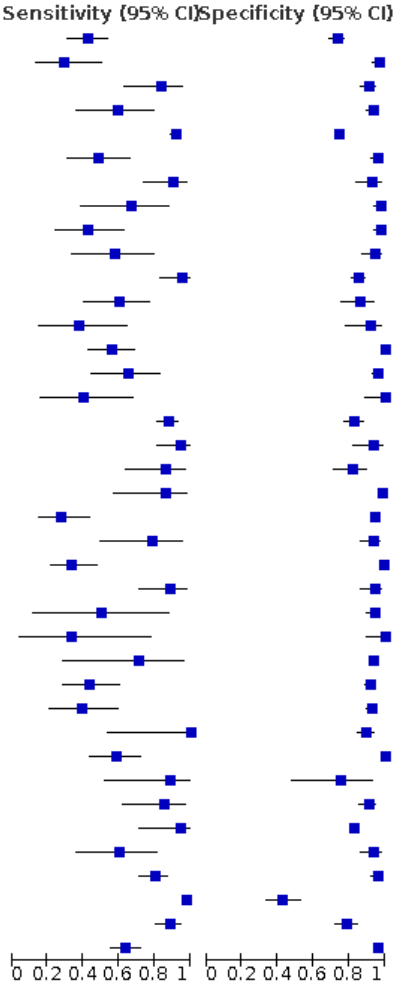
Wang 2009	156	89	8	24	0.95 [0.91, 0.98]	0.21 [0.14, 0.30]
Wang 2013a	19	6	29	34	0.40 [0.26, 0.55]	0.85 [0.70, 0.94]
Wang 2013b	113	8	13	107	0.90 [0.83, 0.94]	0.93 [0.87, 0.97]
Wang 2014a	25	7	15	27	0.63 [0.46, 0.77]	0.79 [0.62, 0.91]
Wang 2014b	51	20	33	60	0.61 [0.49, 0.71]	0.75 [0.64, 0.84]
Wang 2016a	64	5	52	88	0.55 [0.46, 0.64]	0.95 [0.88, 0.98]
Wang 2016b	21	5	28	97	0.43 [0.29, 0.58]	0.95 [0.89, 0.98]
Wang 2016c	181	22	251	416	0.42 [0.37, 0.47]	0.95 [0.92, 0.97]
Wang 2016d	43	29	70	557	0.38 [0.29, 0.48]	0.95 [0.93, 0.97]
Wang 2016e	259	40	166	764	0.61 [0.56, 0.66]	0.95 [0.93, 0.96]
Wang 2017	73	43	40	118	0.65 [0.55, 0.73]	0.73 [0.66, 0.80]
Wang 2019a	114	82	62	277	0.65 [0.57, 0.72]	0.77 [0.72, 0.81]
Wang 2019b	71	31	19	59	0.79 [0.69, 0.87]	0.66 [0.55, 0.75]
Weiss 2019	5	4	15	36	0.25 [0.09, 0.49]	0.90 [0.76, 0.97]
Wong 2008	363	51	109	56	0.77 [0.73, 0.81]	0.52 [0.42, 0.62]
Wong 2009	24	0	13	37	0.65 [0.47, 0.80]	1.00 [0.91, 1.00]
Wong 2014a	46	289	11	1185	0.81 [0.68, 0.90]	0.80 [0.78, 0.82]
Wong 2014b	44	138	9	233	0.83 [0.70, 0.92]	0.63 [0.58, 0.68]
Wu 2009	21	2	8	28	0.72 [0.53, 0.87]	0.93 [0.78, 0.99]
Wu 2017	225	8	141	24	0.61 [0.56, 0.66]	0.75 [0.57, 0.89]
Wu 2018	57	10	86	70	0.40 [0.32, 0.48]	0.88 [0.78, 0.94]
Wu 2020	122	34	76	92	0.62 [0.54, 0.68]	0.73 [0.64, 0.81]
Xing 2019	109	21	78	93	0.58 [0.51, 0.65]	0.82 [0.73, 0.88]
Xu 2018	61	16	27	119	0.69 [0.59, 0.79]	0.88 [0.81, 0.93]
Yan 2018	11	4	13	58	0.46 [0.26, 0.67]	0.94 [0.84, 0.98]
Yang 2013a	66	12	113	68	0.37 [0.30, 0.44]	0.85 [0.75, 0.92]
Yang 2014	79	9	44	48	0.64 [0.55, 0.73]	0.84 [0.72, 0.93]
Yang 2017	20	48	11	92	0.65 [0.45, 0.81]	0.66 [0.57, 0.74]
Yang 2019	64	4	47	176	0.58 [0.48, 0.67]	0.98 [0.94, 0.99]
Yao 2016	597	144	199	905	0.75 [0.72, 0.78]	0.86 [0.84, 0.88]
Ye 2019a	115	10	94	40	0.55 [0.48, 0.62]	0.80 [0.66, 0.90]
Ye 2019b	135	11	69	49	0.66 [0.59, 0.73]	0.82 [0.70, 0.90]
Yoon 2009	61	12	45	88	0.58 [0.48, 0.67]	0.88 [0.80, 0.94]
Youns 2013	32	14	8	26	0.80 [0.64, 0.91]	0.65 [0.48, 0.79]
Yu 2011	103	60	92	263	0.53 [0.46, 0.60]	0.81 [0.77, 0.86]
Yu 2016	38	18	13	120	0.75 [0.60, 0.86]	0.87 [0.80, 0.92]
Yu 2020a	108	8	50	42	0.68 [0.60, 0.76]	0.84 [0.71, 0.93]
Yu 2020b	92	8	60	42	0.61 [0.52, 0.68]	0.84 [0.71, 0.93]
Yu 2020c	189	11	101	69	0.65 [0.59, 0.71]	0.86 [0.77, 0.93]
Zekri 2013	35	29	5	61	0.88 [0.73, 0.96]	0.68 [0.57, 0.77]
Zhan 2020	30	4	18	61	0.63 [0.47, 0.76]	0.94 [0.85, 0.98]
Zhang 2020	47	26	16	48	0.75 [0.62, 0.85]	0.65 [0.53, 0.76]
Zheng 2017	211	39	122	125	0.63 [0.58, 0.69]	0.76 [0.69, 0.83]
Zhou 2012	34	14	84	127	0.29 [0.21, 0.38]	0.90 [0.84, 0.94]
Zhou 2019	27	11	36	152	0.43 [0.30, 0.56]	0.93 [0.88, 0.97]
Zhu 2013	95	18	157	437	0.38 [0.32, 0.44]	0.96 [0.94, 0.98]
Zhu 2020	55	42	46	179	0.54 [0.44, 0.64]	0.81 [0.75, 0.86]
Ziada 2016	78	11	25	400	0.76 [0.66, 0.84]	0.97 [0.95, 0.99]
Zinkin 2008	29	14	12	37	0.71 [0.54, 0.84]	0.73 [0.58, 0.84]
Zuo 2016	52	10	38	20	0.58 [0.47, 0.68]	0.67 [0.47, 0.83]



Test 2. Ultrasound

Ultrasound

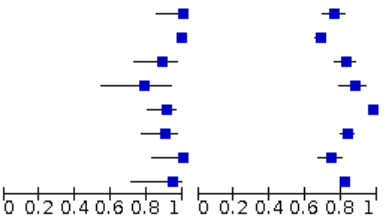
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Atiq 2017	33	119	45	326	0.42 [0.31, 0.54]	0.73 [0.69, 0.77]
Bennett 2002	8	6	19	167	0.30 [0.14, 0.50]	0.97 [0.93, 0.99]
Buffet 1988	20	16	4	168	0.83 [0.63, 0.95]	0.91 [0.86, 0.95]
Chalasani 1999	13	18	9	245	0.59 [0.36, 0.79]	0.93 [0.89, 0.96]
Chang 2015	334	318	29	916	0.92 [0.89, 0.95]	0.74 [0.72, 0.77]
Choi 2019	17	6	18	162	0.49 [0.31, 0.66]	0.96 [0.92, 0.99]
Cottone 1983	27	5	3	65	0.90 [0.73, 0.98]	0.93 [0.84, 0.98]
Cottone 1988	10	3	5	140	0.67 [0.38, 0.88]	0.98 [0.94, 1.00]
Dodd 1992	12	4	16	168	0.43 [0.24, 0.63]	0.98 [0.94, 0.99]
Gambarin-Gelwan 2000	11	5	8	82	0.58 [0.33, 0.80]	0.94 [0.87, 0.98]
Garretti 1988	38	54	2	306	0.95 [0.83, 0.99]	0.85 [0.81, 0.89]
Jalli 2015	18	9	12	57	0.60 [0.41, 0.77]	0.86 [0.76, 0.94]
Kim 2001	6	3	10	33	0.38 [0.15, 0.65]	0.92 [0.78, 0.98]
Kim 2019b	36	0	28	328	0.56 [0.43, 0.69]	1.00 [0.99, 1.00]
Kudo 2019	17	9	9	240	0.65 [0.44, 0.83]	0.96 [0.93, 0.98]
Libbrecht 2002	6	0	9	31	0.40 [0.16, 0.68]	1.00 [0.89, 1.00]
Maringhini 1988	128	37	18	180	0.88 [0.81, 0.93]	0.83 [0.77, 0.88]
Mauduit Astolfi 1987	33	3	2	42	0.94 [0.81, 0.99]	0.93 [0.82, 0.99]
Mok 2004	18	15	3	67	0.86 [0.64, 0.97]	0.82 [0.72, 0.89]
Okazaki 1984	12	3	2	228	0.86 [0.57, 0.98]	0.99 [0.96, 1.00]
Park 2020	12	56	31	958	0.28 [0.15, 0.44]	0.94 [0.93, 0.96]
Pateron 1994	11	7	3	97	0.79 [0.49, 0.95]	0.93 [0.87, 0.97]
Pinero 2015	19	2	37	514	0.34 [0.22, 0.48]	1.00 [0.99, 1.00]
Powell-Jackson 1987	24	5	3	79	0.89 [0.71, 0.98]	0.94 [0.87, 0.98]
Raff 2014	3	9	3	148	0.50 [0.12, 0.88]	0.94 [0.89, 0.97]
Saada 1997	2	0	4	33	0.33 [0.04, 0.78]	1.00 [0.89, 1.00]
Sherman 1995	5	33	2	498	0.71 [0.29, 0.96]	0.94 [0.91, 0.96]
Singal 2012	18	34	23	367	0.44 [0.28, 0.60]	0.92 [0.88, 0.94]
Son 2019	11	27	17	352	0.39 [0.22, 0.59]	0.93 [0.90, 0.95]
Sutherland 2017	6	19	0	167	1.00 [0.54, 1.00]	0.90 [0.85, 0.94]
Tanaka 1986	31	6	22	5286	0.58 [0.44, 0.72]	1.00 [1.00, 1.00]
Teefey 2003	8	4	1	12	0.89 [0.52, 1.00]	0.75 [0.48, 0.93]
Tremolada 1989	17	18	3	176	0.85 [0.62, 0.97]	0.91 [0.86, 0.94]
Ungtrakul 2016	16	397	1	1879	0.94 [0.71, 1.00]	0.83 [0.81, 0.84]
Van Thiel 2004	12	5	8	77	0.60 [0.36, 0.81]	0.94 [0.86, 0.98]
Villacastin Ruiz 2016	86	6	21	157	0.80 [0.72, 0.87]	0.96 [0.92, 0.99]
Wong 2008	463	61	9	46	0.98 [0.96, 0.99]	0.43 [0.33, 0.53]
Yang 2019	80	37	10	136	0.89 [0.81, 0.95]	0.79 [0.72, 0.84]
Yu 2011	88	11	50	281	0.64 [0.55, 0.72]	0.96 [0.93, 0.98]



Test 3. US + AFP

US + AFP

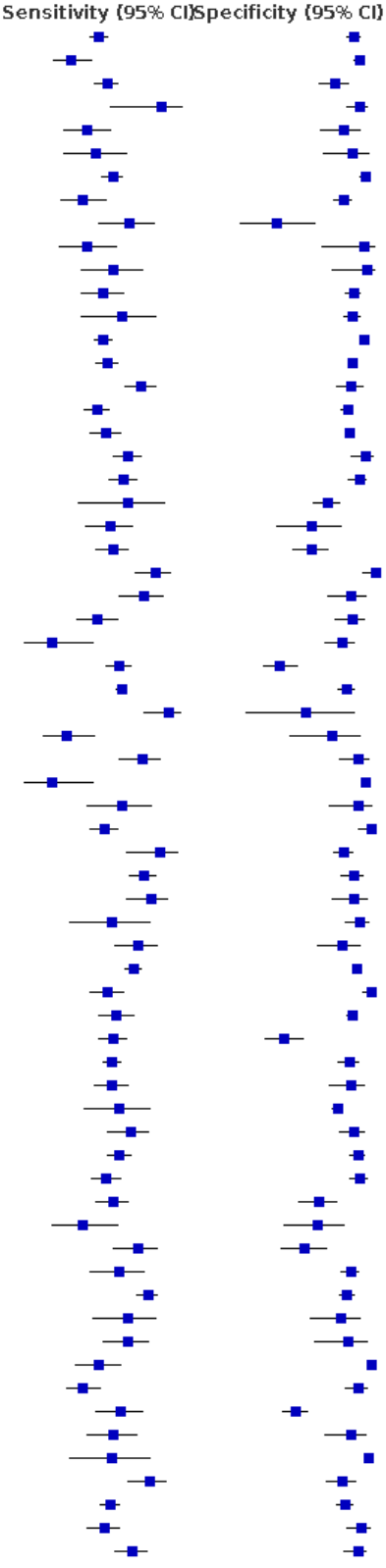
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Buffet 1988	23	44	0	140	1.00 [0.85, 1.00]	0.76 [0.69, 0.82]
Chang 2015	360	391	3	843	0.99 [0.98, 1.00]	0.68 [0.66, 0.71]
Choi 2019	31	29	4	139	0.89 [0.73, 0.97]	0.83 [0.76, 0.88]
Gambarin-Gelwan 2000	15	11	4	76	0.79 [0.54, 0.94]	0.87 [0.79, 0.94]
Kim 2019b	58	7	6	321	0.91 [0.81, 0.96]	0.98 [0.96, 0.99]
Singal 2012	37	67	4	334	0.90 [0.77, 0.97]	0.83 [0.79, 0.87]
Tremolada 1989	20	50	0	144	1.00 [0.83, 1.00]	0.74 [0.67, 0.80]
Ungtrakul 2016	16	404	1	1872	0.94 [0.71, 1.00]	0.82 [0.81, 0.84]



Test 4. AFP cut-off around 20 ng/mL

AFP cut-off around 20 ng/mL

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahn 2016	185	45	181	321	0.51 [0.45, 0.56]	0.88 [0.84, 0.91]
Atiq 2017	27	51	51	511	0.35 [0.24, 0.46]	0.91 [0.88, 0.93]
Baek 2009	130	23	107	77	0.55 [0.48, 0.61]	0.77 [0.68, 0.85]
Bell 1982	12	10	2	100	0.86 [0.57, 0.98]	0.91 [0.84, 0.96]
Beneduce 2004	26	9	34	41	0.43 [0.31, 0.57]	0.82 [0.69, 0.91]
Beneduce 2008	16	4	17	27	0.48 [0.31, 0.66]	0.87 [0.70, 0.96]
Best 2016	166	24	119	378	0.58 [0.52, 0.64]	0.94 [0.91, 0.96]
Bolondi 2001	25	46	36	206	0.41 [0.29, 0.54]	0.82 [0.76, 0.86]
Bon 1998	25	13	12	10	0.68 [0.50, 0.82]	0.43 [0.23, 0.66]
Brunello 1993	17	1	22	15	0.44 [0.28, 0.60]	0.94 [0.70, 1.00]
Capurro 2003	20	1	14	19	0.59 [0.41, 0.75]	0.95 [0.75, 1.00]
Cedrone 2000	39	34	35	242	0.53 [0.41, 0.64]	0.88 [0.83, 0.91]
Chalasani 1999	14	34	8	229	0.64 [0.41, 0.83]	0.87 [0.82, 0.91]
Chang 2015	192	83	171	1151	0.53 [0.48, 0.58]	0.93 [0.92, 0.95]
Chen 2003	142	466	115	2989	0.55 [0.49, 0.61]	0.87 [0.85, 0.88]
Chen 2015	77	13	26	82	0.75 [0.65, 0.83]	0.86 [0.78, 0.93]
Chen 2018	99	70	103	371	0.49 [0.42, 0.56]	0.84 [0.80, 0.87]
Cheng 2012	71	229	60	1362	0.54 [0.45, 0.63]	0.86 [0.84, 0.87]
Chimparlee 2015	105	4	52	69	0.67 [0.59, 0.74]	0.95 [0.87, 0.98]
Chuaypen 2018	97	14	53	136	0.65 [0.56, 0.72]	0.91 [0.85, 0.95]
Cottone 1988	10	39	5	104	0.67 [0.38, 0.88]	0.73 [0.65, 0.80]
Cui 2002	34	11	26	19	0.57 [0.43, 0.69]	0.63 [0.44, 0.80]
Cui 2003	70	33	50	57	0.58 [0.49, 0.67]	0.63 [0.53, 0.73]
da Costa 2015a	49	0	10	49	0.83 [0.71, 0.92]	1.00 [0.93, 1.00]
da Costa 2015b	38	7	12	43	0.76 [0.62, 0.87]	0.86 [0.73, 0.94]
da Costa 2015c	37	10	38	65	0.49 [0.38, 0.61]	0.87 [0.77, 0.93]
da Costa 2015d	5	18	16	76	0.24 [0.08, 0.47]	0.81 [0.71, 0.88]
Dong 2015	118	62	72	52	0.62 [0.55, 0.69]	0.46 [0.36, 0.55]
Edoo 2019	680	46	395	237	0.63 [0.60, 0.66]	0.84 [0.79, 0.88]
El Gawad 2014	36	4	4	6	0.90 [0.76, 0.97]	0.60 [0.26, 0.88]
el-Houseini 2005	14	5	30	15	0.32 [0.19, 0.48]	0.75 [0.51, 0.91]
Elnemr 2012	45	6	15	54	0.75 [0.62, 0.85]	0.90 [0.79, 0.96]
El-Serag 2017	5	32	16	511	0.24 [0.08, 0.47]	0.94 [0.92, 0.96]
El-Sherif 2012	19	3	11	27	0.63 [0.44, 0.80]	0.90 [0.73, 0.98]
Ezzikouri 2015	85	2	74	72	0.53 [0.45, 0.61]	0.97 [0.91, 1.00]
Fabris 1991	23	38	4	173	0.85 [0.66, 0.96]	0.82 [0.76, 0.87]
Fang 2010	110	16	35	112	0.76 [0.68, 0.83]	0.88 [0.80, 0.93]
Fujii1995	40	6	10	44	0.80 [0.66, 0.90]	0.88 [0.76, 0.95]
Gambarin-Gelwan 2000	11	8	8	79	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]
Gani 2015	43	9	16	38	0.73 [0.60, 0.84]	0.81 [0.67, 0.91]
Gopal 2014	316	69	135	603	0.70 [0.66, 0.74]	0.90 [0.87, 0.92]
Grazi 1995	61	3	50	113	0.55 [0.45, 0.64]	0.97 [0.93, 0.99]
Hallager 2018	63	71	41	469	0.61 [0.51, 0.70]	0.87 [0.84, 0.90]
Han 2014	93	46	67	42	0.58 [0.50, 0.66]	0.48 [0.37, 0.59]
Hu 2018	213	26	156	150	0.58 [0.53, 0.63]	0.85 [0.79, 0.90]
Iizuka 2010a	62	8	46	48	0.57 [0.48, 0.67]	0.86 [0.74, 0.94]
Ishii 2000	18	153	11	552	0.62 [0.42, 0.79]	0.78 [0.75, 0.81]
Ismail 2017a	45	12	21	87	0.68 [0.56, 0.79]	0.88 [0.80, 0.94]
Jang 2016	129	19	79	174	0.62 [0.55, 0.69]	0.90 [0.85, 0.94]
Jeon 2016	86	14	71	142	0.55 [0.47, 0.63]	0.91 [0.85, 0.95]
Ji 2016	70	24	50	51	0.58 [0.49, 0.67]	0.68 [0.56, 0.78]
Kanmura 2007	12	11	17	22	0.41 [0.24, 0.61]	0.67 [0.48, 0.82]
Kim 2006b	40	25	15	37	0.73 [0.59, 0.84]	0.60 [0.46, 0.72]
Kim 2006c	26	26	16	159	0.62 [0.46, 0.76]	0.86 [0.80, 0.91]
Kim 2012	154	57	42	297	0.79 [0.72, 0.84]	0.84 [0.80, 0.88]
Kim 2014	20	7	10	28	0.67 [0.47, 0.83]	0.80 [0.63, 0.92]
Kim 2018	36	4	18	22	0.67 [0.53, 0.79]	0.85 [0.65, 0.96]
Kim 2019b	32	7	32	321	0.50 [0.37, 0.63]	0.98 [0.96, 0.99]
Kumada 2014	43	10	61	94	0.41 [0.32, 0.51]	0.90 [0.83, 0.95]
Lee 2004	34	91	20	108	0.63 [0.49, 0.76]	0.54 [0.47, 0.61]
Li 2016a	31	6	22	36	0.58 [0.44, 0.72]	0.86 [0.71, 0.95]
Li 2017b	11	10	8	271	0.58 [0.33, 0.80]	0.96 [0.94, 0.98]
Liao 2012	47	18	12	78	0.80 [0.67, 0.89]	0.81 [0.72, 0.88]
Lim 2015	205	48	156	228	0.57 [0.52, 0.62]	0.83 [0.78, 0.87]
Lin 2000	65	6	57	70	0.53 [0.44, 0.62]	0.92 [0.84, 0.97]
Lin 2015	75	10	33	88	0.69 [0.60, 0.78]	0.90 [0.82, 0.95]

