Research Articles

Screening for hepatocellular carcinoma in chronic liver disease: a systematic review and meta-analysis of randomized controlled trials comparing screening methodologies

James O’Connell, Siobhan Rooney

1 Department of International Health and Tropical Medicine, Royal College of Surgeons in Ireland, Dublin, Ireland, 2 University College Hospital Galway, Saolta Healthcare Group, Health Service Executive, Galway, Ireland

Keywords: chronic liver disease, liver cancer, hepatocellular carcinoma

Journal of Global Health Reports
Vol. 4, 2020

Background
Globally, hepatocellular carcinoma (HCC) is the fifth most prevalent cancer and the second most common cause of cancer-related mortality. HCC is often asymptomatic until an advanced stage. Current guidelines recommend ultrasound surveillance with or without measurement of serum alpha-fetoprotein. Our objective was to determine if screening for HCC is beneficial or harmful in patients with chronic liver disease. Primary outcomes were all-cause mortality and quality of life. Secondary outcomes were mortality due to HCC, the number of cases of HCC detected, and adverse events.

Methods
This is a systematic review and meta-analysis of data from randomized controlled trials. To be included trials had to randomize patients to either an HCC screening group or non-screening group, randomize patients to different screening frequencies or randomize patients to different screening methods. All published reports of randomized trials on screening for HCC were eligible for inclusion, irrespective of the language of publication. Studies had to include patients with chronic liver disease. Data extraction were performed independently by two reviewers.

Results
When screening with six-monthly alpha-fetoprotein and ultrasound abdomen was compared to no screening there was no evidence of difference in HCC related mortality when adjusted for clustering across a range of intracluster correlation coefficients (Intracluster coefficient (ICC) 0.02, odds ratio (OR) 0.60, 95% confidence interval (CI) 0.31-1.15). Screening with six-monthly alpha-fetoprotein when compared to a single alpha-fetoprotein check did not result in a statistically significant difference in all-cause mortality (OR 1.02, 95% confidence interval (CI) 0.65-1.60), mortality due to HCC (OR 1.01, 95% CI 0.57-1.78) or the number of HCC detected (OR 1.11 95% CI 0.64-1.92). There was no evidence of difference in all-cause mortality (OR 0.81, 95% CI 0.26-2.53), mortality due to hepatocellular carcinoma (OR 0.81, 95% CI 0.26-2.53) or the number of patients with HCC detected (OR 1.09 95% CI 0.40-2.99) when twice-a-year ultrasound was compared with annual CT. There was no statistically significant difference when screening more frequently was compared to less frequently in terms of all-cause mortality (OR 0.86, 95% CI 0.56-1.32), mortality due to hepatocellular carcinoma (OR 1.42, 95% CI 0.55-3.64) and the number of cases of hepatocellular carcinoma detected (OR 0.90 95% CI 0.47-1.71).

Conclusion
There is currently insufficient evidence from randomized controlled trials to support routine screening for HCC in patients with chronic liver disease.
non-alcoholic fatty liver disease are major risk factors for HCC.\textsuperscript{3,4} It usually occurs in patients where liver disease has progressed to liver cirrhosis.\textsuperscript{5,6} The incidence of HCC is increasing in certain regions such as the United States (incidence men: 8.11/100,000, women: 2.47/100,000). This may be related to an increase in the prevalence of hepatitis C in the United States of America and increased immigration from high prevalence regions.\textsuperscript{7} The incidence in China decreased over the time period 2000-2014 (incidence men: 38.5/100,000, women: 14.5/100,000).\textsuperscript{8,9} The median age at diagnosis varies between populations. According to data from the Surveillance Epidemiology and End Results (SEER) program the median age of HCC diagnosis by region of birth in those living in the United States was 70-74 years for people from Europe, 65-69 years for people from Asia and 40-45 years for people from West Africa.\textsuperscript{10} HCC is often asymptomatic until an advanced stage. Early-stage tumors are more likely to be amenable to treatment and have better overall survival.\textsuperscript{10} There is no curative treatment for intermediate or late-stage tumors. Patients with symptomatic HCC have a 3-year survival of 8%.\textsuperscript{11} Patients with well-compensated liver disease may be considered for surgical resection. The survival rates are 58% at 3 years and 42% at 5 years in non-cirrhotic patients who undergo surgical resection.\textsuperscript{12} Liver transplantation is the main course of therapy in patients with cirrhosis of the liver. 5-year survival rates after liver transplantation are 69%, with a tumor recurrence rate of 7%.\textsuperscript{13} Patients with intermediate-stage hepatoma treated with transcatheter arterial chemo-embolization (TACE) have a median overall survival of 19-20 months.\textsuperscript{10}

