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

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

Cochrane Library

Cochrane Database of Systematic Reviews

Abdominal ultrasound and alpha-fetoprotein for the diagnosis of hepatocellular carcinoma (Protocol)

Colli A, Nadarević T, Miletić D, Giljaca V, Fraquelli M, Štimac D, Casazza G

Colli A, Nadarević T, Miletić D, Giljaca V, Fraquelli M, Štimac D, Casazza G.
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.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

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

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

[1]Department of Internal Medicine, A Manzoni Hospital ASST Lecco, Lecco, Italy. [2]Department of Radiology, Clinical Hospital Centre Rijeka, Rijeka, Croatia. [3]Directorate of Surgery, Department of Gastroenterology, Heart of England NHS Foundation Trust, Birmingham Heartlands Hospital, Birmingham, UK. [4]Gastroenterology and Endoscopy Unit, Fondazione IRCCS Cà Granda - Ospedale Maggiore Policlinico, Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy. [5]Department of Gastroenterology, Clinical Hospital Centre Rijeka, Rijeka, Croatia. [6]Dipartimento di Scienze Biomediche e Cliniche "L. Sacco", Università degli Studi di Milano, Milan, Italy

Contact address: Agostino Colli, Department of Internal Medicine, A Manzoni Hospital ASST Lecco, Via dell'Eremo, 9/11, Lecco, 23900, Italy. colliagostino@gmail.com.

Editorial group: Cochrane Hepato-Biliary Group.
Publication status and date: New, published in Issue 6, 2019.

Citation: Colli A, Nadarević T, Miletić D, Giljaca V, Fraquelli M, Štimac D, Casazza G. 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.

ABSTRACT

This is a protocol for a Cochrane Review (Diagnostic test accuracy). The objectives are as follows:

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

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). HCC ranks fifth in terms of global cases of cancer and second in terms of cancer deaths for males (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 of liver disease because of the absence of specific symptoms of the malignancy, other than those related to chronic liver disease. Only 20% of patients 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 it presents as either a single lesion with a maximum diameter of less than five cm, or up to two lesions, each with a maximum diameter of three cm (Mazzaferro 1996; EASL-EORTC 2012; Omata 2017; EASL 2018; Heimbach 2018). Furthermore, curative treatment options are unfeasible 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, accurate and early diagnosis of HCC is considered of high importance.

Abdominal ultrasound 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 ultrasound surveillance in detecting HCC in people without any symptoms (Tzartzeva 2018). However, the same study showed a poor result for ultrasound 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 ultrasound can be influenced by the expertise of the operator and the quality of the equipment.

Alpha-fetoprotein (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 ultrasound 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; Tzartzeva 2018), but to our knowledge, there is no recent systematic review which compares AFP alone, ultrasound alone, and combination of ultrasound and AFP 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 first stage of HCC, 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 two 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 magnetic resonance imaging features: focal liver lesions which show non-rim-like hyperenhancement in 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 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). It is comprised of four elements: tumour extension, liver functional reserve, physical status, and cancer-related symptoms. According to the Barcelona Clinic Liver Cancer 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 five cm, or up to three HCC lesions, each with a diameter equal to or less than three 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 ultrasound 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 ultrasound categories for diagnosing suspected liver lesions: US-1 (negative), US-2 (subthreshold), and US-3 (positive). Since ultrasound is an operator-dependent imaging modality and limitations due to patient characteristics can occur, an ultrasound 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 2017).

Alpha-fetoprotein is a glycoprotein of 591 amino acids and a carbohydrate moiety which is assessed in serum by enzyme immunoassays. 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 advanced chronic liver disease, a surveillance programme is usually recommended. The surveillance programmes among the different scientific societies have minimal variations (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 are in need of surveillance. It is suggested that surveillance be performed using abdominal ultrasound, 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 ultrasound every six months. A three- to four-month interval is recommended in people where a nodule of less than one 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 are therefore 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 ultrasound 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.

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 chronic advanced liver disease is considered equivalent to a histological diagnosis of cirrhosis (de Franchis 2015). No test is recommended by the guidelines, prior to a surveillance programme for HCC detection.

Role of index test(s)

Both abdominal ultrasound and AFP (independently, or in combination, or in sequence) are used as first-line tests to exclude the presence of focal liver lesions suspected of being HCC. Further testing is required to confirm the diagnosis as well as for staging.

Alternative test(s)

Contrast-enhanced ultrasound (CEUS) is an advanced form of ultrasound 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 to 40 seconds), then portal (40 to 120 seconds), 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 hyperenhancement and subsequent washout appearance. The advantages of ultrasound contrast 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 multidetector 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 phases, 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 hyperenhancement 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 advanced 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 ultrasound and AFP serum concentration may clarify whether the absence of benefit in surveillance programmes 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 programmes.