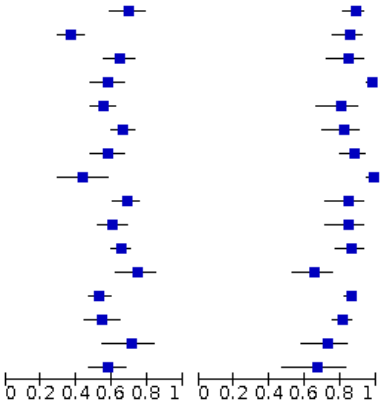


Test 4. (Continued)

Lin 2000	65	6	57	70	0.53 [0.44, 0.62]	0.92 [0.84, 0.97]
Lin 2015	75	10	33	88	0.69 [0.60, 0.78]	0.90 [0.82, 0.95]
Lin 2016	17	23	9	27	0.65 [0.44, 0.83]	0.54 [0.39, 0.68]
Liu 2007	161	29	66	51	0.71 [0.65, 0.77]	0.64 [0.52, 0.74]
Liu 2017	136	16	104	79	0.57 [0.50, 0.63]	0.83 [0.74, 0.90]
Lok 2010	24	15	15	62	0.62 [0.45, 0.77]	0.81 [0.70, 0.89]
Long 2011	76	13	35	56	0.68 [0.59, 0.77]	0.81 [0.70, 0.90]
Luo 2018a	22	18	14	23	0.61 [0.43, 0.77]	0.56 [0.40, 0.72]
Luo 2018b	183	27	142	99	0.56 [0.51, 0.62]	0.79 [0.70, 0.85]
Luo 2018c	78	38	77	105	0.50 [0.42, 0.58]	0.73 [0.65, 0.80]
Mao 2017	37	13	45	44	0.45 [0.34, 0.57]	0.77 [0.64, 0.87]
Maringhini 1988	114	46	32	171	0.78 [0.70, 0.84]	0.79 [0.73, 0.84]
Marrero 2009	247	42	172	375	0.59 [0.54, 0.64]	0.90 [0.87, 0.93]
Moriya 2013	4	13	11	170	0.27 [0.08, 0.55]	0.93 [0.88, 0.96]
Na 2013	36	13	21	51	0.63 [0.49, 0.76]	0.80 [0.68, 0.89]
Nakamura 2006	844	24	517	324	0.62 [0.59, 0.65]	0.93 [0.90, 0.96]
Nguyen 2002	103	30	60	119	0.63 [0.55, 0.71]	0.80 [0.73, 0.86]
Nomura 1996	17	17	10	84	0.63 [0.42, 0.81]	0.83 [0.74, 0.90]
Nomura 1999	21	12	15	37	0.58 [0.41, 0.74]	0.76 [0.61, 0.87]
Oka 1994	21	48	34	150	0.38 [0.25, 0.52]	0.76 [0.69, 0.82]
Oka 2001	266	108	122	104	0.69 [0.64, 0.73]	0.49 [0.42, 0.56]
Okazaki 1984	11	44	3	187	0.79 [0.49, 0.95]	0.81 [0.75, 0.86]
Ozkan 2011	30	1	45	54	0.40 [0.29, 0.52]	0.98 [0.90, 1.00]
Passos-Castilho 2015	12	2	20	28	0.38 [0.21, 0.56]	0.93 [0.78, 0.99]
Paul 2007	68	29	33	165	0.67 [0.57, 0.76]	0.85 [0.79, 0.90]
Pompili 2003	70	9	61	50	0.53 [0.45, 0.62]	0.85 [0.73, 0.93]
Raedle 1995	6	20	1	120	0.86 [0.42, 1.00]	0.86 [0.79, 0.91]
Raedle 1998	52	52	23	584	0.69 [0.58, 0.79]	0.92 [0.89, 0.94]
Raff 2014	8	20	20	308	0.29 [0.13, 0.49]	0.94 [0.91, 0.96]
Reichl 2015	171	2	138	28	0.55 [0.50, 0.61]	0.93 [0.78, 0.99]
Sanai 2010	118	34	88	165	0.57 [0.50, 0.64]	0.83 [0.77, 0.88]
Sarwar 2014	125	14	48	88	0.72 [0.65, 0.79]	0.86 [0.78, 0.92]
Shaheen 2015	25	14	15	16	0.63 [0.46, 0.77]	0.53 [0.34, 0.72]
Shang 2012a	21	5	19	68	0.53 [0.36, 0.68]	0.93 [0.85, 0.98]
Shang 2012b	71	1	20	22	0.78 [0.68, 0.86]	0.96 [0.78, 1.00]
Shen 2012a	245	42	179	95	0.58 [0.53, 0.63]	0.69 [0.61, 0.77]
Shen 2012b	140	82	69	45	0.67 [0.60, 0.73]	0.35 [0.27, 0.44]
Sherman 1995	9	91	5	964	0.64 [0.35, 0.87]	0.91 [0.90, 0.93]
Shimizu 2002	33	5	23	29	0.59 [0.45, 0.72]	0.85 [0.69, 0.95]
Shu 2010	113	44	49	86	0.70 [0.62, 0.77]	0.66 [0.57, 0.74]
Singal 2012	27	38	14	363	0.66 [0.49, 0.80]	0.91 [0.87, 0.93]
Song 2002	18	8	20	23	0.47 [0.31, 0.64]	0.74 [0.55, 0.88]
Song 2014	374	3	176	82	0.68 [0.64, 0.72]	0.96 [0.90, 0.99]
Soroida 2012	131	64	121	391	0.52 [0.46, 0.58]	0.86 [0.82, 0.89]
Sterling 2009	45	86	29	212	0.61 [0.49, 0.72]	0.71 [0.66, 0.76]
Sterling 2012	25	219	21	590	0.54 [0.39, 0.69]	0.73 [0.70, 0.76]
Sultanik 2017	29	21	17	95	0.63 [0.48, 0.77]	0.82 [0.74, 0.88]
Sun 2010	48	26	40	38	0.55 [0.44, 0.65]	0.59 [0.46, 0.71]
Takikawa 1992	81	62	35	191	0.70 [0.61, 0.78]	0.75 [0.70, 0.81]
Tan 2012	160	54	102	96	0.61 [0.55, 0.67]	0.64 [0.56, 0.72]
Tanglijvanich 2010	73	23	27	77	0.73 [0.63, 0.81]	0.77 [0.68, 0.85]
Teng 2016	63	22	48	44	0.57 [0.47, 0.66]	0.67 [0.54, 0.78]
Tian 2017	66	56	54	90	0.55 [0.46, 0.64]	0.62 [0.53, 0.70]
Tremolada 1989	15	39	5	155	0.75 [0.51, 0.91]	0.80 [0.74, 0.85]
Trevisani 2001	102	16	68	154	0.60 [0.52, 0.67]	0.91 [0.85, 0.95]
Tsai 2017	253	58	240	435	0.51 [0.47, 0.56]	0.88 [0.85, 0.91]
Tsuda 2004	29	11	27	21	0.52 [0.38, 0.65]	0.66 [0.47, 0.81]
Ungtrakul 2016	7	29	10	2247	0.41 [0.18, 0.67]	0.99 [0.98, 0.99]
Vongsuvanh 2016	37	6	49	166	0.43 [0.32, 0.54]	0.97 [0.93, 0.99]
Wang 2005	36	15	25	51	0.59 [0.46, 0.71]	0.77 [0.65, 0.87]
Wang 2013a	19	6	29	34	0.40 [0.26, 0.55]	0.85 [0.70, 0.94]
Wang 2014a	25	7	15	27	0.63 [0.46, 0.77]	0.79 [0.62, 0.91]
Weiss 2019	5	4	15	36	0.25 [0.09, 0.49]	0.90 [0.76, 0.97]
Wong 2008	363	51	109	56	0.77 [0.73, 0.81]	0.52 [0.42, 0.62]
Wong 2009	24	0	13	37	0.65 [0.47, 0.80]	1.00 [0.91, 1.00]
Wu 2017	225	8	141	24	0.61 [0.56, 0.66]	0.75 [0.57, 0.89]
Wu 2020	122	34	76	92	0.62 [0.54, 0.68]	0.73 [0.64, 0.81]
Xu 2018	61	16	27	119	0.69 [0.59, 0.79]	0.88 [0.81, 0.93]
Yang 2013a	66	12	113	68	0.37 [0.30, 0.44]	0.85 [0.75, 0.92]

Test 4. (Continued)

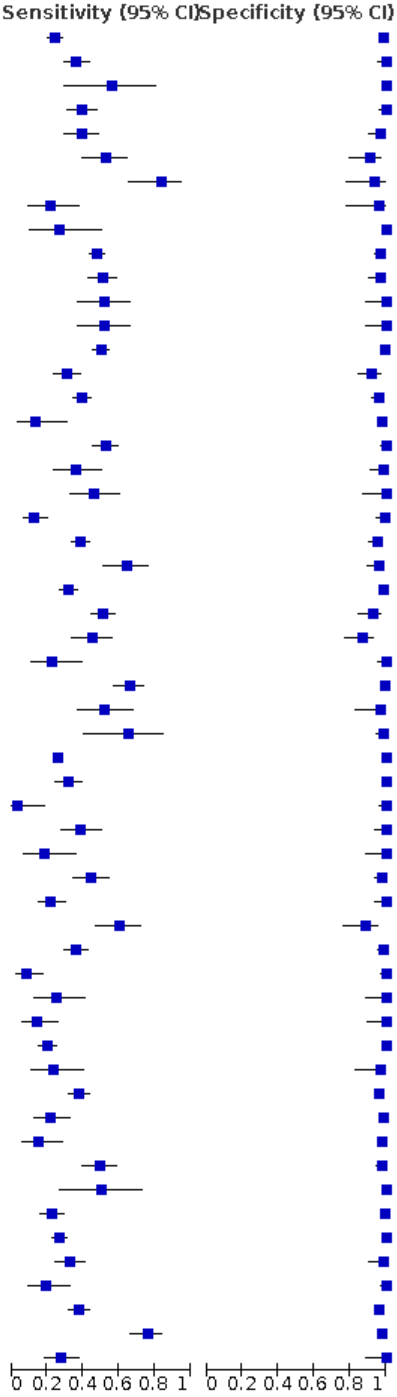
Xu 2018	61	16	27	119	0.69 [0.59, 0.79]	0.88 [0.81, 0.93]
Yang 2013a	66	12	113	68	0.37 [0.30, 0.44]	0.85 [0.75, 0.92]
Yang 2014	79	9	44	48	0.64 [0.55, 0.73]	0.84 [0.72, 0.93]
Yang 2019	64	4	47	176	0.58 [0.48, 0.67]	0.98 [0.94, 0.99]
Ye 2019a	115	10	94	40	0.55 [0.48, 0.62]	0.80 [0.66, 0.90]
Ye 2019b	135	11	69	49	0.66 [0.59, 0.73]	0.82 [0.70, 0.90]
Yoon 2009	61	12	45	88	0.58 [0.48, 0.67]	0.88 [0.80, 0.94]
Yu 2016	22	2	29	136	0.43 [0.29, 0.58]	0.99 [0.95, 1.00]
Yu 2020a	108	8	50	42	0.68 [0.60, 0.76]	0.84 [0.71, 0.93]
Yu 2020b	92	8	60	42	0.61 [0.52, 0.68]	0.84 [0.71, 0.93]
Yu 2020c	189	11	101	69	0.65 [0.59, 0.71]	0.86 [0.77, 0.93]
Zhang 2020	47	26	16	48	0.75 [0.62, 0.85]	0.65 [0.53, 0.76]
Zhu 2013	137	61	121	374	0.53 [0.47, 0.59]	0.86 [0.82, 0.89]
Zhu 2020	55	42	46	179	0.54 [0.44, 0.64]	0.81 [0.75, 0.86]
Zinkin 2008	29	14	12	37	0.71 [0.54, 0.84]	0.73 [0.58, 0.84]
Zuo 2016	52	10	38	20	0.58 [0.47, 0.68]	0.67 [0.47, 0.83]



Test 5. AFP cut-off around 200 ng/mL

AFP cut-off around 200 ng/mL

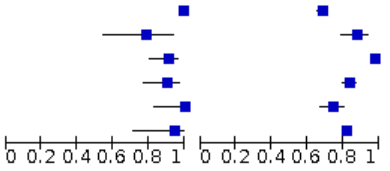
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahn 2016	89	6	277	360	0.24 [0.20, 0.29]	0.98 [0.96, 0.99]
Arrieta 2007	70	0	123	74	0.36 [0.29, 0.43]	1.00 [0.95, 1.00]
Arrigoni 1988	9	0	7	148	0.56 [0.30, 0.80]	1.00 [0.98, 1.00]
Attallah 2011	59	0	91	100	0.39 [0.31, 0.48]	1.00 [0.96, 1.00]
Attallah 2018	43	2	67	70	0.39 [0.30, 0.49]	0.97 [0.90, 1.00]
Bachtiar 2009	34	5	31	49	0.52 [0.40, 0.65]	0.91 [0.80, 0.97]
Badr 2014	25	2	5	28	0.83 [0.65, 0.94]	0.93 [0.78, 0.99]
Bon 1998	8	1	29	22	0.22 [0.10, 0.38]	0.96 [0.78, 1.00]
Chalasani 1999	6	0	16	263	0.27 [0.11, 0.50]	1.00 [0.99, 1.00]
Chan 2014	268	7	294	236	0.48 [0.43, 0.52]	0.97 [0.94, 0.99]
Chimparlee 2015	80	2	77	71	0.51 [0.43, 0.59]	0.97 [0.90, 1.00]
Eissa 2013	26	0	24	30	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
El Moety 2011	26	0	24	30	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
Gopal 2014	226	5	226	667	0.50 [0.45, 0.55]	0.99 [0.98, 1.00]
Han 2014	49	7	111	81	0.31 [0.24, 0.38]	0.92 [0.84, 0.97]
Hu 2018	146	7	223	169	0.40 [0.35, 0.45]	0.96 [0.92, 0.98]
Ishii 2000	4	18	25	687	0.14 [0.04, 0.32]	0.97 [0.96, 0.98]
Iyer 2018	107	0	96	119	0.53 [0.46, 0.60]	1.00 [0.97, 1.00]
Kim 2006b	20	1	35	61	0.36 [0.24, 0.50]	0.98 [0.91, 1.00]
Kim 2018	25	0	29	26	0.46 [0.33, 0.60]	1.00 [0.87, 1.00]
Kumada 2014	13	1	91	103	0.13 [0.07, 0.20]	0.99 [0.95, 1.00]
Lee 2004	143	9	226	169	0.39 [0.34, 0.44]	0.95 [0.91, 0.98]
Liao 2012	38	4	21	92	0.64 [0.51, 0.76]	0.96 [0.90, 0.99]
Lim 2015	115	4	246	272	0.32 [0.27, 0.37]	0.99 [0.96, 1.00]
Liu 2007	116	6	111	74	0.51 [0.44, 0.58]	0.93 [0.84, 0.97]
Liu 2018	36	11	44	71	0.45 [0.34, 0.57]	0.87 [0.77, 0.93]
Lok 2010	9	0	30	77	0.23 [0.11, 0.39]	1.00 [0.95, 1.00]
Maringhini 1988	96	1	50	216	0.66 [0.57, 0.73]	1.00 [0.97, 1.00]
Mashaly 2018	23	1	21	30	0.52 [0.37, 0.68]	0.97 [0.83, 1.00]
Matievskaya 2003	13	2	7	137	0.65 [0.41, 0.85]	0.99 [0.95, 1.00]
Nakamura 2006	354	0	1007	348	0.26 [0.24, 0.28]	1.00 [0.99, 1.00]
Nguyen 2002	52	0	111	149	0.32 [0.25, 0.40]	1.00 [0.98, 1.00]
Nomura 1996	1	0	26	101	0.04 [0.00, 0.19]	1.00 [0.96, 1.00]
Ozkan 2011	29	0	46	55	0.39 [0.28, 0.51]	1.00 [0.94, 1.00]
Passos-Castilho 2015	6	0	26	30	0.19 [0.07, 0.36]	1.00 [0.88, 1.00]
Paul 2007	45	5	56	189	0.45 [0.35, 0.55]	0.97 [0.94, 0.99]
Pompili 2003	29	0	102	59	0.22 [0.15, 0.30]	1.00 [0.94, 1.00]
Poon 2001	39	6	26	45	0.60 [0.47, 0.72]	0.88 [0.76, 0.96]
Sanai 2010	74	3	132	196	0.36 [0.29, 0.43]	0.98 [0.96, 1.00]
Sassa 1999	5	0	56	134	0.08 [0.03, 0.18]	1.00 [0.97, 1.00]
Shaheen 2015	10	0	30	30	0.25 [0.13, 0.41]	1.00 [0.88, 1.00]
Shimizu 2002	8	0	48	34	0.14 [0.06, 0.26]	1.00 [0.90, 1.00]
Snowberger 2007	48	0	191	1696	0.20 [0.15, 0.26]	1.00 [1.00, 1.00]
Song 2002	9	1	29	30	0.24 [0.11, 0.40]	0.97 [0.83, 1.00]
Soroida 2012	95	18	157	437	0.38 [0.32, 0.44]	0.96 [0.94, 0.98]
Sterling 2009	16	4	58	274	0.22 [0.13, 0.33]	0.99 [0.96, 1.00]
Sterling 2012	7	17	39	792	0.15 [0.06, 0.29]	0.98 [0.97, 0.99]
Takikawa 1992	57	6	59	247	0.49 [0.40, 0.59]	0.98 [0.95, 0.99]
Tremolada 1989	10	0	10	194	0.50 [0.27, 0.73]	1.00 [0.98, 1.00]
Trevisani 2001	38	1	132	169	0.22 [0.16, 0.29]	0.99 [0.97, 1.00]
Tsai 2017	134	0	359	493	0.27 [0.23, 0.31]	1.00 [0.99, 1.00]
Yang 2013a	40	1	83	56	0.33 [0.24, 0.42]	0.98 [0.91, 1.00]
Yu 2016	10	0	41	138	0.20 [0.10, 0.33]	1.00 [0.97, 1.00]
Zhu 2013	95	18	157	437	0.38 [0.32, 0.44]	0.96 [0.94, 0.98]
Ziada 2016	78	11	25	400	0.76 [0.66, 0.84]	0.97 [0.95, 0.99]
Zuo 2016	25	0	65	30	0.28 [0.19, 0.38]	1.00 [0.88, 1.00]