Screening is the periodic application of a test in people at risk of developing a given disease to identify an early or latent stage. Screening for HCC fits many of the requirements of a screening program because it follows a known clinical course, has an early treatable stage, and is an important health problem. The rationale for screening is that patients at high risk of HCC, such as those with chronic liver diseases, can be identified and invited to participate. However, HCC often occurs in those with undiagnosed liver cirrhosis.\textsuperscript{14} Patients with undiagnosed liver cirrhosis or chronic liver disease will not be captured by screening programs founded on this rationale. Screening may result in harm. Invasive procedures may be performed based on a positive screening test. Liver biopsy has a 0.5% incidence of hematoa, a 0.1% risk of infection, and a 0.05% risk of death associated with liver biopsy.\textsuperscript{15} Negative psychosocial consequences occur in 5-20% of patients undergoing screening.\textsuperscript{16} In the absence of an effective means of identifying patients with an early-stage HCC however, most patients will die. Observational studies have shown that patients undergoing screening had earlier stage disease compared to patients who did not undergo screening.\textsuperscript{17,18} Non-randomized studies are subject to lead and length time bias. Lead-time is the time by which diagnosis is anticipated by screening with respect to the symptomatic detection of disease. Length time bias is an overestimation of survival duration due to the relative excess of cases detected that are slowly progressing. Even when non-randomized studies account for length time bias the effects of screening on survival vary with the assumed tumor doubling time.\textsuperscript{19} Survival time in one study was significantly longer in a screened group compared to a non-screened group when the tumor doubling time was assumed to be less than 90 days, but this was not evident if the tumor doubling time was assumed to be greater than 90 days.\textsuperscript{20} Most clinical guidelines recommend screening for HCC (Table 1). The purpose of this review is to determine if there is evidence from randomized controlled trials evaluating the efficacy of screening for HCC in patients with chronic liver disease. The objective was to determine if screening for HCC is effective in reducing mortality while being safe and acceptable to the screening population.

METHODS

In our review, only randomized controlled trials were considered for inclusion. Other types of studies were excluded. This was to minimize confounding due to selection bias, performance bias, and detection bias in non-randomized studies. Studies that evaluated the diagnostic accuracy of a test in the confirmation of suspected HCC were excluded. There were no restrictions based on language, publication status, or year of study. Only studies of Individuals with chronic liver disease were included for review. Studies including individuals with a history of HCC were excluded. To be included studies had to compare screening with no screening, compare different screening methodologies, or compare different screening intervals. Screening methodologies included for review were alpha-fetoprotein, ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI). A protocol for this systematic review and meta-analysis is available in Appendix 1 of the Online Supplementary Document.

Primary outcomes chosen were: i) all-cause mortality, and ii) quality of Life (any reported measure of quality of life was accepted). Secondary outcomes were: i) mortality due to HCC, ii) the number of cases of HCC detected, iii) adverse events. According to the International Conference of Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use Guideline for Good Clinical Practice (ICH 1997) an adverse event was defined as any untoward medical occurrence in the participant in the clinical trial, which does not necessarily have a causal relationship with this treatment.\textsuperscript{26} A serious adverse event is any untoward medical occurrence that results in death, is life-threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability or incapacity, or is a congenital anomaly/birth defect.\textsuperscript{26}

SEARCH METHODS FOR IDENTIFICATION OF STUDIES

Electronic searches were performed on the 20th of May 2018 using Medline (1946- May 2018), Embase (1974- May 2018), Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library (1974- May 2018) and the Web of Science Citation Index Expanded (1900- May 2018) were all searched according to the search strategy (Appendix 2 of the Online Supplementary Document). The databases used in the search strategy were chosen based on guidance contained within the Cochrane Handbook. It states that at least two databases should be used. They sug-
Table 1. Summary of current HCC screening guidelines

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Year of publication</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NICE20,21</td>
<td>2016</td>
<td>Ultrasound screening every 6 months for adults with cirrhosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ultrasound screening every 6 months with alpha-fetoprotein measurement for adults with chronic hepatitis B</td>
</tr>
<tr>
<td>AASLD22</td>
<td>2018</td>
<td>Ultrasound with or without AFP every 6 months</td>
</tr>
<tr>
<td>ACR23</td>
<td>2017</td>
<td>Ultrasound recommended except in subgroups with technical limitations where CT or MRI is recommended (e.g., obesity, fatty liver, nodular cirrhotic liver)</td>
</tr>
<tr>
<td>APASL24</td>
<td>2017</td>
<td>Ultrasound and AFP every 6 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AFP alone not recommended</td>
</tr>
<tr>
<td>EASL25</td>
<td>2018</td>
<td>Ultrasound every 6 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No recommendation for AFP</td>
</tr>
<tr>
<td>JSH26</td>
<td>2014</td>
<td>Extremely high-risk patients: ultrasound every 3-4 months, AFP every 3-4 months, Computed tomography/magnetic resonance imaging (optional) every 6-12 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High risk: ultrasound and AFP every 6 months</td>
</tr>
<tr>
<td>LAASL27</td>
<td>2014</td>
<td>In cirrhotic patients ultrasound should be performed every 6 months If quality ultrasound is not available, AFP level may be considered as a biomarker</td>
</tr>
</tbody>
</table>