People with previous diagnoses of, and who had previous treatments for, HCC make up a distinct group and 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, 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. A review assessing the accuracy of MRI for diagnosing HCC is in progress (Tang 2017). We are planning to produce an overview of the reviews that assess abdominal ultrasound and AFP, CEUS, CT, and MRI for the diagnosis of HCC.

OBJECTIVES

To assess the diagnostic accuracy of abdominal ultrasound and alpha-fetoprotein (AFP), alone or in combination, for the diagnosis

of hepatocellular carcinoma (HCC) of any size and at any stage in people with chronic advanced liver disease, either in a surveillance programme or in a clinical setting.

Secondary objectives

- To assess the diagnostic accuracy of abdominal ultrasound and AFP, alone or in combination, for the diagnosis of resectable HCC in people with chronic advanced 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 five cm, or fewer than three lesions with a maximum diameter of three 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 the Americas compared to Europe compared to Asia;
  * 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 screening programmes 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 (hepatitis C or hepatitis B virus associated cirrhosis compared to all other aetiologies);
  * different severity of the underlying chronic liver disease (per cent of participants with MELD (model for end-stage liver disease) score less than 15 or Child Pugh score A);
  * different AFP positivity cut-off values in studies using ultrasound and AFP in combination.

METHODS

Criteria for considering studies for this review

Types of studies

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

We will consider studies of cross-sectional design which include participants with clinical suspicion of HCC, or cohort studies which include high-risk participants in a surveillance programme, as well as studies with a case-control design that compare people with known HCC to a matched control (participants with chronic liver disease without evidence of HCC). We will include studies assessing one index test (AFP or ultrasound, or both in combination) if all the participants underwent at least one of the acceptable reference standards. We plan to exclude studies that analysed data only per-lesion, that is, those that considered the number of lesions rather than participants, unless participant data are made available by study authors.

Participants

Eligibility criteria

We will include 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 must have been treatment-naive for HCC when enrolled in the respective study.

Exclusion criteria

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

Index tests

We will include abdominal ultrasound alone, AFP alone, and a combination of abdominal ultrasound and AFP for the detection of HCC in people with advanced chronic liver disease.

Target conditions

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

Reference standards

We will accept 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 resected or biopsied focal liver lesion(s) with a follow-up period of at least six months to exclude 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 that

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 will search the Cochrane Hepato-Biliary Group Controlled Trials Register (Cochrane Hepato-Biliary Group Module), the Cochrane Hepato-Biliary Group Diagnostic Test of Accuracy Studies Register (Cochrane Hepato-Biliary Group Module), the Cochrane Library, MEDLINE Ovid, Embase Ovid, LILACS (Bireme), Science Citation Index - Expanded (Web of Science), and Conference Proceedings Citation Index – Science (Web of Science) (Royle 2003). See Appendix 1 for the preliminary search strategies, with the expected time spans of the searches.

We will apply no language or document type restrictions.

Searching other resources

We will attempt to identify additional references by manually searching articles retrieved from digital databases and relevant review articles. We will seek information on unpublished studies by contacting experts in the field. In addition, we will handsearch 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 will also search 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 will follow available guidelines as provided in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (DTA Handbook 2013).

Selection of studies

We will retrieve publications if they are potentially eligible for inclusion on the basis of abstract review, or if they are relevant review articles for a manual reference search. Two review authors (AC and TN) will independently review publications for eligibility. To determine eligibility, we will assess each publication to determine whether participants met the inclusion criteria detailed above. We will include abstracts only if they provide sufficient data for analysis. We will resolve disagreements by consensus.

Data extraction and management

Two review authors (AC and TN) will complete a piloted data extraction form for each included study. Each review author will independently retrieve study data. In cases of disagreement, we will reach consensus through discussion.

We will retrieve the following data.

- General information: title, journal, year, publication status, and study design (prospective versus retrospective), surveillance programme or clinical cohorts.
- Sample size: number of participants meeting the criteria and total number of participants screened.
- Baseline characteristics: baseline diagnosis, age, sex, race, 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, and false negative findings. We will extract these data for each presented cut-off value and for either HCC of any size, stage, and resectable HCC.

We will summarise 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 will enter the data into Review Manager 5.3 software.

Missing data

We will contact primary authors by email to ask for missing data which are needed to complete the 2 × 2 tables. If we receive no reply, we will send a second email after two weeks. If no reply is received, we will place the study in question in the list of studies awaiting classification.