Test 6. US + AFP cut-off 20 ng/mL

US + AFP cut-off 20 ng/mL

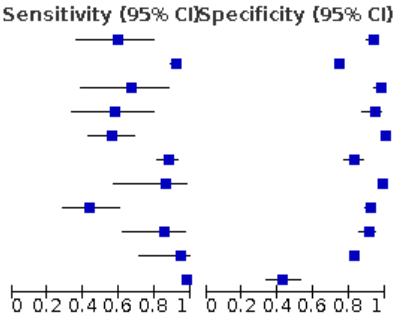
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chang 2015	360	391	3	843	0.99 [0.98, 1.00]	0.68 [0.66, 0.71]
Gambarin-Gelwan 2000	15	11	4	76	0.79 [0.54, 0.94]	0.87 [0.79, 0.94]
Kim 2019b	58	7	6	321	0.91 [0.81, 0.96]	0.98 [0.96, 0.99]
Singal 2012	37	67	4	334	0.90 [0.77, 0.97]	0.83 [0.79, 0.87]
Tremolada 1989	20	50	0	144	1.00 [0.83, 1.00]	0.74 [0.67, 0.80]
Ungtrakul 2016	16	404	1	1872	0.94 [0.71, 1.00]	0.82 [0.81, 0.84]



Test 7. US for direct comparison AFP 20 ng/mL

US for direct comparison AFP 20 ng/mL

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chalasani 1999	13	18	9	245	0.59 [0.36, 0.79]	0.93 [0.89, 0.96]
Chang 2015	334	318	29	916	0.92 [0.89, 0.95]	0.74 [0.72, 0.77]
Cottone 1988	10	3	5	140	0.67 [0.38, 0.88]	0.98 [0.94, 1.00]
Gambarin-Gelwan 2000	11	5	8	82	0.58 [0.33, 0.80]	0.94 [0.87, 0.98]
Kim 2019b	36	0	28	328	0.56 [0.43, 0.69]	1.00 [0.99, 1.00]
Maringhini 1988	128	37	18	180	0.88 [0.81, 0.93]	0.83 [0.77, 0.88]
Okazaki 1984	12	3	2	228	0.86 [0.57, 0.98]	0.99 [0.96, 1.00]
Singal 2012	18	34	23	367	0.44 [0.28, 0.60]	0.92 [0.88, 0.94]
Tremolada 1989	17	18	3	176	0.85 [0.62, 0.97]	0.91 [0.86, 0.94]
Ungtrakul 2016	16	397	1	1879	0.94 [0.71, 1.00]	0.83 [0.81, 0.84]
Wong 2008	463	61	9	46	0.98 [0.96, 0.99]	0.43 [0.33, 0.53]



ADDITIONAL TABLES

Table 1. Guideline recommendations for surveillance for hepatocellular carcinoma

GUIDELINE	INDICATION TO SURVEILANCE	TEST	INTERVAL
American Association for the Study of Liver Disease (AASLD; (Heimbach 2018))	Cirrhosis	Abdominal ultrasound alone or plus AFP	6 months
European Association for the Study of the Liver with European Organization for Research and Treatment of Cancer (EASL-EORTC; (EASL-EORTC 2012; EASL 2018))	Cirrhosis in Child Pugh stages A and B; cirrhosis in Child C stage awaiting liver transplantation; non-cirrhotic hepatitis B virus (HBV) carriers with active hepatitis or family history of HCC; non-cirrhotic chronic hepatitis C with advanced liver fibrosis stage 3 (F3)	Abdominal ultrasound	6 months 3 to 4 months: people with a nodule less than 1 cm or after resection or loco-regional therapies
Asian Pacific Association for the Study of the Liver (APASL; (Omata 2017))	Cirrhosis and chronic HBV infection at risk of HCC	Abdominal ultrasound with serum AFP	6 months

AFP: alpha-foetoprotein; **HCC:** hepatocellular carcinoma

Table 2. Heterogeneity and sensitivity analyses for alpha-foetoprotein (AFP) cut-off value around 20 ng/mL

Subgroup	N of studies	Sensitivity (95% CI)	Specificity (95% CI)	P value

Table 2. Heterogeneity and sensitivity analyses for alpha-foetoprotein (AFP) cut-off value around 20 ng/mL *(Continued)*

All	147	60% (58% to 62%)	84% (82% to 86%)	
case-control	111	60% (58% to 62%)	83% (81% to 85%)	0.133
cross-sectional	36	57% (52% to 62%)	88% (84% to 91%)	
prospective	29	59% (54% to 63%)	86% (81% to 90%)	0.828
retrospective	118	60% (58% to 62%)	84% (82% to 86%)	
before 2000	22	65% (59% to 71%)	85% (81% to 88%)	0.264
after 2000	125	59% (57% to 61%)	84% (82% to 86%)	
cirrhosis > 10%	94	59% (56% to 61%)	85% (82% to 87%)	§
cirrhosis < 10%	2	61% (51% to 70%)* 57% (50% to 63%)**	87% (84% to 90%)* 83% (74% to 90%)**	
Europe	22	60% (54% to 65%)	87% (83% to 90%)	0.447
America	19	56% (50% to 61%)	89% (85% to 92%)	
Asia	98	60% (58% to 62%)	83% (80%to 86%)	
Africa	7	68% (54% to 80%)	81% (71% to 89%)	
HCC prevalence < 10%	16	54% (47% to 62%)	89% (84% to 93%)	0.147
HCC prevalence > 10%	131	60% (58% to 62%)	84% (81% to 86%)	
clinical suspect	117	61% (59% to 63%)	83% (80% to 85%)	0.005
surveillance	30	54% (49% to 60%)	89% (86% to 92%)	
HCC resectable < 20%	4	61% (48% to 72%)	82% (64% to 92%)	0.909

Table 2. Heterogeneity and sensitivity analyses for alpha-foetoprotein (AFP) cut-off value around 20 ng/mL *(Continued)*

HCC resectable > 20%	25	56% (51% to 61%)	87% (81% to 91%)	
biopsy	22	63% (58% to 68%)	82% (77% to 87%)	0.832
other reference standard	124	59% (57% to 61%)	85% (82% to 87%)	
viral < 80%	35	59% (55% to 63%)	87% (83% to 90%)	0.694
viral > 80%	84	59% (57% to 62%)	84% (81% to 86%)	
Child A < 50%	17	59% (52% to 67%)	86% (82%to 89%)	0.746
Child A > 50%	34	59% (55% to 62%)	83% (77% to 87%)	
Full text	142	60% (58% to 62%)	84% (82% to 86%)	

* Hallager 2018 ; ** Liu 2017
§ Model failed to converge
HCC: hepatocellular carcinoma

Table 3. Heterogeneity and sensitivity analyses for alpha-foetoprotein (AFP) cut-off value around 200 ng/mL

Subgroup	N of studies	Sensitivity (95% CI)	Specificity (95% CI)	P value
All	56	36% (31% to 41%)	99% (98% to 100%)	-
case-control	42	35% (30% to 40%)	99% (98% to 100%)	0.874
cross-sectional	14	39% (28% to 51%)	99% (98% to 100%)	
prospective	9	42% (27% to 58%)	99% (97% to 100%)	0.713
retrospective	47	35% (30% to 40%)	99% (98% to 100%)	
before 2000	9	28% (15% to 47%)	100% (98% to 100%)	0.336
after 2000	47	37% (33% to 42%)	99% (98% to 100%)	

Table 3. Heterogeneity and sensitivity analyses for alpha-foetoprotein (AFP) cut-off value around 200 ng/mL *(Continued)*

cirrhosis > 10%	41	40% (28% to 40%)	99% (99% to 100%)	-
cirrhosis < 10%	0	-	-	
Europe	8	40% (28% to 54%)	99% (98% to 100%)	0.020
America	9	27% (21% to 35%)	100% (98% to 100%)	
Asia	31	34% (29% to 40%)	98% (97% to 99%)	
Africa	8	53% (39% to 66%)	99% (97% to 100%)	
HCC prevalence < 10%	5	30% (16% to 48%)	100% (95% to 100%)	0.805
HCC prevalence > 10%	51	36% (32% to 41%)	99% (98% to 99%)	
clinical suspect	49	36% (32% to 41%)	99% (98% to100%)	0.995
surveillance	7	34% (18% to 54%)	99% (96% to 100%)	
HCC resectable < 20%	2	42% (8% to 85%)	99% (82% to 100%)	0.931
HCC resectable > 20%	8	27% (12% to 50%)	99% (97% to 100%)	
biopsy	9	31% (24% to 39%)	100% (97% to 100%)	0.140
other reference standard	46	37% (32% to 43%)	99% (98% to 100%)	
viral < 80%	11	37% (29% to 46%)	99% (97% to 100%)	0.705
viral > 80%	30	32% (26% to 39%)	98% (98% to 100%)	
Child A < 50%	13	42% (31% to 54%)	99% (99% to 100%)	0.008
Child A > 50%	11	24% (19% to 29%)	99% (97 to 100%)	
Full text	54	36% (31% to 41%)	99% (98% to 100%)	-

HCC: hepatocellular carcinoma

Table 4. Heterogeneity and sensitivity analyses for ultrasonography (US)

Subgroup	N of studies	Sensitivity (95% CI)	Specificity (95% CI)	P value
All	39	72% (63% to 79%)	94% (91% to 96%)	-
case-control	3	82% (64% to 92%)	87% (77% to 93%)	0.737
cross-sectional	36	71% (62% to 79%)	95% (92% to 97%)	
prospective	18	72% (60% to 81%)	94% (90% to 96%)	1.000
retrospective	21	72% (58% to 82%)	94% (89% to 97%)	
before 2000	16	79% (70% to 86%)	96% (92% to 98%)	0.091
after 2000	23	67% (54% to 78%)	93% (88% to 96)	
cirrhosis > 10%	33	70% (60% to 78%)	94% (91% to 96%)	-
cirrhosis < 10%	0			
Europe	12	82% (73% to 89%)	94% (90% to 97%)	0.186
America	13	57% (45% to 68%)	94% (89% to 96%)	
Asia	13	76% (58% to 88%)	94% (85% to 98%)	
Africa	0	-	-	
HCC prevalence < 10%	15	69% (54% to 81%)	96% (92% to 98%)	0.660
HCC prevalence > 10%	24	74% (62% to 82%)	93% (88% to 96%)	
clinical suspect	19	74% (61% to 84%)	93% (89% to 96%)	0.898
surveillance	20	69% (57% to 79%)	95% (91% to 98%)	

Table 4. Heterogeneity and sensitivity analyses for ultrasonography (US) *(Continued)*

HCC resectable < 20%	4	90% (75% to 97%)	82% (60% to 94%)	0.088
HCC resectable > 20%	16	66% (52% to 77%)	95% (91% to 97%)	
biopsy	7	81% (64% to 91%)	90% (84% to 94%)	0.379
OLT	10	55% (41% to 69%)	97% (93% to 96%)	
other reference standard	22	76% (64% to 84%)	94% (89% to 97%)	
viral < 80%	17	70% (57% to 80%)	94% (90% to 96%)	0.777
viral > 80%	9	79% (58% to 91%)	91% (79% to 97%)	
Child A < 50%	5	50% (33% to 68%)	91% (83% to 95%)	0.346
Child A > 50%	9	74% (52% to 88%)	93% (82 to 98%)	
US positivity criteria predefined	25	74% (63% to 83%)	93% (89% to 96%)	-
Uninterpretable test results reported	3	80% (71% to 81%)	76% (71% to 81%)	-
Full text	38	72% (64% to 80%)	94% (91% to 96%)	-

OLT: orthotopic liver transplantation; **HCC**: hepatocellular carcinoma

Table 5. Heterogeneity and sensitivity analyses for the combination of alpha-foetoprotein (AFP) (cut-off 20 ng/mL) and ultrasonography (US)

Subgroup	N of studies	Sensitivity (95% CI)	Specificity (95% CI)	P value
All	6	96% (88% to 98%)	85% (73% to 93%)	-
case-control	0			-
cross-sectional	6	96% (88% to 98%)	85% (73% to 93%)	

Table 5. Heterogeneity and sensitivity analyses for the combination of alpha-foetoprotein (AFP) (cut-off 20 ng/mL) and ultrasonography (US) *(Continued)*

prospective	3	91% (84% to 95%)	91% (75% to 97%)	0.578
retrospective	3	97% (83% to 99%)	77% (66% to 85%)	
before 2000	2	95% (44% to 100%)	81% (69% to 89%)	0.703
after 2000	4	96% (89% to 99%)	87% (69% to 95%)	
cirrhosis > 10%	6	96% (88% to 98%)	85% (73% to 93%)	-
cirrhosis < 10%	0			
Europe	1	100% (83% to 100%)	74% (67% to 80%)	§
America	2	79% (54% to 94%)	87% (79% to 94%)	
		90% (77% to 97%)	83% (79% to 87%)	
Asia	3	99% (98% to 100%)	68% (66% to 71%)	
		91% (81% to 96%)	98% (96% to 99%)	
		94% (71% to 100%)	82% (81% to 84%)	
Africa	0			
HCC prevalence < 10%	3	96% (78% to 99%)	80% (76% to 84%)	0.100
HCC prevalence > 10%	3	95% (79% to 99%)	90% (68% to 97%)	
clinical suspect	1	79% (54% to 94%)	87% (78% to 93%)	0.289
surveillance	5	97% (92% to 99%)	85% (70% to 93%)	
HCC resectable < 20%	0			-
HCC resectable > 20%	4	95% (84% to 99%)	88% (72% to 96%)	
biopsy	1	99% (98% to 100%)	68% (66% to 71%)	§

Table 5. Heterogeneity and sensitivity analyses for the combination of alpha-foetoprotein (AFP) (cut-off 20 ng/mL) and ultrasonography (US) *(Continued)*

OLT	1	79% (54% to 94%)	87% (78% to 93%)	
other reference standard	4	93% (86% to 97%)	88% (72% to 95%)	
viral < 80%	1	79% (54% to 94%)	87% (79% to 94%)	*
viral > 80%	3	99% (98% to 100%)	68% (66% to 71%)	
		91% (81% to 96%)	98% (96% to 99%)	
		94% (71% to 100%)	82% (81% to 84%)	
Child A < 50%	1	100% (83% to 100%)	74% (67% to 80%)	*§
Child A > 50%	2	99% (98% to 100%)	68% (66% to 71%)	
		91% (81% to 96%)	98% (96% to 99%)	
US positivity criteria predefined	2	90% (77% to 97%)	83% (79% to 87%)	§
		94% (71% to 100%)	82% (81% to 84%)	
Full text	6	96% (88% to 98%)	85% (73% to 93%)	-

* Sparse and missing data. Meta-analysis not conducted
§ Model failed to converge
OLT: orthotopic liver transplantation; **HCC**: hepatocellular carcinoma; **US**: ultrasonography

APPENDICES

Appendix 1. Search strategies

Database	Time span	Search strategy
The Cochrane Hepato-Biliary Group Controlled Trials Register	June 2020	((ultrason* or ultrasound* or echograph* or echotomograph* or doppler* or B-mode or B-scan or grey*scale) or (alpha or alfa) AND (fetoprotein* or foetoprotein or fetalprotein)) and diagnos* and (((liver or hepato*) and (carcinom* or cancer* or neoplasm* or malign* or tumo*)) or HCC) and (liver OR hepat* OR cirrhosis OR fibrosis)
The Cochrane Hepato-Biliary Group Diagnostic Test Accuracy Studies Register	June 2020	((ultrason* or ultrasound* or echograph* or echotomograph* or doppler* or B-mode or B-scan or grey*scale) or (alpha or alfa) AND (fetoprotein* or foetoprotein or fetalprotein)) and diagnos* and (((liver or hepato*) and (carcinom* or

(Continued)

		cancer* or neoplasm* or malign* or tumo*)) or HCC) and (liver OR hepat* OR cirrhosis OR fibrosis)
The Cochrane Library	2020, Issue 6	#1 MeSH descriptor: [Ultrasonography] explode all trees #2 (ultrason* or ultrasound* or echograph* or echotomograph* or doppler* or B-mode or B-scan or grey*scale) #3 #1 or #2 #4 MeSH descriptor: [alpha-Fetoproteins] explode all trees #5 (alpha or alfa) AND (fetoprotein* or foetoprotein or fetalprotein) #6 #4 or #5 #7 MeSH descriptor: [Diagnostic Techniques and Procedures] explode all trees #8 diagnos* #9 #7 or #8 #10 MeSH descriptor: [Carcinoma, Hepatocellular] explode all trees #11 MeSH descriptor: [Liver Neoplasms] explode all trees #12 ((liver or hepato*) and (carcinom* or cancer* or neoplasm* or malign* or tumo*)) or HCC #13 #10 or #11 or #12 #14 MeSH descriptor: [Liver Diseases] explode all trees #15 liver OR hepat* OR cirrhosis OR fibrosis #16 #14 or #15 #17 (#3 or #6) and #9 and #13 and #16
MEDLINE Ovid	1946 to June 2020	1. exp ULTRASONOGRAPHY/ 2. (ultrason* or ultrasound* or echograph* or echotomograph* or doppler* or B-mode or B-scan or grey*scale).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] 3. 1 or 2 4. exp alpha-Fetoproteins/ 5. ((alpha or alfa) and (fetoprotein* or foetoprotein or fetalprotein)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] 6. 4 or 5 7. exp "Diagnostic Techniques and Procedures"/ 8. diagnos*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, or-

(Continued)

		ganism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] 9. 7 or 8 10. exp Carcinoma, Hepatocellular/ 11. exp Liver Neoplasms/ 12. (((liver or hepato*) and (carcinom* or cancer* or neoplasm* or malign* or tumo*)) or HCC).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] 13. 10 or 11 or 12 14. exp Liver Diseases/ 15. (liver or hepat* or cirrhosis or fibrosis).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] 16. 14 or 15 17. (3 or 6) and 9 and 13 and 16 18. limit 17 to (humans and ("all adult (19 plus years)" or "young adult (19 to 24 years)" or "adult (19 to 44 years)" or "young adult and adult (19-24 and 19-44)" or "middle age (45 to 64 years)" or "middle aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)"))
Embase Ovid	1974 to June 2020	1. exp echography/ 2. (ultrason* or ultrasound* or echograph* or echotomograph* or doppler* or B-mode or B-scan or grey*scale).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] 3. 1 or 2 4. exp alpha fetoprotein/ 5. ((alpha or alfa) and (fetoprotein* or foetoprotein or fetalprotein)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] 6. 4 or 5 7. exp diagnostic test/ 8. diagnos*.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] 9. 7 or 8 10. exp liver cell carcinoma/ 11. exp liver tumor/

(Continued)