NICE - National Institute of Clinical Excellence, AASLD - American Association for the Study of Liver Diseases, ACR - American College of Radiology, APASL - Asian Pacific Association for the Study of the Liver, JSH - Japan Society of Hepatology, LAASL - Latin American Association for the Study of the Liver, AFP - Alpha-fetoprotein

gest CENTRAL, Medline, or Embase. The handbook states that CENTRAL is the most comprehensive source of controlled trials. Clinical trial registries were also searched (clinicaltrials.gov, World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) search portal). The references of identified relevant randomized controlled trials were searched.

DATA COLLECTION AND ANALYSIS

The first systematic reviewer screened titles and abstracts for inclusion. The full articles of studies deemed eligible were then assessed by both first and second systematic reviewers for inclusion. All studies which were excluded were recorded with the relevant explanation. Any disagreements between the first and second systematic reviewers were resolved by discussion. It was planned that for any discrepancies that arose during data extraction where agreement could not be reached between the first and second reviewers, a third reviewer would be invited to resolve the discrepancy. Data extraction was performed independently by two reviewers. A standardized data extraction spreadsheet was used. Any disagreement between systematic reviewers about data extracted was resolved by discussion. The assessment of the risk of bias in included studies was performed using the Cochrane Collaboration’s recommended domain-based evaluation. The odds ratio was used as the measure of treatment effect for all-cause mortality, mortality secondary to HCC, and the number of cases of HCC. Using a random effect model the Mantel-Haenszel method was used to calculate an odds ratio as the measure of treatment effect. Studies that were cluster randomized controlled trials that had not accounted for this in their statistical analyses were adjusted for clustering using guidance from the Cochrane Consumers and Communication Group. An intracluster correlation coefficient (ICC) of 0.02 was used for adjustment. There were no intracluster correlation coefficients referenced in the included studies. An intracluster coefficient of 0.02 was chosen based on empirical estimates for randomized controlled trials with continuous and binary outcome measures which were generally less than 0.064 with a median value of 0.02.

Missing data was deemed to be acceptable if the explanation provided by the study authors was deemed likely not to impact on the study outcomes (e.g. when a participant with a history of HCC was randomized but subsequently excluded in studies which randomized participants in clusters which had pre-specified the exclusion of anyone with a history of HCC) or if the number missing was negligible. We accepted intention-to-treat analyses and modified intention-to-treat analyses but not per-protocol analyses.

Heterogeneity was assessed using forest plots of the measures of treatment effect. We looked for overlap of the confidence intervals of the measure of treatment effect and whether the point estimates of treatment effect were on the same side of the line of no effect. If all point estimates were on the same side of the line of no effect, we looked for the magnitude of those treatment effects. The chi-square test was used to assess for heterogeneity. A P-value of less than 0.10 was used as a cut-off for the detection of heterogeneity. The I² test for heterogeneity was also used. An I² of 30-60% was considered to represent moderate heterogeneity, 50-90% was considered to represent substantial heterogeneity and 75-90% was thought to be considerable heterogeneity. Interpretation of the I² also required concomitant interpretation of the chi-square test as well as the direction and magnitude of treatment effects point estimates and their confidence intervals.
If sufficient studies (10 or more) met the review inclusion criteria it was planned to generate a funnel plot and obtain the Egger test and Begg test to check for asymmetry as an indicator of reporting bias. Data were synthesized and analyzed using RevMan (Review Manager) 5.2.

Any heterogeneity detected between studies was investigated by considering both clinical and methodological factors. Clinical factors to consider were the baseline risk of HCC of study participants, whether the study was performed in primary care, secondary care, or both and the aetiology of cirrhosis of trial participants. The investigation of heterogeneity due to any potential methodological factors was performed by assessing the risk of bias in each study. For any heterogeneity detected we planned on changing the measure of treatment effect from odds ratio to risk difference to determine if it altered heterogeneity. It was planned to perform a subgroup analysis for all outcomes of participants who had hepatitis B or hepatitis C where studies reported the necessary data to explore any heterogeneity and determine the primary and secondary outcomes in these subgroups. It was planned to perform a sensitivity analysis to examine the effect of non-hepatitis B or C participants on the review outcomes where studies reported the necessary data. The quality of evidence was assessed using guidance from the GRADE handbook for grading the quality of evidence and strength of recommendations and GRADEpro software.34,35

RESULTS

RESULTS OF THE SEARCH

We identified a total of 7753 references through the electronic searches of the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library (n = 709), MedlineE (n= 4255), Embase (n = 720), and Web of Science Citation Index Expanded (n =2069) (flow diagram Figure 1). We excluded 152 duplicates. We excluded 7464 records through screening of references. Sixteen articles were retrieved for further assessment. Of the 16 articles, we excluded seven. In total, nine publications describing six studies fulfilled the inclusion criteria. No references were identified through searching reference lists of the identified studies.