Assessment of methodological quality

Two review authors (AC and TN) will independently assess 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 will reach a consensus through discussion. We will address 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, a number of inadequate results should be reported and those should be assessed as false in the analyses. If non-evaluable results are not reported or are excluded from the analysis, we will consider the study to be at high risk of bias. Ultrasound visualisation can often be suboptimal due to patient characteristics; therefore, lack of reporting or exclusion of those patients from analyses could overestimate the accuracy of ultrasound. We will classify a study as having high risk of bias if at least one of the domains of QUADAS-2 is judged as being at high or unclear risk of bias (Appendix 2).

Statistical analysis and data synthesis

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

Ultrasound

Abdominal ultrasound is considered positive when a lesion of more than 10 mm with no definitely benign features is observed, or a new venous thrombus has been detected according to defined criteria (LI-RADS 2017). Subthreshold lesions of less than 10 mm are noted

only when no definitely benign features have been observed (LI-RADS 2017). Firstly, we will perform a graphical descriptive analysis of the included studies. We will report forest plots (sensitivity and specificity separately, with their 95% confidence intervals (CIs)), and we will provide a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Secondly, we will perform a meta-analysis using the bivariate model, and we will provide estimates of summary sensitivity and specificity. We will use the pooled estimates obtained from the fitted models to calculate summary estimates of positive and negative likelihood ratios (LR+ and LR-, respectively).

Alpha-fetoprotein

Alpha-fetoprotein is considered positive when higher than a defined cut-off (threshold) value is noted (Colli 2006; Marrero 2009; Lok 2010). Firstly, we will perform a graphical descriptive analysis of the included studies. We will report forest plots (sensitivity and specificity separately, with their 95% CIs), and provide a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Secondly, we will perform a meta-analysis. In the case that primary studies reported accuracy estimates of AFP using different cut-off values, we will use 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 will use the bivariate model, and we will provide estimates of summary sensitivity and specificity. We will use 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 will report sensitivities and specificities for all cut-off values, but we will use a single cut-off value for each study in HSROC or bivariate analysis. The most common cut-off values are 10, 20, 100, or 400 nanograms per millilitre (ng/mL).

Non-evaluable index test results

In case of non-evaluable index test results (especially relevant for ultrasound), we plan to analyse data according to the intention to diagnose (ITD) principle (Schuetz 2012). We will classify participants with non-evaluable results as false positive if they had a negative reference standard, or false negative result on a positive reference standard. If data for the ITD analyses are not retrievable from the text, we will contact the study authors. If we receive no response, we will include the study in the analyses with data retrievable from the published manuscript and consider the study to be at high risk of bias (see Data extraction and management).

Combination of ultrasound and alpha-fetoprotein

The index test obtained by the combination of ultrasound and AFP tests is considered positive when at least one of the two tests is positive. Firstly, we will perform a graphical descriptive analysis of the included studies. We will report forest plots (sensitivity and specificity separately, with their 95% CIs), and we will provide a graphical representation of studies in the receiver operating characteristic (ROC) space (sensitivity against 1 - specificity). Secondly, we will perform 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 will use 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 will use the bivariate model and will provide estimates of summary sensitivity and specificity. We will use 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 will report sensitivities and specificities for all cut-off values, but we will use a single cut-off value for each study in HSROC or bivariate analysis.

Comparisons

The combination of the two tests, ultrasound and AFP, is considered positive when at least one of the two tests is positive. We plan to make 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 HSROC (for comparisons of SROC curves) or bivariate (for comparisons of sensitivity and specificity at fixed cut-off value) model. We plan to assess 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 will perform both indirect and direct comparisons, if sufficient data are available.

We will consider P values less than 0.05 as two-sided and statistically significant. We will perform 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 plan to investigate the effects of the following predefined sources of heterogeneity.

- Study design (prospective compared to retrospective, case-control compared to cross-sectional cohort studies).
- 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 Americas compared to Europe compared to Asia.
- 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 screening programmes 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 (hepatitis C or hepatitis B virus associated cirrhosis compared to all other aetiologies).
- Different severity of the underlying chronic liver disease (per cent of participants with MELD score less than 15 or Child Pugh score A).

- AFP positivity cut-off values in studies using ultrasound and AFP in combination.

We will estimate effects by adding covariates, categorical or continuous, to the bivariate or to the HSROC models. We will assess 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 plan to assess 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 have defined the following signalling questions as most relevant, and plan to conduct a sensitivity analyses in which we exclude 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 suspect 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 also plan to conduct 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 do not plan to use a filter search strategy or to implement any language or sample limitations. We do not plan to test for publication bias due to the lack of validated methods for diagnostic test accuracy reviews.