		12. (((liver or hepato*) and (carcinom* or cancer* or neoplasm* or malign* or tumo*)) or HCC).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] 13. 10 or 11 or 12 14. exp liver disease/ 15. (liver or hepat* or cirrhosis or fibrosis).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] 16. 14 or 15 17. (3 or 6) and 9 and 13 and 16 18. limit 17 to (human and (adult <18 to 64 years> or aged <65+ years>))
LILACS (Bireme)	1982 to June 2020	(ultrason$ or ultrasound$ or echograph$ or echotomograph$ or doppler$ or B-mode or B-scan or grey$scale) or (alpha or alfa) AND (fetoprotein$ or foetoprotein or fetalprotein) [Words] and diagnos$ [Words] and (((liver or hepato$) and (carcinom$ or cancer$ or neoplasm$ or malign$ or tumo$)) or HCC) AND (liver OR hepat$ OR cirrhosis OR fibrosis) [Words]
Science Citation Index Expanded (Web of Science)	1900 to June 2020	#6 (#1 or #2) AND #3 AND #4 AND #5 #5 TS=(liver or hepat* or cirrhosis or fibrosis) #4 TS=(((liver or hepato*) and (carcinom* or cancer* or neoplasm* or malign* or tumo*)) or HCC) #3 TS=(diagnos*) #2 TS=((alpha or alfa) and (fetoprotein* or foetoprotein or fetalprotein)) #1 TS=(ultrason* or ultrasound* or echograph* or echotomograph* or doppler* or B-mode or B-scan or grey*scale)
Conference Proceedings Citation Index – Science (Web of Science)	1990 to June 2020	#6 (#1 or #2) AND #3 AND #4 AND #5 #5 TS=(liver or hepat* or cirrhosis or fibrosis) #4 TS=(((liver or hepato*) and (carcinom* or cancer* or neoplasm* or malign* or tumo*)) or HCC) #3 TS=(diagnos*) #2 TS=((alpha or alfa) and (fetoprotein* or foetoprotein or fetalprotein)) #1 TS=(ultrason* or ultrasound* or echograph* or echotomograph* or doppler* or B-mode or B-scan or grey*scale)

Appendix 2. QUADAS-2

Domain	1. Participant selection	2. Index test	3. Reference standard	4. Flow and timing

(Continued)

Signalling questions and criteria	Q1: "*Was a consecutive or random sample of participants enrolled?*" Yes - if the study reports on a consecutive or a random selection of participants. No - if the study reports on another form of selection of participants. Unclear - if the study does not report on how the participants were enrolled. Q2: "*Was a case-control design avoided?*" Yes - if a case-control design was avoided. No - if the study was a case-control. Unclear - if the study design was not clear. Q.3: "*Did the study avoid inappropriate exclusions*?" Yes - if definitions of exclusion criteria are appropriate (i.e. previous surgery or treatment for hepatocellular carcinoma; patients with cholangiocarcinoma) and all exclusions are reported. No - if exclusion criteria are inappropriate and exclusions are not reported. Unclear - if the study does not report causes of exclusions.	Q1: "*Were the index test results interpreted without knowledge of the results of the reference standard?*" **For ultrasonography (US) and AFP:** Yes - if the study reports that the results of the index test were interpreted without the knowledge of the results of the reference standard. No - if the study reports that results of the index test were interpreted with the results of the reference standard. Unclear - if the study does not report information about blinding of the results of the index test and reference standard. Q2: "*If a threshold was used, was it pre-specified?*" **Only for AFP:** Yes - if the threshold used was reported in the methods section. No - if the study reports that the threshold was chosen during the data analysis stage (e.g. maximum of Youden index). Unclear - if the study does not report information about threshold selection. Q3: "*Were positivity criteria clearly defined?*" **Only for US:** Yes - if the study clearly reports positivity criteria (i.e. the minimum diameter of a detectable lesion, exclusion of benign criteria). No - if the study does not report the positivity criteria.	Q1: "*Is the reference standard likely to correctly classify the target condition?*" Yes - if the reference standard correctly defines the presence/absence of HCC (pathology of explanted liver in a transplant cohort or CT MRI or histology of resected or biopsied focal lesions with adequate follow up). No - if other reference tests than pathology of explanted liver or CT MRI or histology of resected or biopsied focal lesions with adequate follow up were used. Unclear - if the study does not report on the reference standard used. Q2: "*Were the reference standard results interpreted without the knowledge of the results of the index test?*" Yes - if the study reports that the results of the reference standard were interpreted without the knowledge of the results of the index test. No - if the study reports that the results of the reference standard were interpreted with the knowledge of the results of the index test. Unclear - if the study does not report information about blinding of the results of the reference standard and the index test.	Q1: "*Was there an appropriate interval between the index test and the reference standard?*" Yes - if the interval between the index test and the reference standard was less than 3 months. No - if the interval was longer than 3 months. Unclear - if the study does not report the interval between the index test and the reference standard. Q2: "*Did all participants receive the same reference standard?*" Yes - if the study has only one reference standard for all the participants. No - if the study has more than one reference standard. Unclear - if the study information regarding the use of reference standard are unclear. Q3: "*Were all participants included in the analysis and analysed according to intention-to-diagnose principle (uninterpretable results considered as false)?*" Yes - if all enrolled participants were included in the analysis and uninterpretable index test results were analysed according to the intention-to-diagnose principle).

(Continued)

				No - if any participant was excluded from the analysis for any reason or uninterpretable index test results were not analysed according to intention-to-diagnose principle. Unclear - if the exclusion of participants from the analysis is unclear.
Risk of bias	*Could the selection of participants have introduced bias?* Low risk: "Yes" for all signalling questions. High risk: "No" or "Unclear" for at least one signalling question.	*Could the conduct or interpretation of the index test have introduced bias?* Low risk: "Yes" for all signalling questions. High risk: "No" or "Unclear" for at least one signalling question.	*Could the reference standard, its conduct, or its interpretation have introduced bias?* Low risk: "Yes" for all signalling questions. High risk: "No" or "Unclear" for at least one signalling question.	*Could the participant flow have introduced bias?* Low risk: "Yes" for all signalling questions. High risk: "No" or "Unclear" for at least one signalling question.
Concerns about applicability	*Are there concerns that included participants and setting do not match the review question?* Low concern: the participants included in the review represent the participants in whom the tests is used in clinical practice (i.e. surveillance programme in patients with cirrhosis; clinical cohort of patients with cirrhosis). High concern: the participants included in the review differ from the participants in whom the tests is used in clinical practice.	*Are there concerns that the index test, its conduct, or interpretation differ from the review question?* Low concern: the index test, its conduct or its interpretation does not differ from the way it is used in clinical practice. High concern: the index test, its conduct or its interpretation differs from the way it is used in clinical practice.	*Are there concerns that the target condition as defined by the reference standard does not match the question?* Low concern: the definition of the target condition as defined by the reference standard does match the question as CT scan or MR for all included patients. High concern: the definition of the target condition as defined by the reference standard does not match the question (i.e. pathology of the explanted liver is feasible only in the case of liver transplant; the natural history and prognosis of HCC detected in explanted liver might be different.)	-

Appendix 3. Study level assessments of study quality

Figure 13

Figure 13. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study

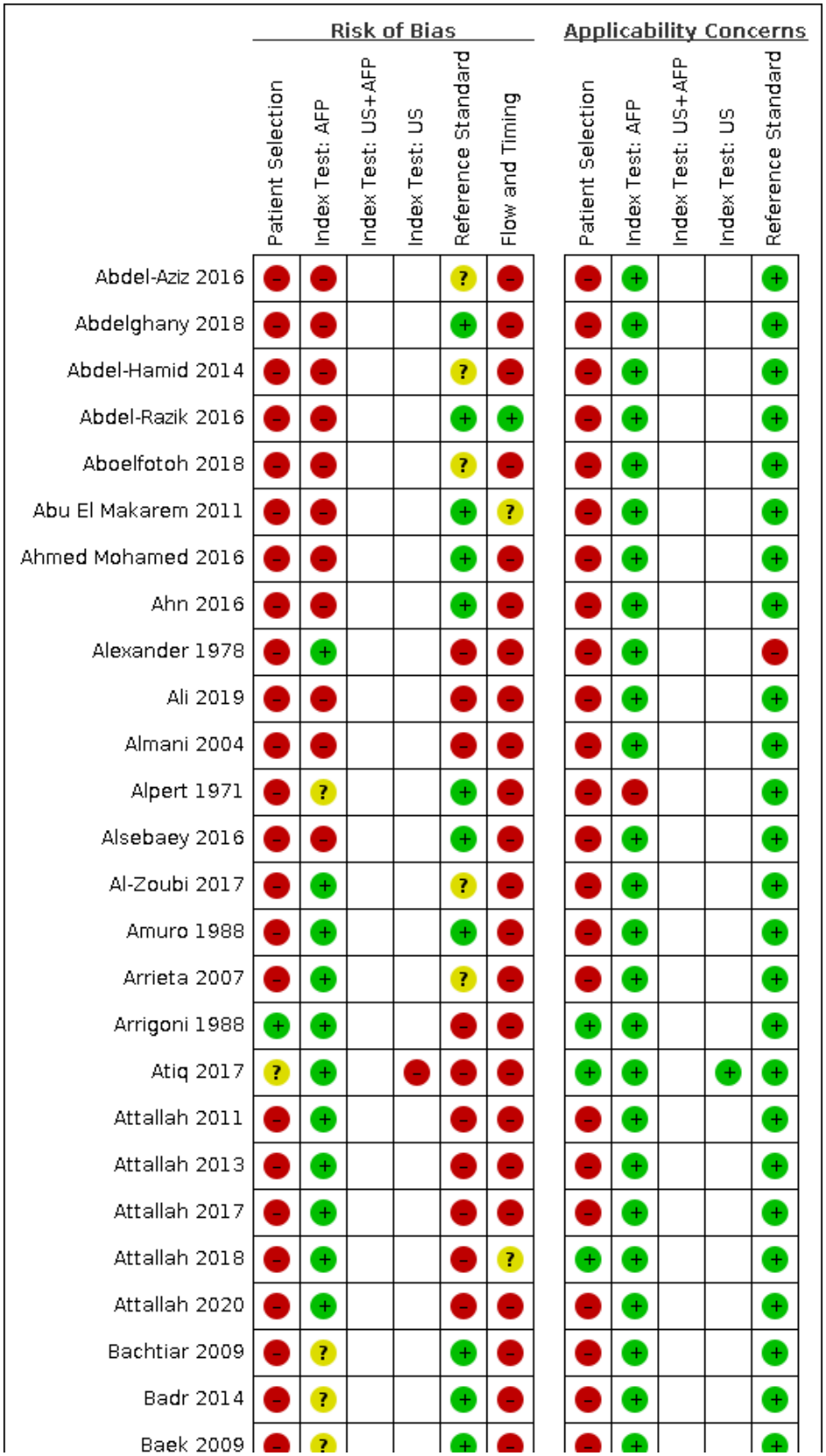


Figure 13. (Continued)

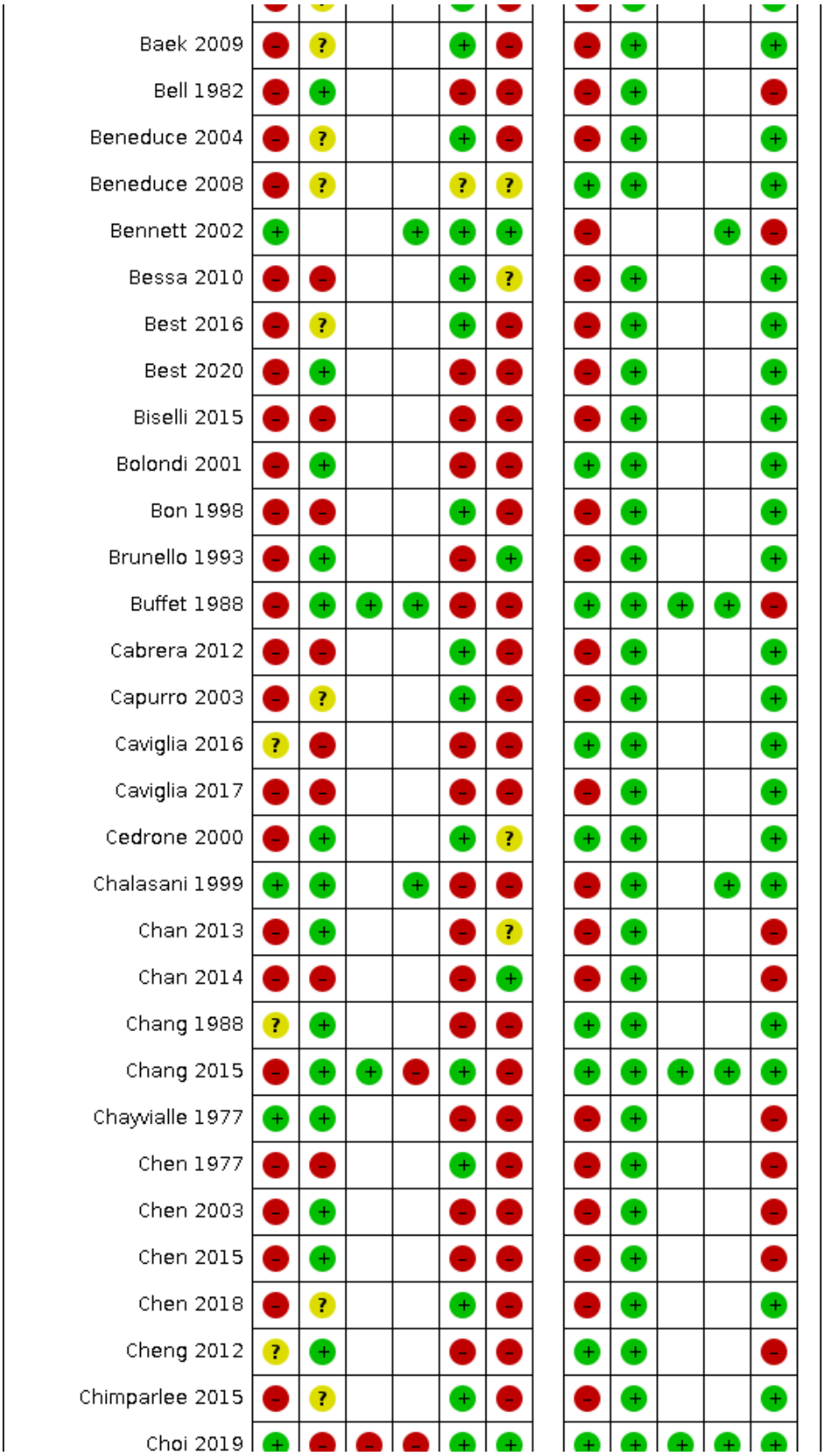


Figure 13. (Continued)

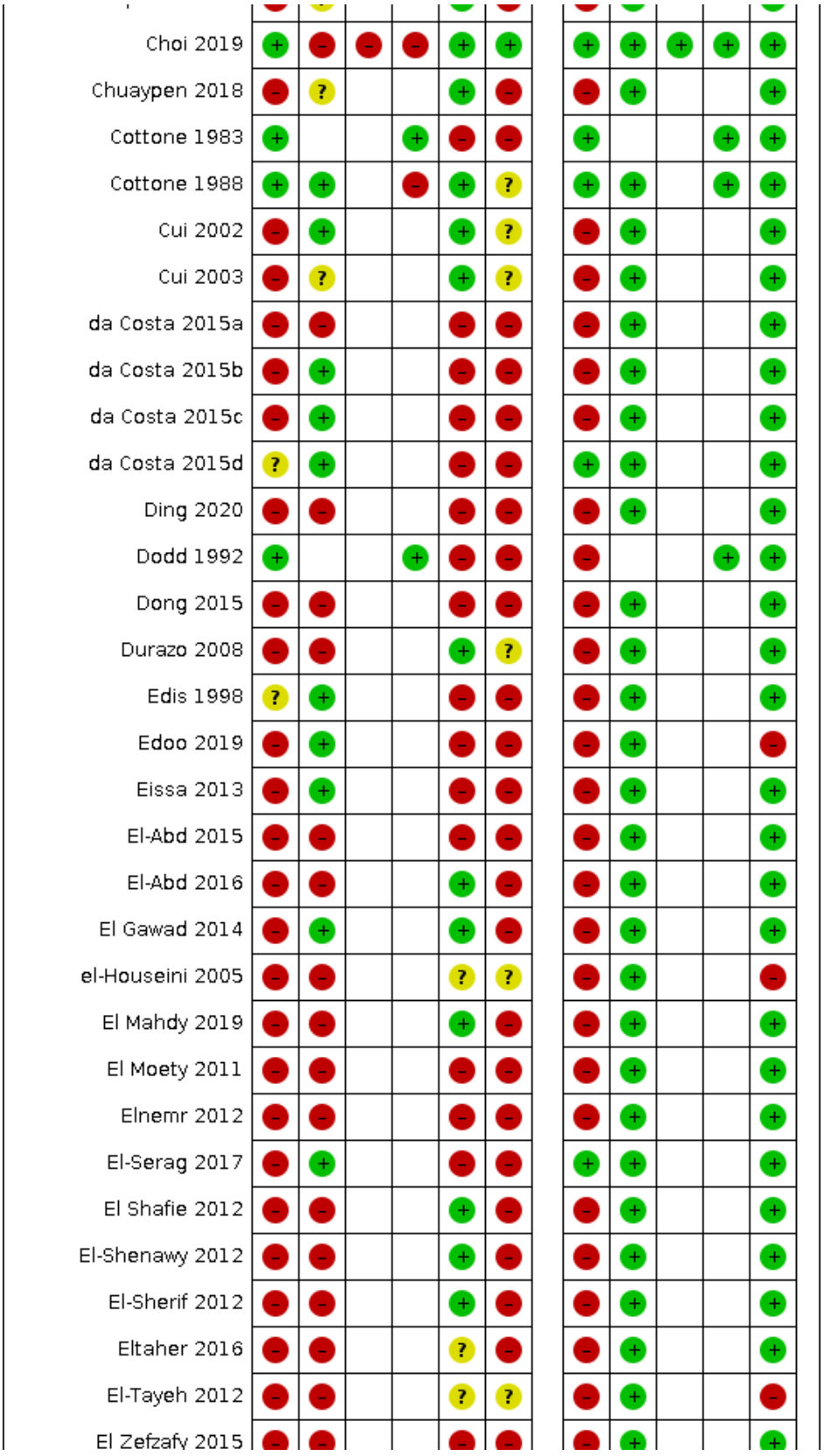


Figure 13. (Continued)