INCLUDED STUDIES

Included studies are shown in Table 2 (summarised from Appendix 3 Table 2 in the Online Supplementary Document).

EXCLUDED STUDIES

Seven studies were excluded compromising of one protocol for a multi-center prospective clinical trial evaluating the effectiveness of ultrasound plus MRI in the detection of HCC,42 one randomized controlled trial evaluating the efficacy of contrast-enhanced ultrasonography against B-mode ultrasonography for the detection of HCC,43 two review articles,44,45 one randomized controlled trial comparing symptomatic treatment versus interferon alfa for the treatment of HCC,46 one retrospective cohort study which evaluated the usefulness of regular screening by ultrasonography and contrast-enhanced imaging for detection of HCC47 and one protocol for a trial evaluating contrast-enhanced ultrasound with perfluorobutane (Sonazoid) for the detection of HCC.48

RISK OF BIAS IN INCLUDED STUDIES

The risk of bias assessment was performed using the Cochrane Risk of Bias tool for Randomized Controlled Trials (Table 5).50
### Table 2. Characteristics of studies included

<table>
<thead>
<tr>
<th>Study</th>
<th>Year of publication</th>
<th>Study design</th>
<th>Country</th>
<th>Patients</th>
<th>Comparison</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al[36]</td>
<td>2004</td>
<td>Cluster randomized controlled trial</td>
<td>China</td>
<td>Adults aged 35-59 years with hepatitis B virus infection or a history of chronic hepatitis</td>
<td>Alpha-fetoprotein and ultrasound every 6 months versus no screening</td>
<td>Number of HCC detected, HCC related mortality, 5-year survival</td>
</tr>
<tr>
<td>Chen et al[37]</td>
<td>2003</td>
<td>Cluster randomized controlled trial</td>
<td>China</td>
<td>Men aged 30-69 who were chronic carriers of the hepatitis-B virus during the period 1989-1995</td>
<td>Alpha-fetoprotein every 6 months versus no screening</td>
<td>Number of HCC detected, HCC related mortality and all-cause mortality</td>
</tr>
<tr>
<td>Pocha et al[38]</td>
<td>2013</td>
<td>Randomised controlled trial</td>
<td>United States of America</td>
<td>Patients aged 18-70. Child’s A cirrhosis, documented cirrhosis, potential candidates for HCC treatment if diagnosed</td>
<td>Ultrasound and alpha-fetoprotein every 6 months versus annual CT plus alpha-fetoprotein every 6 months</td>
<td>Number of HCC detected, HCC related mortality and all-cause mortality</td>
</tr>
<tr>
<td>Sherman et al[39]</td>
<td>1995</td>
<td>Randomised controlled trial</td>
<td>Canada</td>
<td>Adults aged over 18 years old and hepatitis B surface antigen-positive for more than 6 months</td>
<td>Serial alpha-fetoprotein alone every 6 months versus alpha-fetoprotein in combination with ultrasound every 6 months</td>
<td>Incidence of HCC is reported but not reported separately for each arm of the study</td>
</tr>
<tr>
<td>Trinchet et al[40]</td>
<td>2011</td>
<td>Cluster randomized controlled trial</td>
<td>France</td>
<td>Adults aged 18 years or older, histologically proven cirrhosis related to hepatitis B, hepatitis C, alcohol or hereditary hemochromatosis, Child-Pugh A or B</td>
<td>Ultrasound every 3- months versus ultrasound every 6-months</td>
<td>Number of HCC detected, all-cause mortality, HCC related mortality, 5-year survival</td>
</tr>
<tr>
<td>Wang et al[41]</td>
<td>2013</td>
<td>Cluster randomized controlled trial</td>
<td>Taiwan</td>
<td>Adults aged 40 years or older, platelets less than or equal to 150x10^9/L, positive hepatitis B surface antigen or positive hepatitis C antibody</td>
<td>Ultrasound every 4- months versus ultrasound every 12-months</td>
<td>Number of HCC detected, tumor size, 4-year survival</td>
</tr>
</tbody>
</table>

HCC – Hepatocellular carcinoma CT – Computed tomography
Figure 1. PRISMA flow diagram of search results.