ACKNOWLEDGEMENTS

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, Rigshospitalet, Copenhagen University Hospital, 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.

Peer Reviewers: names of peer reviewers to be obtained by the Cochrane Diagnostic Accuracy Reviews Editorial Team, UK
Contact Editor from the Cochrane Diagnostic Accuracy Reviews Editorial Team: Susan Mallett, UK
Contact Editor: Christian Gluud, Denmark
Sign-off Editor: Christian Gluud, Denmark

REFERENCES

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.

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* 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* 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.

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.

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-CTrial. *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.

GBD 2017

Global Burden of Disease Liver Cancer Collaboration. The burden of primary liver cancer and underlying ethiologies 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.

Gopal 2014

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**:870-7.

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* 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 JNM, 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. [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 editor(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.

Kumada 2014

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 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* 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 2017

The American College of Radiology. Liver imaging reporting and data system. www.acr.org/quality-safety/resources/LIRADS (accessed 7 July 2018).

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 27 July 2018).

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 2010

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**(2):493-502.

Marrero 2009

Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, et al. Alpha-fetoprotein, des-gamma carboxy prothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 2009;**137**(1):110-8.

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.

Nakamura 2006

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

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.

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* 2018;**67**(1):401-21.

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.

Shah 2014

Shah S, Shukla A, Paunipagar B. Radiological features of hepatocellular carcinoma. *Journal of Clinical and Experimental Hepatology* 2014;**4**:63-6.

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.

Tang 2017

Tang A, McInnes M, Hope TA, Vu KN, Amre D, Wolfson T, et al. Magnetic resonance imaging performed with gadoxetate disodium for the diagnosis of hepatocellular carcinoma in cirrhotic and non-cirrhotic patients. *Cochrane Database of Systematic Reviews* 2017, Issue 8. [DOI: 10.1002/14651858.CD012766]

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.

ADDITIONAL TABLES

Table 1. Guideline recommendations for surveillance for hepatocellular carcinoma

GUIDELINE	INDICATION TO SURVEILLANCE	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-fetoprotein; HCC: hepatocellular carcinoma

APPENDICES

Appendix 1. Search strategies

Database	Time span	Search strategy
The Cochrane Hepato-Biliary Group Controlled Trials Register	Date will be given at review stage	((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 of Accuracy Studies Register	Date will be given at review stage	((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 Library	Latest issue	#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

(Continued)

		#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 the date of search	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, organism 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]

(Continued)

		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, key-word 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 the date of search	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 sub-heading word, candidate term word] 9. 7 or 8 10. exp liver cell carcinoma/ 11. exp liver tumor/ 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 sub-heading 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

(Continued)

		18. limit 17 to (human and (adult <18 to 64 years> or aged <65+ years>))
LILACS (Bireme)	1982 to the date of search	(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	1900 to the date of search	#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	1990 to the date of search	#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
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?*"	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.	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). No - if other reference tests than pathology of explanted liver were used.	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.

(Continued)

	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.	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.	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.	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 (non-evaluable results considered as false)?"* Yes - if all enrolled participants were included in the analysis and non-evaluable index test results were analysed according to the intention to diagnose principle). No - if any participant was excluded from the analysis for any reason or non-evaluable 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?*	*Are there concerns that the index test, its conduct, or interpretation differ from the review question?*	*Are there concerns that the target condition as defined by the reference standard does not match the question?*	-

(Continued)

	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.	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.	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.)

CONTRIBUTIONS OF AUTHORS

AC co-ordinated the protocol design, and will evaluate studies for inclusion, extract data from studies, assess the risk of bias, and design the final review.
TN wrote the protocol, and will perform searches for references, evaluate references for obtaining the full reports, evaluate studies for inclusion, extract data from studies, assess the risk of bias, and write the final review.
MF will perform searches for references and critically comment on the review.
VG commented on the protocol, and will evaluate references for obtaining the full-text reports, and write the final review.
DM commented on the protocol, and will critically comment on the final review.
DŠ critically commented on the protocol, and will act as arbiter if review authors cannot reach a consensus, and will critically comment on the final review.
GC wrote the protocol and will critically comment on the final review.

All authors have seen and approved the final version of the protocol prior to publication.

DECLARATIONS OF INTEREST

AC: none known.
TN: none known.
MF: none known.
VG: none known.
DM: none known.
DŠ: none known.
GC: none known.

SOURCES OF SUPPORT

Internal sources

- None, Other.

External sources

- None, Other.