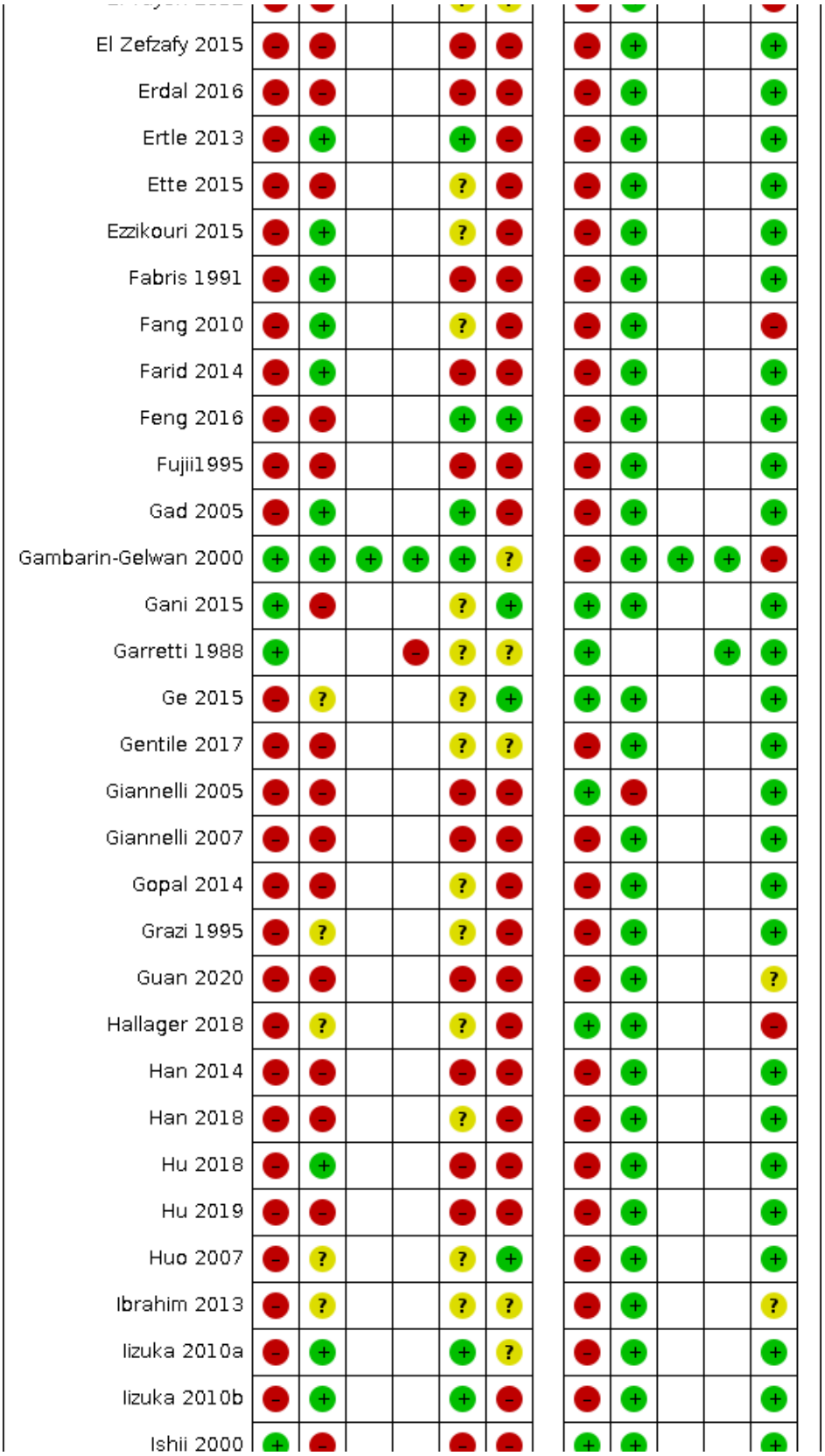


Figure 13. (Continued)

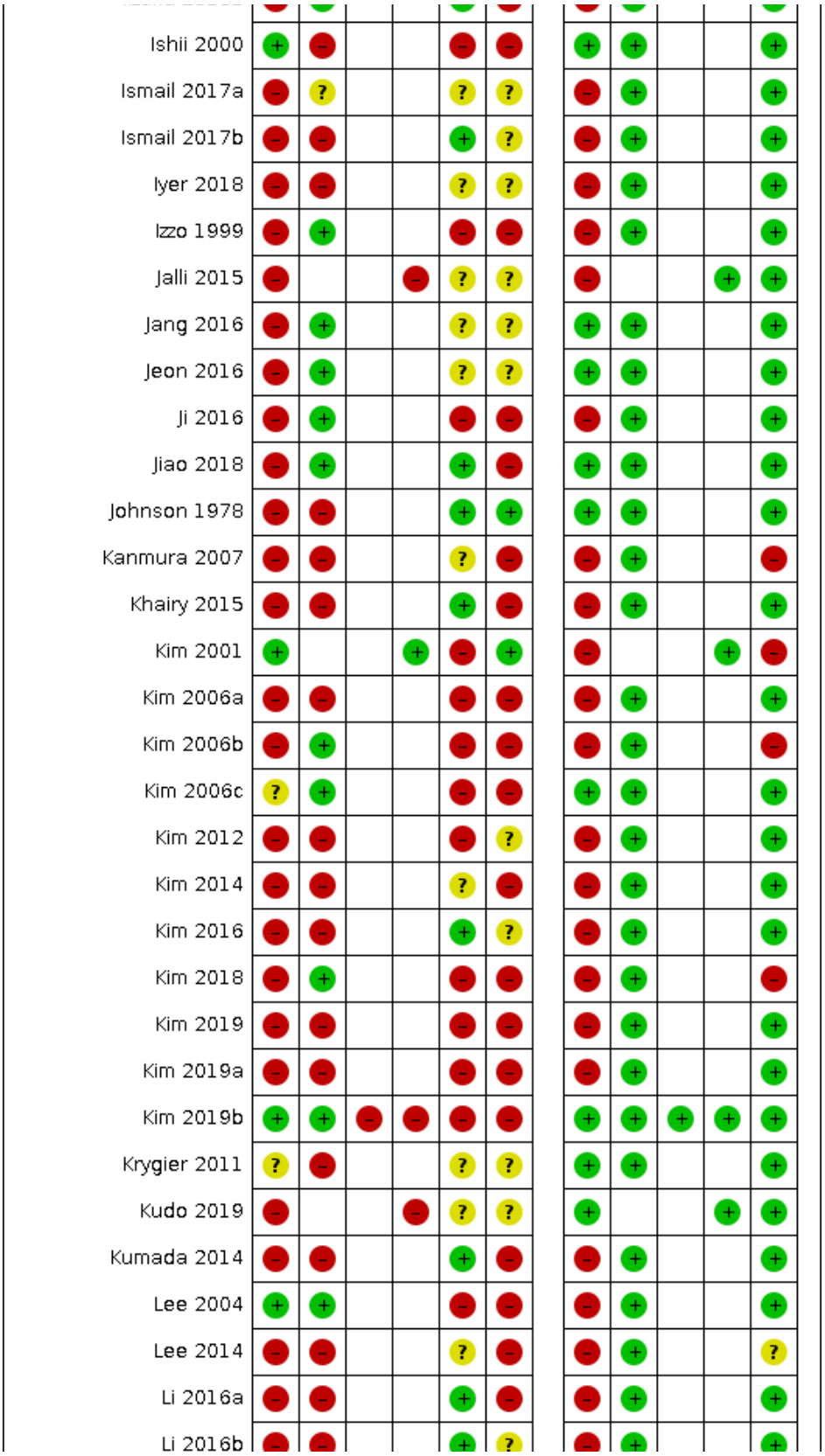


Figure 13. (Continued)

Figure 13. (Continued)

Figure 13. (Continued)

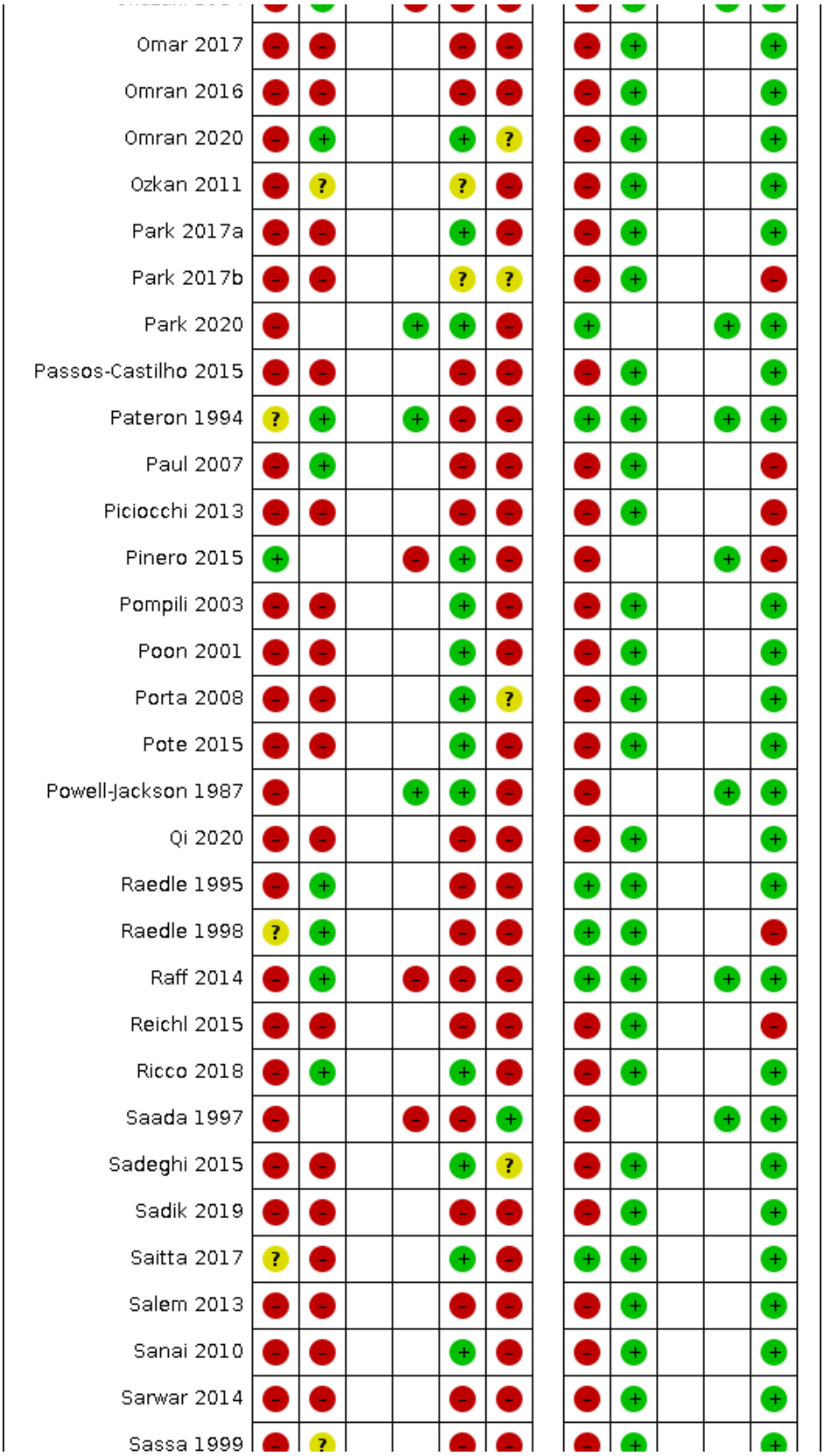


Figure 13. (Continued)

Figure 13. (Continued)

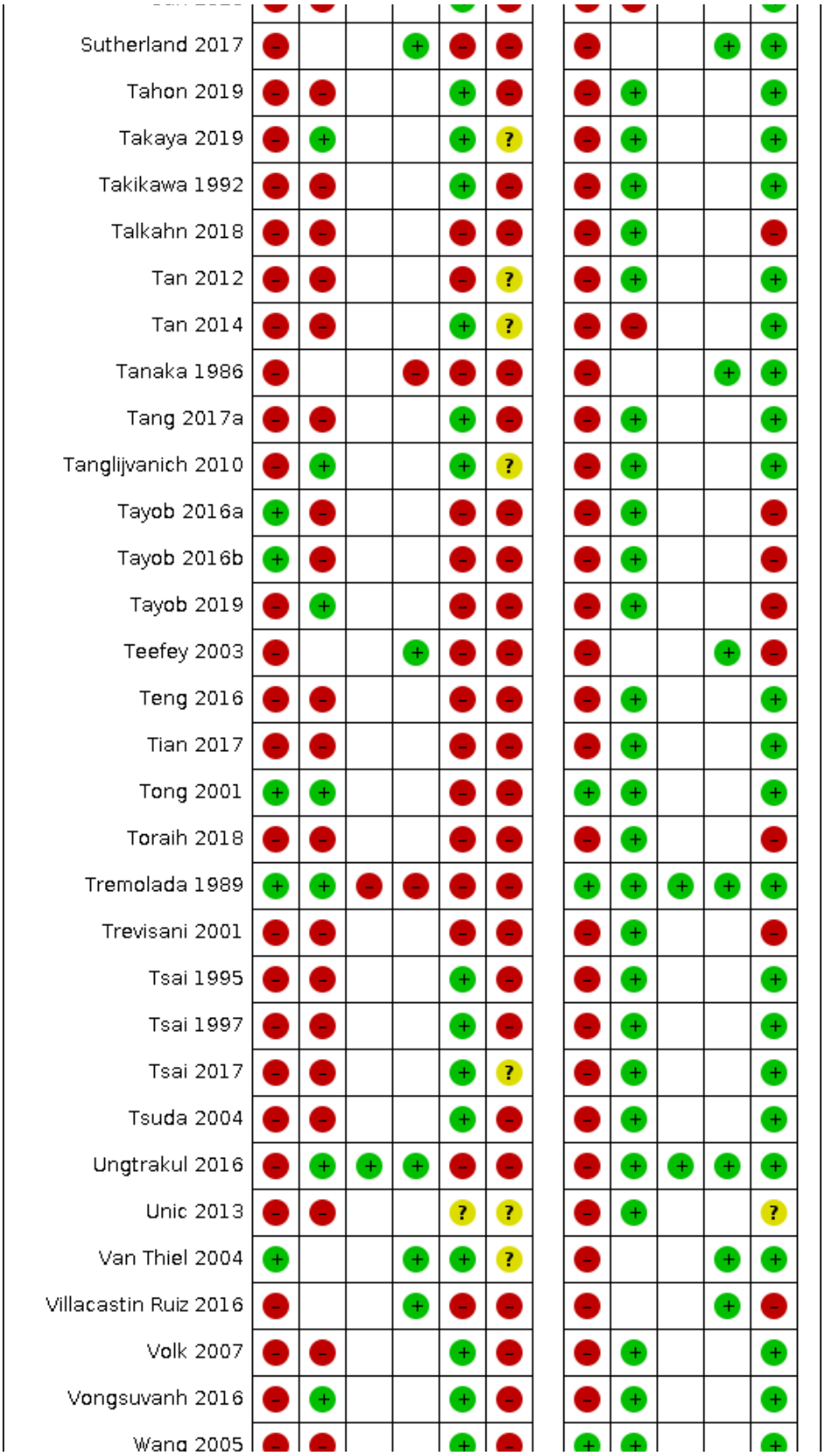


Figure 13. (Continued)

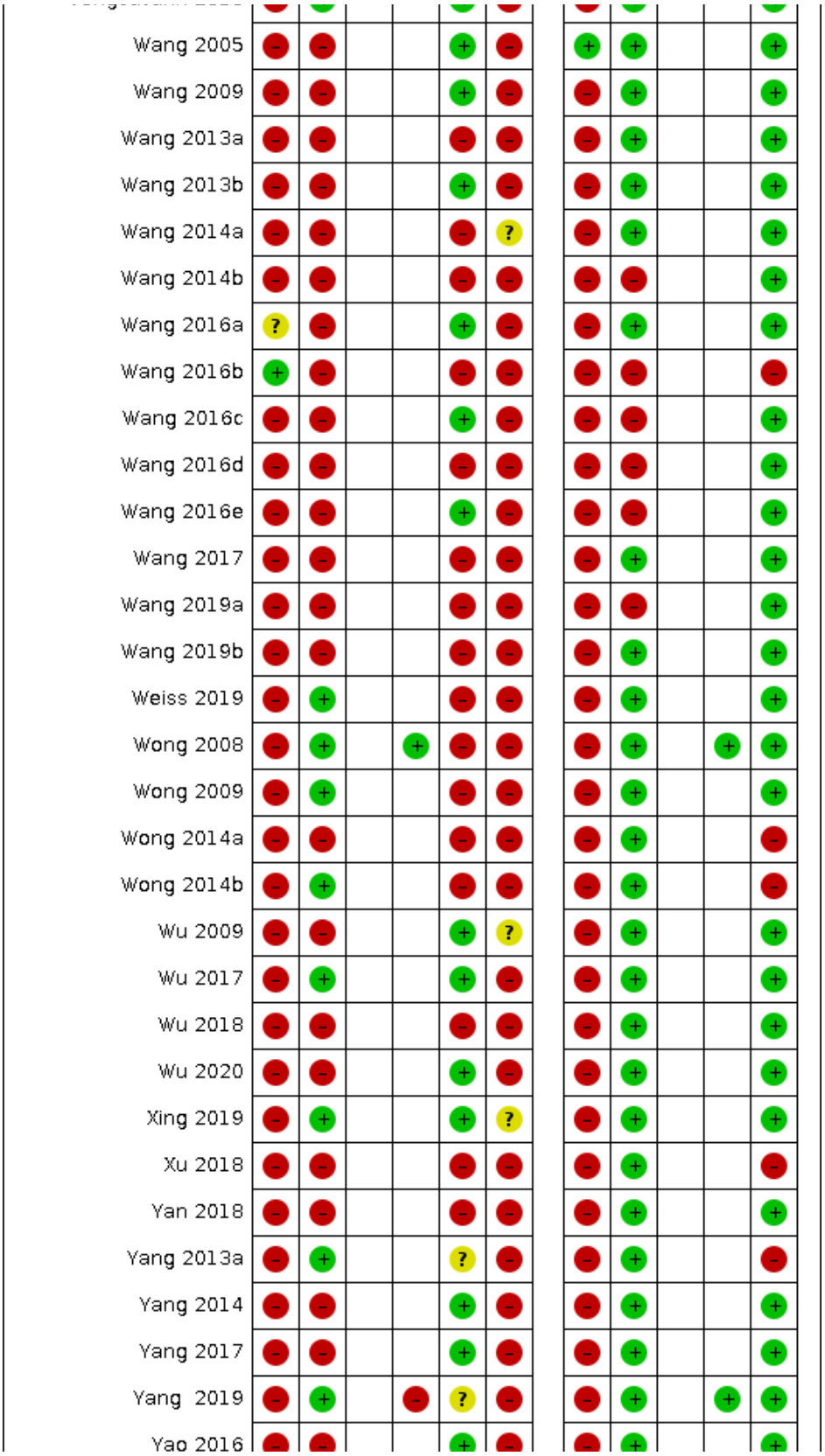


Figure 13. (Continued)

Appendix 4. Results of studies of alpha-fetoprotein with any cut-off value

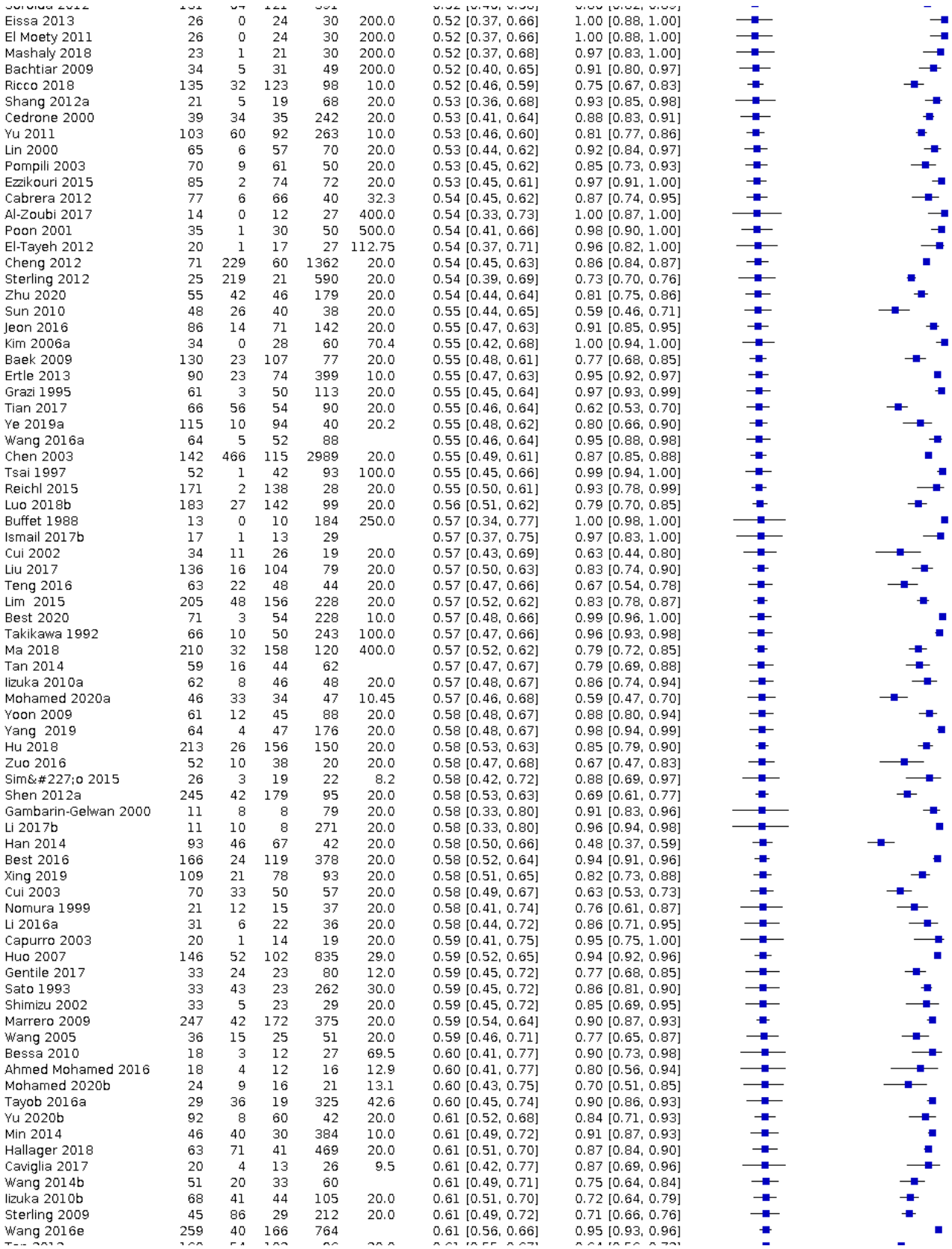
Figure 14

Figure 14. Forest plots of sensitivity and specificity of alpha-fetoprotein with any cut-off value against different reference standards in 326 studies ordered by increasing sensitivity. Reference standards were: the pathology of the explanted liver in case of transplantation.;the histology of resected focal liver lesions, or the histology of biopsied focal liver lesions with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity

and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line).