7753 records identified from database and registry searching

152 duplicate records removed

7601 records retained after duplicates removed

7464 records removed after screening of titles

137 records retained for abstract review

121 records removed after abstract review

16 records retained for full text eligibility review

7 records removed because they did not meet inclusion criteria:
(2) review articles
(2) study protocols
(1) retrospective cohort study
(1) trial evaluating treatments for hepatitis C
(1) trial comparing different modalities of ultrasound

9 full text articles retained for qualitative synthesis

6 full text articles retained for quantitative synthesis
<table>
<thead>
<tr>
<th>Bias</th>
<th>Zhang et al&lt;sup&gt;36&lt;/sup&gt;</th>
<th>Chen et al&lt;sup&gt;37&lt;/sup&gt;</th>
<th>Pocha et al&lt;sup&gt;38&lt;/sup&gt;</th>
<th>Sherman et al&lt;sup&gt;29&lt;/sup&gt;</th>
<th>Trinchet et al&lt;sup&gt;40&lt;/sup&gt;</th>
<th>Wang et al&lt;sup&gt;41&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Unclear risk</td>
<td>Low risk</td>
<td>Low risk. A computer-generated random number list was used to allocate and randomize subjects</td>
<td>Unclear risk</td>
<td>Low risk. Randomization was computer-generated, a permuted block design was used</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Low risk. Allocation was concealed using a centralized phone procedure to the data-management center.</td>
<td>Unclear risk*</td>
</tr>
<tr>
<td>Blinding of participants and personnel (performance bias)</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
</tr>
<tr>
<td>Blinding of outcome assessment (detection bias)</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>Unclear risk*</td>
<td>High risk. Outcome assessors not blinded</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>Unclear risk*</td>
<td>High risk. In the screening group, 400 subjects (10.8%) were tested at enrolment only and did not attend for further screening. It is unclear how these were followed up.</td>
<td>High risk. There were post-randomization drop-outs. Non-liver related death (6), Non-adherence to the protocol (12), withdrew from participation (8), followed by transplant center (10), others (8)</td>
<td>High risk. 72 subjects did not return for the first follow up screening visit. During the study period, an additional 182 subjects withdrew from the screening program having completed at least one follow up visit.</td>
<td>High risk. Compliance was estimated as inadequate in 143 (11.9%) of patients.</td>
<td>Unclear risk*</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>High risk. Different reports of this study exist. All-cause mortality not reported.</td>
<td>Low risk. Detection rates, the characteristics of the screening test and the outcome with respect to incidence, stage, survival and mortality in the screened and control group were reported.</td>
<td>Low risk. Most important outcomes were reported</td>
<td>Low risk. No reporting of all-cause mortality and disease-specific mortality in individual arms however the authors state the study was not designed or powered to detect these outcomes, it was designed as a feasibility study.</td>
<td>Low risk. Most important outcomes were reported</td>
<td>High risk. Did not report on all-cause mortality or disease-specific mortality</td>
</tr>
<tr>
<td>Other bias</td>
<td>High risk. Follow up in the control and screening group was not the same.</td>
<td>High risk. Alpha-fetoprotein testing performed using RPHA method which has a low sensitivity for detecting HCC.</td>
<td>High risk. Supported by a grant from Schering Canada Inc. Contains patients enrolled from hospitals and the community of unclear proportions.</td>
<td>High risk. Supported by a grant from Schering Canada Inc. Contains patients enrolled from hospitals and the community of unclear proportions.</td>
<td>High risk. Supported by a grant from Schering Canada Inc. Contains patients enrolled from hospitals and the community of unclear proportions.</td>
<td>High risk. Supported by a grant from Schering Canada Inc. Contains patients enrolled from hospitals and the community of unclear proportions.</td>
</tr>
</tbody>
</table>

HCC – Hepatocellular carcinoma  
RPHA= Reversed passive hemagglutination assay  
*Unclear risk stated when information to judge the risk of bias was not available in the report.
ALLOCATION (SELECTION BIAS)

Information regarding allocation concealment was not available in the studies by Zhang et al,36 Chen et al,37 Pocha et al,38 Sherman et al,39 Wang et al.41 The risks of selection bias in these studies are therefore unclear. Allocation concealment in the study by Trinchet et al.,40 was achieved using a centralized phone procedure to the data-management centre. The risk of selection bias in this study is low.

BLINDING (PERFORMANCE BIAS AND DETECTION BIAS)

In the study by Wang et al.41 the public health nurses followed up patients with newly detected nodules with regard to survival. It is unclear if the nurses were blinded as to which group the patients belonged to. The risk of performance bias here is unclear. Information regarding blinding was not available in the studies by Zhang et al36 Chen et al,37 Pocha et al,38 Sherman et al,39 Trinchet et al.40 The risks of performance bias and detection bias in these studies, as a result, are unclear.

INCOMPLETE OUTCOME DATA (ATTRITION BIAS)

In the study by Chen et al,37 there was considerable attrition bias. 400/3712 (10.8%) subjects in the intervention group did not attend for any further follow up after their enrolment screening alpha-fetoprotein. It is unclear what follow up if any, these subjects had. Compliance was poor in the intervention group. 1070/3712 (28.8%) of participants were considered to not have complete compliance, meaning they missed 1 or more of their scheduled screening examinations. As a result, the risk of bias in this study is high. Information regarding incomplete outcome data was not available in the studies by Zhang et al,36 Pocha et al,38 Sherman et al,39 Trinchet et al.40 Wang et al.41 The risks of bias due to incomplete outcome data are unclear in these studies.

SELECTIVE REPORTING (REPORTING BIAS)

In the studies by Chen et al57 and Pocha et al38 detection rates, the characteristics of the screening test and the outcomes with respect to incidence, stage, survival, and mortality in the screened and control group were reported. The risks of selective reporting bias in these studies are low. In the study by Sherman et al39 the authors have not reported on all-cause mortality and disease-specific mortality in the screening group compared to the control group however they state that this study was not designed to evaluate these outcomes and that this trial was a preliminary study preparatory to a larger study. The risk of selective reporting bias in this study is low. Trinchet et al40 reported on most important outcomes and the risk of reporting bias is low. Wang et al41 reported on overall survival in patients with HCC but did not report on all-cause mortality. The risk of reporting bias in this study is therefore high. Similarly, Zhang et al36 did not report all-cause mortality and the risk of reporting bias is also high.

OTHER POTENTIAL SOURCES OF BIAS

Four of the studies, Zhang et al36 Chen et al,37 Trinchet et al.38 and Wang et al.,41 were cluster randomized controlled trials with no adjustment for clustering or use an intracluster correlation coefficient reported in their statistical analysis. There may be a correlation of observations within the clusters and the resulting standard errors of the effect of the interventions will be too small, along with the confidence intervals and P-values.

Effects of interventions (Appendix 5 Quality of Evidence and Summary of Findings and Appendix 6 Forest Plots in the Online Supplementary Document).49

Zhang et al compared screening with six-monthly alpha-fetoprotein to no screening.36 There was no difference in the number of cases of HCC detected across a range of intracluster coefficients (ICC) (ICC 0.02; OR 1.30; 95% CI 0.81-2.08) (ICC 0.05; OR 1.30; 95% CI 0.68-2.48) (ICC 0.1; OR 1.30; 95% CI 0.10-3.07). In the same study there was no evidence of difference in HCC related mortality using a range of intracluster coefficients (ICC 0.02; OR 0.60; 95% CI 0.31-1.15), (ICC 0.05; OR 0.60; 95% CI 0.25-1.45) (ICC 0.1; OR 0.60 95% CI 0.19-1.94) (Figure 2). The number of clusters in this study was unclear. The authors report there were more than 300 clusters. When the number of clusters was assumed to be 350 there was no statistically significant difference in either the mortality due to HCC (ICC=0.02; OR 0.5995%; CI 0.32-1.11) or the number of cases of HCC detected (ICC=0.02; OR 1.30; 95% CI 0.83-2.04). When the number of clusters was assumed to be 399 there was no statistically significant difference in either the mortality due to HCC (ICC=0.02; OR 0.59; 95% CI 0.32-1.09) or the number of cases of HCC detected (ICC 0.02; OR 1.30 95% CI 0.83-2.04).

In the study by Chen et al57 there were 5581 participants. There was no statistically significant difference in all-cause mortality when screening with six-monthly alpha-fetoprotein was compared to a single alpha-fetoprotein check (OR 1.02; 95% CI 0.65-1.60) in this study. There was no evidence of difference in mortality due to HCC (OR 1.01; 95% CI 0.57-1.78) and there was no evidence to support any difference in the number of cases of HCC detected in the same study (OR 1.11 95% CI 0.64-1.93).

There was no evidence of difference in all-cause mortality when ultrasound was compared to CT (OR 0.81; 95% CI 0.26-2.53) in the study by Pocha et al, which made this comparison with 163 participants.38 In the same study, there was no statistically significant difference between the two groups in terms of HCC related mortality (OR 0.67 95% CI 0.20-2.20) or the number of cases of HCC detected (OR 1.09 95% CI 0.40-2.99). A subgroup analysis of the number of cases of HCC detected in participants with hepatitis C did not show any evidence of difference between the two groups (OR 1.11; 95% CI 0.40-3.05) (Appendix 6 Figure 17 Online Supplementary Document). There were no cases of HCC in participants with hepatitis B in this study, so it was not possible to do a subgroup analysis of hepatitis B participants.