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Sassa 1999	5	0	56	134	200.0	0.08 [0.03, 0.18]	1.00 [0.97, 1.00]
Shaheen 2018	6	0	34	40	400.0	0.15 [0.06, 0.30]	1.00 [0.91, 1.00]
Nomura 2012	5	14	23	123	40.0	0.18 [0.06, 0.37]	0.90 [0.83, 0.94]
da Costa 2015d	5	18	16	76	20.0	0.24 [0.08, 0.47]	0.81 [0.71, 0.88]
El-Serag 2017	5	32	16	511	20.0	0.24 [0.08, 0.47]	0.94 [0.92, 0.96]
Weiss 2019	5	4	15	36	20.0	0.25 [0.09, 0.49]	0.90 [0.76, 0.97]
Li 2016c	27	30	77	274	17.85	0.26 [0.18, 0.35]	0.90 [0.86, 0.93]
Moriya 2013	4	13	11	170	20.0	0.27 [0.08, 0.55]	0.93 [0.88, 0.96]
Raff 2014	8	20	20	308	20.0	0.29 [0.13, 0.49]	0.94 [0.91, 0.96]
Zhou 2012	34	14	84	127	10.9	0.29 [0.21, 0.38]	0.90 [0.84, 0.94]
Omran 2020	30	0	74	92	400.0	0.29 [0.20, 0.39]	1.00 [0.96, 1.00]
Marrero 2005	43	6	101	146	99.0	0.30 [0.23, 0.38]	0.96 [0.92, 0.99]
Loglio 2019	11	1	24	222	7.0	0.31 [0.17, 0.49]	1.00 [0.98, 1.00]
el-Houseini 2005	14	5	30	15	19.8	0.32 [0.19, 0.48]	0.75 [0.51, 0.91]
Edis 1998	13	4	26	67	10.0	0.33 [0.19, 0.50]	0.94 [0.86, 0.98]
Atiq 2017	27	51	51	511	20.0	0.35 [0.24, 0.46]	0.91 [0.88, 0.93]
Liu 2020	38	5	67	49	400.0	0.36 [0.27, 0.46]	0.91 [0.80, 0.97]
Arrieta 2007	70	0	123	74	200.0	0.36 [0.29, 0.43]	1.00 [0.95, 1.00]
Yang 2013a	66	12	113	68	20.0	0.37 [0.30, 0.44]	0.85 [0.75, 0.92]
Li 2016b	24	17	41	88	17.85	0.37 [0.25, 0.50]	0.84 [0.75, 0.90]
Passos-Castilho 2015	12	2	20	28	20.0	0.38 [0.21, 0.56]	0.93 [0.78, 0.99]
Zhu 2013	95	18	157	437	200.0	0.38 [0.32, 0.44]	0.96 [0.94, 0.98]
Wang 2016d	43	29	70	557		0.38 [0.29, 0.48]	0.95 [0.93, 0.97]
Oka 1994	21	48	34	150	20.0	0.38 [0.25, 0.52]	0.76 [0.69, 0.82]
Attallah 2018	43	2	67	70	200.0	0.39 [0.30, 0.49]	0.97 [0.90, 1.00]
Attallah 2011	59	0	91	100	200.0	0.39 [0.31, 0.48]	1.00 [0.96, 1.00]
Wang 2013a	19	6	29	34	20.0	0.40 [0.26, 0.55]	0.85 [0.70, 0.94]
Omran 2016	21	1	32	19	400.0	0.40 [0.26, 0.54]	0.95 [0.75, 1.00]
Wu 2018	57	10	86	70	209.2	0.40 [0.32, 0.48]	0.88 [0.78, 0.94]
Attallah 2020	59	0	89	133	400.0	0.40 [0.32, 0.48]	1.00 [0.97, 1.00]
Sadik 2019	12	7	18	24	38.1	0.40 [0.23, 0.59]	0.77 [0.59, 0.90]
Shaheen 2015	16	1	24	29	88.5	0.40 [0.25, 0.57]	0.97 [0.83, 1.00]
Giannelli 2007	203	29	296	433	24.81	0.41 [0.36, 0.45]	0.94 [0.91, 0.96]
Attallah 2013	93	3	134	338	400.0	0.41 [0.35, 0.48]	0.99 [0.97, 1.00]
Bolondi 2001	25	46	36	206	20.0	0.41 [0.29, 0.54]	0.82 [0.76, 0.86]
Ungtrakul 2016	7	29	10	2247	20.0	0.41 [0.18, 0.67]	0.99 [0.98, 0.99]
Kumada 2014	43	10	61	94	20.0	0.41 [0.32, 0.51]	0.90 [0.83, 0.95]
Kanmura 2007	12	11	17	22	20.0	0.41 [0.24, 0.61]	0.67 [0.48, 0.82]
Wang 2016c	181	22	251	416		0.42 [0.37, 0.47]	0.95 [0.92, 0.97]
Amuro 1988	22	1	30	41	400.0	0.42 [0.29, 0.57]	0.98 [0.87, 1.00]
Wang 2016b	21	5	28	97		0.43 [0.29, 0.58]	0.95 [0.89, 0.98]
Zhou 2019	27	11	36	152	18.5	0.43 [0.30, 0.56]	0.93 [0.88, 0.97]
Vongsuvanh 2016	37	6	49	166	20.0	0.43 [0.32, 0.54]	0.97 [0.93, 0.99]
Beneduce 2004	26	9	34	41	20.0	0.43 [0.31, 0.57]	0.82 [0.69, 0.91]
Brunello 1993	17	1	22	15	20.0	0.44 [0.28, 0.60]	0.94 [0.70, 1.00]
Loglio 2018	28	0	36	148	7.0	0.44 [0.31, 0.57]	1.00 [0.98, 1.00]
Attallah 2017	140	0	178	341	400.0	0.44 [0.38, 0.50]	1.00 [0.99, 1.00]
Giannelli 2005	54	13	66	77	13.7	0.45 [0.36, 0.54]	0.86 [0.77, 0.92]
Liu 2018	36	11	44	71	200.0	0.45 [0.34, 0.57]	0.87 [0.77, 0.93]
Mao 2017	37	13	45	44	20.0	0.45 [0.34, 0.57]	0.77 [0.64, 0.87]
Piciocchi 2013	30	13	36	63	14.0	0.45 [0.33, 0.58]	0.83 [0.73, 0.91]
Yan 2018	11	4	13	58	80.5	0.46 [0.26, 0.67]	0.94 [0.84, 0.98]
Shariff 2016	6	1	7	24	24.2	0.46 [0.19, 0.75]	0.96 [0.80, 1.00]
El Zefzafy 2015	14	1	16	29	134.0	0.47 [0.28, 0.66]	0.97 [0.83, 1.00]
Song 2002	18	8	20	23	20.0	0.47 [0.31, 0.64]	0.74 [0.55, 0.88]
Sadeghi 2015	29	9	32	69	24.2	0.48 [0.35, 0.61]	0.88 [0.79, 0.95]
Beneduce 2008	16	4	17	27	20.0	0.48 [0.31, 0.66]	0.87 [0.70, 0.96]
Nabih 2014	17	3	18	31	240.0	0.49 [0.31, 0.66]	0.91 [0.76, 0.98]
Maringhini 1988	71	0	75	217	500.0	0.49 [0.40, 0.57]	1.00 [0.98, 1.00]
Chen 2018	99	70	103	371	20.0	0.49 [0.42, 0.56]	0.84 [0.80, 0.87]
da Costa 2015c	37	10	38	65	20.0	0.49 [0.38, 0.61]	0.87 [0.77, 0.93]
Kim 2019b	32	7	32	321		0.50 [0.37, 0.63]	0.98 [0.96, 0.99]
Pateron 1994	7	15	7	89	15.0	0.50 [0.23, 0.77]	0.86 [0.77, 0.92]
Luo 2018c	78	38	77	105	20.0	0.50 [0.42, 0.58]	0.73 [0.65, 0.80]
Alpert 1971	59	0	58	337	1000.0	0.50 [0.41, 0.60]	1.00 [0.99, 1.00]
Tayob 2019	1744	3363	1712	30264	400.0	0.50 [0.49, 0.52]	0.90 [0.90, 0.90]
Guan 2020	154	12	148	152	55.87	0.51 [0.45, 0.57]	0.93 [0.88, 0.96]
Song 2020a	51	5	49	62	10.0	0.51 [0.41, 0.61]	0.93 [0.83, 0.98]
Jiao 2018	92	2	88	158	400.0	0.51 [0.44, 0.59]	0.99 [0.96, 1.00]
Tsai 2017	253	58	240	435	20.0	0.51 [0.47, 0.56]	0.88 [0.85, 0.91]
Minami 2015b	15	4	14	54	17.0	0.52 [0.33, 0.71]	0.93 [0.83, 0.98]
Tsuda 2004	29	11	27	21	20.0	0.52 [0.38, 0.65]	0.66 [0.47, 0.81]
Soroida 2012	131	64	121	391		0.52 [0.46, 0.58]	0.86 [0.82, 0.89]
Eissa 2013	26	0	24	30	200.0	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
El Moety 2011	26	0	24	30	200.0	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]

Figure 14. (Continued)



Eissa 2013	26	0	24	30	200.0	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
El Moety 2011	26	0	24	30	200.0	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
Mashaly 2018	23	1	21	30	200.0	0.52 [0.37, 0.68]	0.97 [0.83, 1.00]
Bachtiar 2009	34	5	31	49	200.0	0.52 [0.40, 0.65]	0.91 [0.80, 0.97]
Ricco 2018	135	32	123	98	10.0	0.52 [0.46, 0.59]	0.75 [0.67, 0.83]
Shang 2012a	21	5	19	68	20.0	0.53 [0.36, 0.68]	0.93 [0.85, 0.98]
Cedrone 2000	39	34	35	242	20.0	0.53 [0.41, 0.64]	0.88 [0.83, 0.91]
Yu 2011	103	60	92	263	10.0	0.53 [0.46, 0.60]	0.81 [0.77, 0.86]
Lin 2000	65	6	57	70	20.0	0.53 [0.44, 0.62]	0.92 [0.84, 0.97]
Pompili 2003	70	9	61	50	20.0	0.53 [0.45, 0.62]	0.85 [0.73, 0.93]
Ezzikouri 2015	85	2	74	72	20.0	0.53 [0.45, 0.61]	0.97 [0.91, 1.00]
Cabrera 2012	77	6	66	40	32.3	0.54 [0.45, 0.62]	0.87 [0.74, 0.95]
Al-Zoubi 2017	14	0	12	27	400.0	0.54 [0.33, 0.73]	1.00 [0.87, 1.00]
Poon 2001	35	1	30	50	500.0	0.54 [0.41, 0.66]	0.98 [0.90, 1.00]
El-Tayeh 2012	20	1	17	27	112.75	0.54 [0.37, 0.71]	0.96 [0.82, 1.00]
Cheng 2012	71	229	60	1362	20.0	0.54 [0.45, 0.63]	0.86 [0.84, 0.87]
Sterling 2012	25	219	21	590	20.0	0.54 [0.39, 0.69]	0.73 [0.70, 0.76]
Zhu 2020	55	42	46	179	20.0	0.54 [0.44, 0.64]	0.81 [0.75, 0.86]
Sun 2010	48	26	40	38	20.0	0.55 [0.44, 0.65]	0.59 [0.46, 0.71]
Jeon 2016	86	14	71	142	20.0	0.55 [0.47, 0.63]	0.91 [0.85, 0.95]
Kim 2006a	34	0	28	60	70.4	0.55 [0.42, 0.68]	1.00 [0.94, 1.00]
Baek 2009	130	23	107	77	20.0	0.55 [0.48, 0.61]	0.77 [0.68, 0.85]
Ertle 2013	90	23	74	399	10.0	0.55 [0.47, 0.63]	0.95 [0.92, 0.97]
Grazi 1995	61	3	50	113	20.0	0.55 [0.45, 0.64]	0.97 [0.93, 0.99]
Tian 2017	66	56	54	90	20.0	0.55 [0.46, 0.64]	0.62 [0.53, 0.70]
Ye 2019a	115	10	94	40	20.2	0.55 [0.48, 0.62]	0.80 [0.66, 0.90]
Wang 2016a	64	5	52	88		0.55 [0.46, 0.64]	0.95 [0.88, 0.98]
Chen 2003	142	466	115	2989	20.0	0.55 [0.49, 0.61]	0.87 [0.85, 0.88]
Tsai 1997	52	1	42	93	100.0	0.55 [0.45, 0.66]	0.99 [0.94, 1.00]
Reichl 2015	171	2	138	28	20.0	0.55 [0.50, 0.61]	0.93 [0.78, 0.99]
Luo 2018b	183	27	142	99	20.0	0.56 [0.51, 0.62]	0.79 [0.70, 0.85]
Buffet 1988	13	0	10	184	250.0	0.57 [0.34, 0.77]	1.00 [0.98, 1.00]
Ismail 2017b	17	1	13	29		0.57 [0.37, 0.75]	0.97 [0.83, 1.00]
Cui 2002	34	11	26	19	20.0	0.57 [0.43, 0.69]	0.63 [0.44, 0.80]
Liu 2017	136	16	104	79	20.0	0.57 [0.50, 0.63]	0.83 [0.74, 0.90]
Teng 2016	63	22	48	44	20.0	0.57 [0.47, 0.66]	0.67 [0.54, 0.78]
Lim 2015	205	48	156	228	20.0	0.57 [0.52, 0.62]	0.83 [0.78, 0.87]
Best 2020	71	3	54	228	10.0	0.57 [0.48, 0.66]	0.99 [0.96, 1.00]
Takikawa 1992	66	10	50	243	100.0	0.57 [0.47, 0.66]	0.96 [0.93, 0.98]
Ma 2018	210	32	158	120	400.0	0.57 [0.52, 0.62]	0.79 [0.72, 0.85]
Tan 2014	59	16	44	62		0.57 [0.47, 0.67]	0.79 [0.69, 0.88]
Iizuka 2010a	62	8	46	48	20.0	0.57 [0.48, 0.67]	0.86 [0.74, 0.94]
Mohamed 2020a	46	33	34	47	10.45	0.57 [0.46, 0.68]	0.59 [0.47, 0.70]
Yoon 2009	61	12	45	88	20.0	0.58 [0.48, 0.67]	0.88 [0.80, 0.94]
Yang 2019	64	4	47	176	20.0	0.58 [0.48, 0.67]	0.98 [0.94, 0.99]
Hu 2018	213	26	156	150	20.0	0.58 [0.53, 0.63]	0.85 [0.79, 0.90]
Zuo 2016	52	10	38	20	20.0	0.58 [0.47, 0.68]	0.67 [0.47, 0.83]
Simão 2015	26	3	19	22	8.2	0.58 [0.42, 0.72]	0.88 [0.69, 0.97]
Shen 2012a	245	42	179	95	20.0	0.58 [0.53, 0.63]	0.69 [0.61, 0.77]
Gambarin-Gelwan 2000	11	8	8	79	20.0	0.58 [0.33, 0.80]	0.91 [0.83, 0.96]
Li 2017b	11	10	8	271	20.0	0.58 [0.33, 0.80]	0.96 [0.94, 0.98]
Han 2014	93	46	67	42	20.0	0.58 [0.50, 0.66]	0.48 [0.37, 0.59]
Best 2016	166	24	119	378	20.0	0.58 [0.52, 0.64]	0.94 [0.91, 0.96]
Xing 2019	109	21	78	93	20.0	0.58 [0.51, 0.65]	0.82 [0.73, 0.88]
Cui 2003	70	33	50	57	20.0	0.58 [0.49, 0.67]	0.63 [0.53, 0.73]
Nomura 1999	21	12	15	37	20.0	0.58 [0.41, 0.74]	0.76 [0.61, 0.87]
Li 2016a	31	6	22	36	20.0	0.58 [0.44, 0.72]	0.86 [0.71, 0.95]
Capurro 2003	20	1	14	19	20.0	0.59 [0.41, 0.75]	0.95 [0.75, 1.00]
Huo 2007	146	52	102	835	29.0	0.59 [0.52, 0.65]	0.94 [0.92, 0.96]
Gentile 2017	33	24	23	80	12.0	0.59 [0.45, 0.72]	0.77 [0.68, 0.85]
Sato 1993	33	43	23	262	30.0	0.59 [0.45, 0.72]	0.86 [0.81, 0.90]
Shimizu 2002	33	5	23	29	20.0	0.59 [0.45, 0.72]	0.85 [0.69, 0.95]
Marrero 2009	247	42	172	375	20.0	0.59 [0.54, 0.64]	0.90 [0.87, 0.93]
Wang 2005	36	15	25	51	20.0	0.59 [0.46, 0.71]	0.77 [0.65, 0.87]
Bessa 2010	18	3	12	27	69.5	0.60 [0.41, 0.77]	0.90 [0.73, 0.98]
Ahmed Mohamed 2016	18	4	12	16	12.9	0.60 [0.41, 0.77]	0.80 [0.56, 0.94]
Mohamed 2020b	24	9	16	21	13.1	0.60 [0.43, 0.75]	0.70 [0.51, 0.85]
Tayob 2016a	29	36	19	325	42.6	0.60 [0.45, 0.74]	0.90 [0.86, 0.93]
Yu 2020b	92	8	60	42	20.0	0.61 [0.52, 0.68]	0.84 [0.71, 0.93]
Min 2014	46	40	30	384	10.0	0.61 [0.49, 0.72]	0.91 [0.87, 0.93]
Hallager 2018	63	71	41	469	20.0	0.61 [0.51, 0.70]	0.87 [0.84, 0.90]
Caviglia 2017	20	4	13	26	9.5	0.61 [0.42, 0.77]	0.87 [0.69, 0.96]
Wang 2014b	51	20	33	60		0.61 [0.49, 0.71]	0.75 [0.64, 0.84]
Iizuka 2010b	68	41	44	105	20.0	0.61 [0.51, 0.70]	0.72 [0.64, 0.79]
Sterling 2009	45	86	29	212	20.0	0.61 [0.49, 0.72]	0.71 [0.66, 0.76]
Wang 2016e	259	40	166	764		0.61 [0.56, 0.66]	0.95 [0.93, 0.96]