Wang et al and Trinchet et al compared more frequent screening versus less frequent screening using ultrasound.40,41 When adjusted for clustering, there was no evidence of difference in all-cause mortality when screening...
with ultrasound every three months was compared to screening with ultrasound every six months (OR 0.86; 95% CI 0.56-1.32) in the study by Trinchet et al. There was also no evidence of difference in HCC related mortality (OR 1.42; 95% CI 0.55-3.64) in the same study. There was no significant heterogeneity between the two studies as demonstrated by a significant overlap of the 95% confidence intervals ($I^2 = 0\%$, $\chi^2 = 0.67$, df=1 ($P=0.41$) (Appendix 6 Figure 20, Online Supplementary Document). Meta-analyses of these studies did not find any statistically significant difference in the number of cases of HCC detected by more frequent screening compared to less frequent screening (OR 0.90, 95% CI 0.47-1.71).

A subgroup analysis of patients with hepatitis B in the study by Wang et al did not show any evidence of difference in the number of cases of HCC detected (ICC 0.02; OR 1.45; 95% CI 0.51-6.68). Similarly, for patients with hepatitis C in the same study, a subgroup analysis did not show any evidence of difference in the number of cases of HCC detected (ICC 0.02; OR 1.42; 95% CI 0.23-8.61).

No studies reported on the quality of life or adverse events. It was not possible to perform a subgroup analysis for all outcomes of participants who had hepatitis B or hepatitis C because studies which included these patients had not reported the necessary data to explore any heterogeneity and determine the primary and secondary outcomes in these subgroups. It was not possible to perform a sensitivity analysis to examine the effect of non-hepatitis B or C participants on the review outcomes. There were insufficient studies to meet the review inclusion criteria as planned to generate a funnel plot and obtain the Egger test and the Begg test to check for asymmetry as an indicator of reporting bias.

QUALITY OF THE EVIDENCE

The number of HCC related deaths and the number of cases of HCC detected in the study by Zhang et al was small. The confidence intervals were wide suggesting imprecision. There was no inconsistency or indirectness. The quality of evidence in this study is therefore moderate. In the studies by Chen et al, Pocha et al, Wang et al, and Trinchet et al there was a high risk of attrition bias because of poor compliance with follow up screening tests. The confidence intervals were wide and would likely include a minimal clinically important difference for a trial comparing screening frequencies for HCC or comparing screening methodologies. There was no inconsistency or indirectness in these four studies. For these reasons, the quality of evidence from these studies is low. There were insufficient studies to formally assess for publication bias with a funnel plot.
DISCUSSION

There was no evidence from this review to support screening for HCC in patients with chronic liver disease to reduce mortality. A significant finding was the effect of adjusting for clustering on study outcomes. Understandably, clustering is used in randomized controlled trials for logistical and feasibility reasons. Correlation of observations occurs within clusters and this must be accounted for in the statistical analysis. This is because participants within clusters are more likely to be similar to each other than participants in other clusters. In the study by Zhang et al., they reported a 37% reduction in the HCC related mortality.\textsuperscript{36} When adjusted for clustering there was no evidence of difference in the HCC related mortality ratios. This finding remained true across a range of intracluster correlation coefficients and across a range of assumed number of clusters (the number of clusters in the study was reported as “more than 300”, we performed our analysis assuming there were 301,350 and 399 clusters respectively). A potential bias in the review process was the use of an intracluster coefficient of 0.02 to adjust for clustering. We could not find previous similar studies to give guidance on the estimation of a suitable intracluster coefficient. The choice of too small an intracluster coefficient can have a substantial impact on confidence interval width. We demonstrated that the use of a range of intracluster coefficients (0.02, 0.05, 0.1) did not change the outcome. Of note, there was a lack of studies that included quality of life and adverse events as outcomes. These are important outcomes when considering a screening test. A screening test with significant adverse events or reductions in quality of life may not be acceptable to the target population.

It is important to consider the treatments patients received when discussing the benefits of screening and its impact on mortality. In the study by Zhang et al.,\textsuperscript{36} no reduction in HCC related mortality was seen despite the screening group detecting more early-stage tumors and receiving more liver transplantation compared to the control group (46.5% vs. 7.5%). Similarly, in the study by Trinchet et al.\textsuperscript{41} more patients with HCCs <10mm were detected. More patients received liver transplantation when screened every 3 months with ultrasound compared to 6-monthly screening in this trial (18.9% vs. 4.3%). This did not translate to a reduction in all-cause mortality or HCC related mortality. Wang et al.\textsuperscript{40} found that when comparing screening every 4 months versus every 12 months they detected more early-stage tumors and it resulted in more curative therapy being given but did not confer a benefit in survival in the 4-year follow-up. From these trials it appears despite early diagnosis and treatment, mortality was unchanged, suggesting treatments were not an effective means of cure in HCC detected by screening.