Figure 14. (Continued)

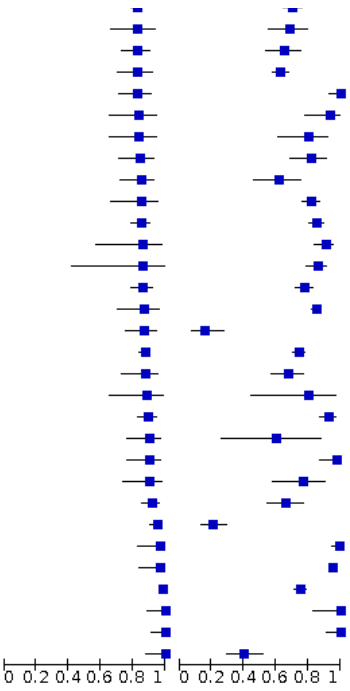
Sterling 2009	45	86	29	212	20.0	0.61 [0.49, 0.72]	0.71 [0.66, 0.76]
Wang 2016e	259	40	166	764		0.61 [0.56, 0.66]	0.95 [0.93, 0.96]
Tan 2012	160	54	102	96	20.0	0.61 [0.55, 0.67]	0.64 [0.56, 0.72]
Luo 2018a	22	18	14	23	20.0	0.61 [0.43, 0.77]	0.56 [0.40, 0.72]
Pote 2015	52	22	33	22	5.5	0.61 [0.50, 0.72]	0.50 [0.35, 0.65]
Wu 2017	225	8	141	24	20.0	0.61 [0.56, 0.66]	0.75 [0.57, 0.89]
Ahn 2016	225	76	141	290	10.8	0.61 [0.56, 0.66]	0.79 [0.75, 0.83]
Lok 2010	24	15	15	62	20.0	0.62 [0.45, 0.77]	0.81 [0.70, 0.89]
Wu 2020	122	34	76	92	20.0	0.62 [0.54, 0.68]	0.73 [0.64, 0.81]
Choi 2019	26	22	16	146	5.0	0.62 [0.46, 0.76]	0.87 [0.81, 0.92]
Kim 2006c	26	26	16	159	20.0	0.62 [0.46, 0.76]	0.86 [0.80, 0.91]
Snowberger 2007	148	339	91	1357	8.9	0.62 [0.55, 0.68]	0.80 [0.78, 0.82]
Ji 2016	124	30	76	67	20.0	0.62 [0.55, 0.69]	0.69 [0.59, 0.78]
Nakamura 2006	844	24	517	324	20.0	0.62 [0.59, 0.65]	0.93 [0.90, 0.96]
Jang 2016	129	19	79	174	20.0	0.62 [0.55, 0.69]	0.90 [0.85, 0.94]
Ishii 2000	18	153	11	552	20.0	0.62 [0.42, 0.79]	0.78 [0.75, 0.81]
Dong 2015	118	62	72	52	20.0	0.62 [0.55, 0.69]	0.46 [0.36, 0.55]
Trevisani 2001	106	18	64	152	16.0	0.62 [0.55, 0.70]	0.89 [0.84, 0.94]
Wang 2014a	25	7	15	27	20.0	0.63 [0.46, 0.77]	0.79 [0.62, 0.91]
Zhan 2020	30	4	18	61	343.3	0.63 [0.47, 0.76]	0.94 [0.85, 0.98]
Ette 2015	39	4	23	53	18.0	0.63 [0.50, 0.75]	0.93 [0.83, 0.98]
Lee 2004	34	91	20	108	20.0	0.63 [0.49, 0.76]	0.54 [0.47, 0.61]
Nomura 1996	17	17	10	84	20.0	0.63 [0.42, 0.81]	0.83 [0.74, 0.90]
Sultanik 2017	29	21	17	95	20.0	0.63 [0.48, 0.77]	0.82 [0.74, 0.88]
Na 2013	36	13	21	51	20.0	0.63 [0.49, 0.76]	0.80 [0.68, 0.89]
Nguyen 2002	103	30	60	119	20.0	0.63 [0.55, 0.71]	0.80 [0.73, 0.86]
Abdel-Aziz 2016	43	8	25	10	120.5	0.63 [0.51, 0.75]	0.56 [0.31, 0.78]
Edoo 2019	680	46	395	191	20.0	0.63 [0.60, 0.66]	0.81 [0.75, 0.85]
El-Sherif 2012	19	3	11	27	19.96	0.63 [0.44, 0.80]	0.90 [0.73, 0.98]
Porta 2008	19	7	11	23	16.94	0.63 [0.44, 0.80]	0.77 [0.58, 0.90]
Zheng 2017	211	39	122	125	30.5	0.63 [0.58, 0.69]	0.76 [0.69, 0.83]
Chalasani 1999	14	34	8	229	20.0	0.64 [0.41, 0.83]	0.87 [0.82, 0.91]
Lee 2014	77	2	43	38	6.0	0.64 [0.55, 0.73]	0.95 [0.83, 0.99]
Yang 2014	79	9	44	48	20.0	0.64 [0.55, 0.73]	0.84 [0.72, 0.93]
El-Shenawy 2012	9	48	5	41	102.0	0.64 [0.35, 0.87]	0.46 [0.35, 0.57]
Sherman 1995	9	91	5	964	20.0	0.64 [0.35, 0.87]	0.91 [0.90, 0.93]
Kim 2016	625	392	345	712	6.8	0.64 [0.61, 0.67]	0.64 [0.62, 0.67]
Yang 2017	20	48	11	92	9.9	0.65 [0.45, 0.81]	0.66 [0.57, 0.74]
Park 2017b	626	392	344	712	6.8	0.65 [0.61, 0.68]	0.64 [0.62, 0.67]
Wang 2017	73	43	40	118	17.56	0.65 [0.55, 0.73]	0.73 [0.66, 0.80]
Chuaypen 2018	97	14	53	136	20.0	0.65 [0.56, 0.72]	0.91 [0.85, 0.95]
Wang 2019a	114	82	62	277	145.65	0.65 [0.57, 0.72]	0.77 [0.72, 0.81]
Wong 2009	24	0	13	37	20.0	0.65 [0.47, 0.80]	1.00 [0.91, 1.00]
Abdel-Hamid 2014	13	1	7	39	121.0	0.65 [0.41, 0.85]	0.97 [0.87, 1.00]
Farid 2014	13	1	7	9	10.0	0.65 [0.41, 0.85]	0.90 [0.55, 1.00]
Matievskaya 2003	13	2	7	137	200.0	0.65 [0.41, 0.85]	0.99 [0.95, 1.00]
Yu 2020c	189	11	101	69	20.0	0.65 [0.59, 0.71]	0.86 [0.77, 0.93]
Lin 2016	17	23	9	27	20.0	0.65 [0.44, 0.83]	0.54 [0.39, 0.68]
Krygier 2011	19	26	10	63	13.6	0.66 [0.46, 0.82]	0.71 [0.60, 0.80]
Singal 2012	27	38	14	363	20.0	0.66 [0.49, 0.80]	0.91 [0.87, 0.93]
Ye 2019b	135	11	69	49	20.2	0.66 [0.59, 0.73]	0.82 [0.70, 0.90]
Song 2020b	53	61	27	99	21.48	0.66 [0.55, 0.76]	0.62 [0.54, 0.69]
Biselli 2015	77	43	39	225	10.0	0.66 [0.57, 0.75]	0.84 [0.79, 0.88]
Chayvialle 1977	12	3	6	179	7.7	0.67 [0.41, 0.87]	0.98 [0.95, 1.00]
Cottone 1988	10	39	5	104	20.0	0.67 [0.38, 0.88]	0.73 [0.65, 0.80]
Kim 2014	20	7	10	28	20.0	0.67 [0.47, 0.83]	0.80 [0.63, 0.92]
Kim 2018	36	4	18	22	20.0	0.67 [0.53, 0.79]	0.85 [0.65, 0.96]
Chimparlee 2015	105	4	52	73	20.0	0.67 [0.59, 0.74]	0.95 [0.87, 0.99]
Shen 2012b	140	82	69	45	20.0	0.67 [0.60, 0.73]	0.35 [0.27, 0.44]
Tsai 1995	68	0	33	101	120.0	0.67 [0.57, 0.76]	1.00 [0.96, 1.00]
Liu 2019	112	52	54	133	31.2	0.67 [0.60, 0.75]	0.72 [0.65, 0.78]
Tahon 2019	27	12	13	18	6.95	0.68 [0.51, 0.81]	0.60 [0.41, 0.77]
Talkahn 2018	27	0	13	30	220.0	0.68 [0.51, 0.81]	1.00 [0.88, 1.00]
Saitta 2017	27	16	13	34	6.5	0.68 [0.51, 0.81]	0.68 [0.53, 0.80]
Seo 2015	106	107	51	991	10.0	0.68 [0.60, 0.75]	0.90 [0.88, 0.92]
Bon 1998	25	13	12	10	20.0	0.68 [0.50, 0.82]	0.43 [0.23, 0.66]
Li 2017a	23	23	11	52	10.28	0.68 [0.49, 0.83]	0.69 [0.58, 0.79]
Song 2014	374	3	176	82	21.0	0.68 [0.64, 0.72]	0.96 [0.90, 0.99]
Ozkan 2011	51	3	24	52	13.0	0.68 [0.56, 0.78]	0.95 [0.85, 0.99]
Omar 2017	878	364	413	708	11.9	0.68 [0.65, 0.71]	0.66 [0.63, 0.69]
Ismail 2017a	45	12	21	87	20.0	0.68 [0.56, 0.79]	0.88 [0.80, 0.94]
Takaya 2019	28	9	13	11	10.0	0.68 [0.52, 0.82]	0.55 [0.32, 0.77]
Yu 2020a	108	8	50	42	20.0	0.68 [0.60, 0.76]	0.84 [0.71, 0.93]
Park 2017a	54	14	25	63	10.0	0.68 [0.57, 0.78]	0.82 [0.71, 0.90]
Sanai 2010	141	49	65	150	11.7	0.68 [0.62, 0.75]	0.75 [0.69, 0.81]
Long 2011	76	13	35	56	20.0	0.68 [0.59, 0.77]	0.81 [0.70, 0.90]
Oka 2001	266	108	122	104	20.0	0.69 [0.64, 0.73]	0.49 [0.42, 0.56]

Figure 14. (Continued)

Sanai 2010	141	49	65	150	11.7	0.68 [0.62, 0.75]	0.75 [0.69, 0.81]
Long 2011	76	13	35	56	20.0	0.68 [0.59, 0.77]	0.81 [0.70, 0.90]
Oka 2001	266	108	122	104	20.0	0.69 [0.64, 0.73]	0.49 [0.42, 0.56]
Unic 2013	22	0	10	28	14.62	0.69 [0.50, 0.84]	1.00 [0.88, 1.00]
Durazo 2008	99	13	45	83	25.0	0.69 [0.61, 0.76]	0.86 [0.78, 0.93]
Miura 2007	44	16	20	24		0.69 [0.56, 0.80]	0.60 [0.43, 0.75]
Chang 2015	250	212	113	1022	9.0	0.69 [0.64, 0.74]	0.83 [0.81, 0.85]
Volk 2007	58	15	26	154	23.0	0.69 [0.58, 0.79]	0.91 [0.86, 0.95]
Han 2018	58	19	26	84	11.7	0.69 [0.58, 0.79]	0.82 [0.73, 0.89]
Xu 2018	61	16	27	119	20.0	0.69 [0.59, 0.79]	0.88 [0.81, 0.93]
Raedle 1998	52	52	23	584	20.0	0.69 [0.58, 0.79]	0.92 [0.89, 0.94]
Lin 2015	75	10	33	88	20.0	0.69 [0.60, 0.78]	0.90 [0.82, 0.95]
Shu 2010	113	44	49	86	20.0	0.70 [0.62, 0.77]	0.66 [0.57, 0.74]
Sun 2020	28	46	12	60		0.70 [0.53, 0.83]	0.57 [0.47, 0.66]
Ibrahim 2013	56	12	24	23	100.0	0.70 [0.59, 0.80]	0.66 [0.48, 0.81]
El-Abd 2015	28	0	12	40	10.0	0.70 [0.53, 0.83]	1.00 [0.91, 1.00]
Mustika 2019	7	0	3	31	200.0	0.70 [0.35, 0.93]	1.00 [0.89, 1.00]
Gopal 2014	317	69	135	607	20.0	0.70 [0.66, 0.74]	0.90 [0.87, 0.92]
Feng 2016	232	26	97	345	21.33	0.71 [0.65, 0.75]	0.93 [0.90, 0.95]
Zinkin 2008	29	14	12	37	20.0	0.71 [0.54, 0.84]	0.73 [0.58, 0.84]
Liu 2007	161	29	66	51	20.0	0.71 [0.65, 0.77]	0.64 [0.52, 0.74]
Ge 2015	64	25	25	76	6.79	0.72 [0.61, 0.81]	0.75 [0.66, 0.83]
Almani 2004	72	11	28	89	8.6	0.72 [0.62, 0.81]	0.89 [0.81, 0.94]
Alsebaey 2016	31	5	12	17	13.65	0.72 [0.56, 0.85]	0.77 [0.55, 0.92]
Chang 1988	26	1	10	26	400.0	0.72 [0.55, 0.86]	0.96 [0.81, 1.00]
Sarwar 2014	125	14	48	88	20.85	0.72 [0.65, 0.79]	0.86 [0.78, 0.92]
Wu 2009	21	2	8	28	125.0	0.72 [0.53, 0.87]	0.93 [0.78, 0.99]
Tayob 2016b	29	58	11	523	22.9	0.72 [0.56, 0.85]	0.90 [0.87, 0.92]
Nomair 2019	16	3	6	19	95.0	0.73 [0.50, 0.89]	0.86 [0.65, 0.97]
Gani 2015	43	9	16	38	20.45	0.73 [0.60, 0.84]	0.81 [0.67, 0.91]
Tanglijvanich 2010	73	23	27	77	20.0	0.73 [0.63, 0.81]	0.77 [0.68, 0.85]
Qi 2020	88	17	32	72	11.88	0.73 [0.64, 0.81]	0.81 [0.71, 0.88]
Song 2011	64	163	23	311	18.76	0.74 [0.63, 0.82]	0.66 [0.61, 0.70]
Tang 2017a	130	32	46	158	12.3	0.74 [0.67, 0.80]	0.83 [0.77, 0.88]
Abu El Makarem 2011	84	0	29	120	43.0	0.74 [0.65, 0.82]	1.00 [0.97, 1.00]
Moriyama 2000	29	12	10	38	18.0	0.74 [0.58, 0.87]	0.76 [0.62, 0.87]
Kim 2019a	61	26	21	54	3.4	0.74 [0.64, 0.83]	0.68 [0.56, 0.78]
Yu 2016	38	18	13	120	5.0	0.75 [0.60, 0.86]	0.87 [0.80, 0.92]
Zhang 2020	47	26	16	48	20.0	0.75 [0.62, 0.85]	0.65 [0.53, 0.76]
Chen 2015	77	13	26	82	20.0	0.75 [0.65, 0.83]	0.86 [0.78, 0.93]
Yao 2016	597	144	199	905	11.62	0.75 [0.72, 0.78]	0.86 [0.84, 0.88]
Tremolada 1989	15	39	5	155	20.0	0.75 [0.51, 0.91]	0.80 [0.74, 0.85]
Arrigoni 1988	12	9	4	139	40.0	0.75 [0.48, 0.93]	0.94 [0.89, 0.97]
Elnemr 2012	45	6	15	54	20.0	0.75 [0.62, 0.85]	0.90 [0.79, 0.96]
Matsuda 2008	36	4	12	17	11.35	0.75 [0.60, 0.86]	0.81 [0.58, 0.95]
Ziada 2016	78	11	25	400	200.0	0.76 [0.66, 0.84]	0.97 [0.95, 0.99]
Fang 2010	110	16	35	112	20.0	0.76 [0.68, 0.83]	0.88 [0.80, 0.93]
Minami 2015a	22	6	7	52	5.0	0.76 [0.56, 0.90]	0.90 [0.79, 0.96]
da Costa 2015b	38	7	12	43	20.0	0.76 [0.62, 0.87]	0.86 [0.73, 0.94]
Mukozu 2013	45	11	14	17	15.0	0.76 [0.63, 0.86]	0.61 [0.41, 0.78]
Marrero 2003	42	22	13	82	11.0	0.76 [0.63, 0.87]	0.79 [0.70, 0.86]
Miura 2010	232	45	71	89	10.0	0.77 [0.71, 0.81]	0.66 [0.58, 0.74]
Ali 2019	46	11	14	49	117.0	0.77 [0.64, 0.87]	0.82 [0.70, 0.90]
Wong 2008	363	51	109	56	20.0	0.77 [0.73, 0.81]	0.52 [0.42, 0.62]
Sharma 2010	54	12	16	56	11.88	0.77 [0.66, 0.86]	0.82 [0.71, 0.91]
Paul 2007	78	42	23	152	10.7	0.77 [0.68, 0.85]	0.78 [0.72, 0.84]
El Shafie 2012	24	14	7	21	28.51	0.77 [0.59, 0.90]	0.60 [0.42, 0.76]
Erdal 2016	31	11	9	43	6.0	0.78 [0.62, 0.89]	0.80 [0.66, 0.89]
Shang 2012b	71	1	20	22	20.0	0.78 [0.68, 0.86]	0.96 [0.78, 1.00]
Abdelghany 2018	11	4	3	6	3.5	0.79 [0.49, 0.95]	0.60 [0.26, 0.88]
Kim 2012	154	57	42	297	20.0	0.79 [0.72, 0.84]	0.84 [0.80, 0.88]
Okazaki 1984	11	44	3	187	20.0	0.79 [0.49, 0.95]	0.81 [0.75, 0.86]
Khairy 2015	37	9	10	8	62.5	0.79 [0.64, 0.89]	0.47 [0.23, 0.72]
Wang 2019b	71	31	19	59	4.93	0.79 [0.69, 0.87]	0.66 [0.55, 0.75]
Liao 2012	47	18	12	78	20.0	0.80 [0.67, 0.89]	0.81 [0.72, 0.88]
Izzo 1999	79	19	20	340	10.0	0.80 [0.71, 0.87]	0.95 [0.92, 0.97]
Youns 2013	32	14	8	26	9.0	0.80 [0.64, 0.91]	0.65 [0.48, 0.79]
Fujii1995	40	6	10	44	20.0	0.80 [0.66, 0.90]	0.88 [0.76, 0.95]
Wong 2014a	46	289	11	1185	6.0	0.81 [0.68, 0.90]	0.80 [0.78, 0.82]
Chan 2013	46	289	11	1185	6.0	0.81 [0.68, 0.90]	0.80 [0.78, 0.82]
Iyer 2018	165	4	38	115	9.75	0.81 [0.75, 0.86]	0.97 [0.92, 0.99]
El Mahdy 2019	49	25	11	50	11.5	0.82 [0.70, 0.90]	0.67 [0.55, 0.77]
Caviglia 2016	45	6	10	38	5.3	0.82 [0.69, 0.91]	0.86 [0.73, 0.95]
El-Abd 2016	41	4	9	26	12.9	0.82 [0.69, 0.91]	0.87 [0.69, 0.96]
Chan 2014	460	72	98	171	10.0	0.82 [0.79, 0.86]	0.70 [0.64, 0.76]
Alexander 1978	29	20	6	43	30.0	0.83 [0.66, 0.93]	0.68 [0.55, 0.79]
Abdel-Razik 2016	68	28	14	52	11.8	0.83 [0.73, 0.90]	0.65 [0.54, 0.75]

Figure 14. (Continued)

Chan 2014	460	72	98	171	10.0	0.82 [0.79, 0.86]	0.70 [0.64, 0.76]
Alexander 1978	29	20	6	43	30.0	0.83 [0.66, 0.93]	0.68 [0.55, 0.79]
Abdel-Razik 2016	68	28	14	52	11.8	0.83 [0.73, 0.90]	0.65 [0.54, 0.75]
Wong 2014b	44	138	9	233	6.0	0.83 [0.70, 0.92]	0.63 [0.58, 0.68]
da Costa 2015a	49	0	10	49	20.0	0.83 [0.71, 0.92]	1.00 [0.93, 1.00]
Badr 2014	25	2	5	28	200.0	0.83 [0.65, 0.94]	0.93 [0.78, 0.99]
Eltaher 2016	25	6	5	24	273.0	0.83 [0.65, 0.94]	0.80 [0.61, 0.92]
Mehinovic 2018	42	9	8	41	23.34	0.84 [0.71, 0.93]	0.82 [0.69, 0.91]
Kim 2019	45	18	8	29	3.4	0.85 [0.72, 0.93]	0.62 [0.46, 0.75]
Fabris 1991	23	38	4	173	20.0	0.85 [0.66, 0.96]	0.82 [0.76, 0.87]
Li 2019a	144	35	25	207	4.05	0.85 [0.79, 0.90]	0.86 [0.80, 0.90]
Bell 1982	12	10	2	100	20.0	0.86 [0.57, 0.98]	0.91 [0.84, 0.96]
Raedle 1995	6	20	1	120	20.0	0.86 [0.42, 1.00]	0.86 [0.79, 0.91]
Gad 2005	95	52	15	182	10.0	0.86 [0.79, 0.92]	0.78 [0.72, 0.83]
Tong 2001	27	86	4	485	8.0	0.87 [0.70, 0.96]	0.85 [0.82, 0.88]
Kim 2006b	48	52	7	10	7.0	0.87 [0.76, 0.95]	0.16 [0.08, 0.28]
Ding 2020	413	131	60	379	7.21	0.87 [0.84, 0.90]	0.74 [0.70, 0.78]
Zekri 2013	35	29	5	61	10.35	0.88 [0.73, 0.96]	0.68 [0.57, 0.77]
Shariff 2010	16	2	2	8	24.2	0.89 [0.65, 0.99]	0.80 [0.44, 0.97]
Wang 2013b	113	8	13	107	4.0	0.90 [0.83, 0.94]	0.93 [0.87, 0.97]
El Gawad 2014	36	4	4	6	20.0	0.90 [0.76, 0.97]	0.60 [0.26, 0.88]
Montaser 2012	36	1	4	40	23.0	0.90 [0.76, 0.97]	0.98 [0.87, 1.00]
Salem 2013	27	7	3	23	10.4	0.90 [0.73, 0.98]	0.77 [0.58, 0.90]
Chen 1977	115	25	10	49	10.0	0.92 [0.86, 0.96]	0.66 [0.54, 0.77]
Wang 2009	156	89	8	24	8.0	0.95 [0.91, 0.98]	0.21 [0.14, 0.30]
Johnson 1978	29	1	1	99	10.0	0.97 [0.83, 1.00]	0.99 [0.95, 1.00]
McMahon 2000	31	70	1	1385	15.0	0.97 [0.84, 1.00]	0.95 [0.94, 0.96]
Hu 2019	422	143	7	434	8.72	0.98 [0.97, 0.99]	0.75 [0.71, 0.79]
Toraih 2018	30	0	0	20	205.0	1.00 [0.88, 1.00]	1.00 [0.83, 1.00]
Aboelfotoh 2018	40	0	0	40	15.0	1.00 [0.91, 1.00]	1.00 [0.91, 1.00]
Liu 2010a	28	47	0	32	400.0	1.00 [0.88, 1.00]	0.41 [0.30, 0.52]

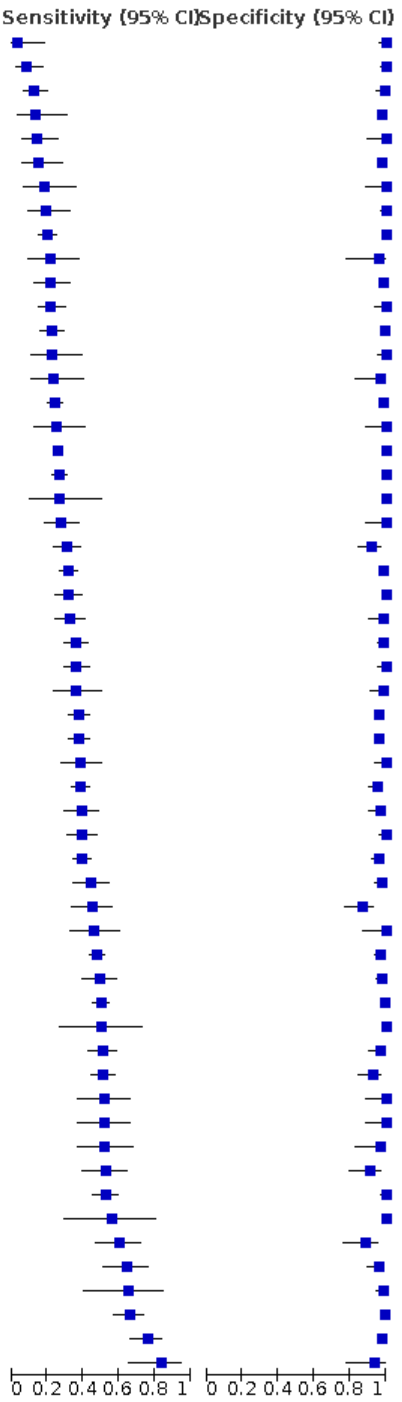


Appendix 5. Results of studies of alpha-fetoprotein with a cut-off value of 200ng/mL

Figure 15

Figure 15. Forest plots of sensitivity and specificity of alpha-foetoprotein with a cut-off value around 200 ng/mL.against different reference standards in 56 studies ordered by increasing sensitivity. Referencev standards were: the pathology of the explanted liver in case of transplantation.;the histology of resected focal liver lesions, or the histology of biopsied focal liver lesions with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months. TP = true positive; FP = false positive; FN = false negative; TN = true negative. Values between brackets are the 95% confidence intervals (CIs) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue square) and its 95% CI (black horizontal line).

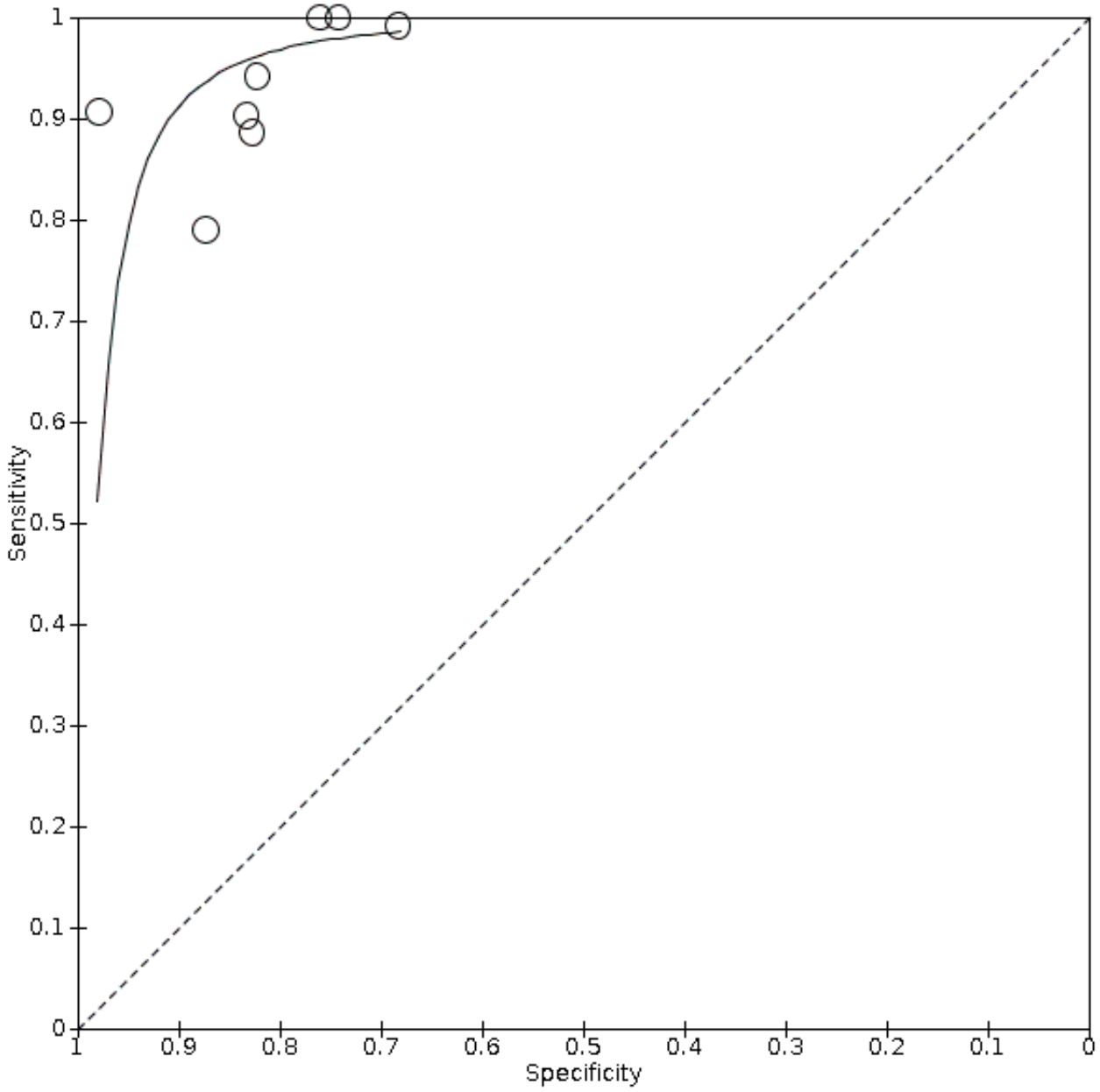
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Nomura 1996	1	0	26	101	0.04 [0.00, 0.19]	1.00 [0.96, 1.00]
Sassa 1999	5	0	56	134	0.08 [0.03, 0.18]	1.00 [0.97, 1.00]
Kumada 2014	13	1	91	103	0.13 [0.07, 0.20]	0.99 [0.95, 1.00]
Ishii 2000	4	18	25	687	0.14 [0.04, 0.32]	0.97 [0.96, 0.98]
Shimizu 2002	8	0	48	34	0.14 [0.06, 0.26]	1.00 [0.90, 1.00]
Sterling 2012	7	17	39	792	0.15 [0.06, 0.29]	0.98 [0.97, 0.99]
Passos-Castilho 2015	6	0	26	30	0.19 [0.07, 0.36]	1.00 [0.88, 1.00]
Yu 2016	10	0	41	138	0.20 [0.10, 0.33]	1.00 [0.97, 1.00]
Snowberger 2007	48	0	191	1696	0.20 [0.15, 0.26]	1.00 [1.00, 1.00]
Bon 1998	8	1	29	22	0.22 [0.10, 0.38]	0.96 [0.78, 1.00]
Sterling 2009	16	4	58	274	0.22 [0.13, 0.33]	0.99 [0.96, 1.00]
Pompili 2003	29	0	102	59	0.22 [0.15, 0.30]	1.00 [0.94, 1.00]
Trevisani 2001	38	1	132	169	0.22 [0.16, 0.29]	0.99 [0.97, 1.00]
Lok 2010	9	0	30	77	0.23 [0.11, 0.39]	1.00 [0.95, 1.00]
Song 2002	9	1	29	30	0.24 [0.11, 0.40]	0.97 [0.83, 1.00]
Ahn 2016	89	6	277	360	0.24 [0.20, 0.29]	0.98 [0.96, 0.99]
Shaheen 2015	10	0	30	30	0.25 [0.13, 0.41]	1.00 [0.88, 1.00]
Nakamura 2006	354	0	1007	348	0.26 [0.24, 0.28]	1.00 [0.99, 1.00]
Tsai 2017	134	0	359	493	0.27 [0.23, 0.31]	1.00 [0.99, 1.00]
Chalasani 1999	6	0	16	263	0.27 [0.11, 0.50]	1.00 [0.99, 1.00]
Zuo 2016	25	0	65	30	0.28 [0.19, 0.38]	1.00 [0.88, 1.00]
Han 2014	49	7	111	81	0.31 [0.24, 0.38]	0.92 [0.84, 0.97]
Lim 2015	115	4	246	272	0.32 [0.27, 0.37]	0.99 [0.96, 1.00]
Nguyen 2002	52	0	111	149	0.32 [0.25, 0.40]	1.00 [0.98, 1.00]
Yang 2013a	40	1	83	56	0.33 [0.24, 0.42]	0.98 [0.91, 1.00]
Sanai 2010	74	3	132	196	0.36 [0.29, 0.43]	0.98 [0.96, 1.00]
Arrieta 2007	70	0	123	74	0.36 [0.29, 0.43]	1.00 [0.95, 1.00]
Kim 2006b	20	1	35	61	0.36 [0.24, 0.50]	0.98 [0.91, 1.00]
Soroida 2012	95	18	157	437	0.38 [0.32, 0.44]	0.96 [0.94, 0.98]
Zhu 2013	95	18	157	437	0.38 [0.32, 0.44]	0.96 [0.94, 0.98]
Ozkan 2011	29	0	46	55	0.39 [0.28, 0.51]	1.00 [0.94, 1.00]
Lee 2004	143	9	226	169	0.39 [0.34, 0.44]	0.95 [0.91, 0.98]
Attallah 2018	43	2	67	70	0.39 [0.30, 0.49]	0.97 [0.90, 1.00]
Attallah 2011	59	0	91	100	0.39 [0.31, 0.48]	1.00 [0.96, 1.00]
Hu 2018	146	7	223	169	0.40 [0.35, 0.45]	0.96 [0.92, 0.98]
Paul 2007	45	5	56	189	0.45 [0.35, 0.55]	0.97 [0.94, 0.99]
Liu 2018	36	11	44	71	0.45 [0.34, 0.57]	0.87 [0.77, 0.93]
Kim 2018	25	0	29	26	0.46 [0.33, 0.60]	1.00 [0.87, 1.00]
Chan 2014	268	7	294	236	0.48 [0.43, 0.52]	0.97 [0.94, 0.99]
Takikawa 1992	57	6	59	247	0.49 [0.40, 0.59]	0.98 [0.95, 0.99]
Gopal 2014	226	5	226	667	0.50 [0.45, 0.55]	0.99 [0.98, 1.00]
Tremolada 1989	10	0	10	194	0.50 [0.27, 0.73]	1.00 [0.98, 1.00]
Chimparlee 2015	80	2	77	71	0.51 [0.43, 0.59]	0.97 [0.90, 1.00]
Liu 2007	116	6	111	74	0.51 [0.44, 0.58]	0.93 [0.84, 0.97]
El Moety 2011	26	0	24	30	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
Eissa 2013	26	0	24	30	0.52 [0.37, 0.66]	1.00 [0.88, 1.00]
Mashaly 2018	23	1	21	30	0.52 [0.37, 0.68]	0.97 [0.83, 1.00]
Bachtiar 2009	34	5	31	49	0.52 [0.40, 0.65]	0.91 [0.80, 0.97]
Iyer 2018	107	0	96	119	0.53 [0.46, 0.60]	1.00 [0.97, 1.00]
Arrigoni 1988	9	0	7	148	0.56 [0.30, 0.80]	1.00 [0.98, 1.00]
Poon 2001	39	6	26	45	0.60 [0.47, 0.72]	0.88 [0.76, 0.96]
Liao 2012	38	4	21	92	0.64 [0.51, 0.76]	0.96 [0.90, 0.99]
Matievskaya 2003	13	2	7	137	0.65 [0.41, 0.85]	0.99 [0.95, 1.00]
Maringhini 1988	96	1	50	216	0.66 [0.57, 0.73]	1.00 [0.97, 1.00]
Ziada 2016	78	11	25	400	0.76 [0.66, 0.84]	0.97 [0.95, 0.99]
Badr 2014	25	2	5	28	0.83 [0.65, 0.94]	0.93 [0.78, 0.99]



Appendix 6. Summary receiver operating characteristic (ROC) comparing the combination of alpha-fetoprotein and ultrasound

Figure 16

Figure 16. Summary receiver operating characteristic (ROC) comparing in 8 studies the combination of alpha-foetoprotein and ultrasound against different reference standards. Reference standards were: the pathology of the explanted liver in case of transplantation, the histology of resected focal liver lesions, or the histology of biopsied focal liver lesions with a follow-up period of at least six months, typical characteristics on cross-sectional multiphasic contrast CT or MRI, with a follow-up period of at least six months.



HISTORY

Protocol first published: Issue 6, 2019
Review first published: Issue 4, 2021

CONTRIBUTIONS OF AUTHORS

AC co-ordinated the protocol design, completed the search for studies, performed data extraction and quality assessment, drafted parts of the review, provided methodological and statistical analyses and expert hepatology opinion, and wrote the final version of the manuscript.
TN wrote the protocol, and completed the search for studies, performed data extraction and quality assessment, drafted parts of the review, provided expert radiology opinion, and wrote the final version of the manuscript.
DM provided expert radiology opinion and commented critically on the protocol and the final review.
VG performed searches for references, provided expert hepatology opinion, and reviewed the final version of the manuscript.
MF performed searches for references, critically comment on the review, and provided expert hepatology opinion.
DŠ provided expert hepatology opinion and critically commented on the protocol and the final review.
GC wrote the protocol, performed data extraction and quality assessment, drafted parts of the manuscript, conducted statistical analyses, provided methodological expertise, and reviewed the final version of the manuscript.

All authors have read and approved the review for publication.

DECLARATIONS OF INTEREST

AC: none known
TN: none known
DM: none known
VG: none known
MF: none known
DŠ: none known
GC: none known

SOURCES OF SUPPORT

Internal sources

- None, Other

External sources

- None, Other

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We expanded the title to include the studied population in our review: i.e. from 'Abdominal ultrasound and alpha-fetoprotein for the diagnosis of hepatocellular carcinoma' (Protocol title) into 'Abdominal ultrasound and alpha-fetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease'.

We did not extract data on 'race', and hence, we deleted the word 'race' from our text on data collection. We considered the usual way of reporting ethnicity data not being informative and potentially misleading.

We removed "Different AFP positivity cut-off values in studies using ultrasound and AFP in combination" as the last secondary outcome. The reason for this is that we did not perform a comparison of studies with a combination of ultrasound and AFP as index test, using different cut-off values of AFP, as all studies, except two, used a cut-off value of 20 ng/mL.

We did not perform a planned sensitivity analysis on whether or not the positivity threshold was pre-specified for the AFP tests because we chose to analyse the results of studies using the most common cut-off values of 20 ng/mL and 200 ng/mL.

INDEX TERMS

Medical Subject Headings (MeSH)

Abdomen [diagnostic imaging]; alpha-Fetoproteins [*analysis]; Bias; Biomarkers, Tumor [blood]; Carcinoma, Hepatocellular [blood] [*diagnosis] [pathology]; Case-Control Studies; Chronic Disease; Confidence Intervals; Cross-Sectional Studies; Liver Diseases [*complications]; Liver Neoplasms [blood] [*diagnosis] [pathology]; Sensitivity and Specificity; Ultrasonography [*methods]

MeSH check words

Adult; Female; Humans; Male