Nil evidence of effect is not evidence of no effect. Reasons why the studies included in this review may not have detected an effect if it was truly there may be related to the patients enrolled. Within some studies, for example, Chen et al.\textsuperscript{37} patients with non-cirrhotic chronic liver disease, and cirrhotic liver disease were included in each arm. The patients with non-cirrhotic liver disease would be at low risk of HCC and may have diminished any potential significant findings in cirrhotic liver disease patients. The studies by Chen et al.\textsuperscript{37} and Zhang et al.\textsuperscript{36} were performed in what the authors note to be high prevalence settings. This may limit the applicability of evidence from these studies to lower prevalence settings. However, given there was a lack of evidence of effect in a high prevalence setting it is unlikely to materialize in a low prevalence setting.

A limitation of this study may have been the choice of all-cause mortality as the primary outcome. All-cause mortality would be a standard primary outcome for studies evaluating screening tests. However, patients with chronic liver disease die for many reasons other than HCC, and often before its development if it was going to be an occurrence. As such, a screening test for HCC may not result in any significant reduction in all-cause mortality. A second limitation was the small number of studies that met the inclusion criteria (n=6). The strengths of this review include rigorous methodology and adherence to PRISMA guidelines for reporting on systematic reviews (Appendix 4 Online Supplementary Document).

Current NICE guidelines state “offer ultrasound (with or without measurement of serum alpha-fetoprotein) every 6 months as surveillance for HCC for people with cirrhosis who do not have hepatitis B virus infection” and “Perform 6-monthly surveillance for HCC by hepatic ultrasound and alpha-fetoprotein testing in people with significant fibrosis (METAVIR stage greater than or equal to F2 or Ishak stage greater than or equal to 3) or cirrhosis”.\textsuperscript{20} This recommendation is based on evidence from a systematic review and meta-analysis performed by the NICE guidelines’ development group. There is very low-quality evidence from two retrospective cohort studies (n=351) which indicated a clinical benefit of surveillance for survival when analyzed using time-to-event data.\textsuperscript{50,51} They also evaluated annual versus six-monthly surveillance and found only very low-quality evidence from one observational study (n=649) which indicated a clinical benefit of six-monthly surveillance for survival.\textsuperscript{52} Low-quality evidence from the same study indicated a clinical benefit of six-monthly surveillance for the detection of HCC beyond a very early stage. Moderate quality evidence from one randomized controlled trial (n=1278) indicated a clinical benefit of three-monthly surveillance for survival and HCC occurrence when compared to six-monthly surveillance.\textsuperscript{40} Evidence ranging from very low to moderate quality from the same randomized controlled trial indicated no clinical difference in HCC diameter >50 mm at detection, the number of HCC nodules detected, or the HCC stage at detection. A Cochrane review performed in 2012 examined whether alpha-fetoprotein and/or ultrasound was effective for screening for HCC.\textsuperscript{53} They concluded there was insufficient evidence. The purpose of our systematic review was to determine if there was any high-quality evidence to support or refute screening for HCC, screening frequencies, and screening methodologies. The systematic review and meta-analysis we have conducted included studies that included patients with all stages of chronic liver disease and was not limited to alpha-fetoprotein and ultrasound as the only interventions. Conducted in 2018 it is more up to date and has broader inclusion criteria. Of note, all studies identified in this review were published after 1995 and all were in English.
An increase in worldwide use of the hepatitis B vaccine and a reduction in global hepatitis C burden due to recent treatment advances may result in reductions in the global incidence of HCC. According to the WHO at the end of 2017, 187 countries had the hepatitis B vaccine nationwide. Global coverage with the uptake of all 3 vaccines is estimated at 87%.54 Should the incidence of HCC decline, in the absence of a high-quality evidence base to support screening, then resource-limited healthcare programs may opt to allocate resources to higher impact evidence-based interventions for other conditions of importance to their population. Screening has the potential to have the greatest benefit in high-risk populations such as those with hepatitis B. There is a need for new effective evidence-based HCC screening methods and treatments to reduce disease burden in these high-risk populations.

Screening according to current guidelines also requires resource allocation for laboratory measurement and reporting of alpha-fetoprotein and radiology staffing and infrastructure to perform ultrasonography. Consideration should be given to the possibility that screening for HCC is potentially harmful to patients and consumers significant healthcare resources. Screening may even detract from other aspects of chronic liver disease care. In the absence of high-quality evidence evaluating adverse events and the psychosocial impact of screening for HCC, it may not be correct to advocate for screening on a population level. Current guidelines that recommend screening may even act as a deterrent to such a high quality randomized controlled trial being undertaken. A feasibility study in Australia found that 99.5% of people would not participate in a randomized controlled trial comparing screening for HCC to no screening. Of this 88% elected for a nonrandomized screening program.55 Notably in this study important information was omitted from the decision aid presented to patients. There was no mention of the potential for invasive investigations such as liver biopsy being performed based on the results of screening tests. There was no mention of the risks associated with invasive investigations such as liver biopsy. Only non-invasive further investigations such as CT or MRI were discussed. There was no discussion of the potential consequences of radiation exposure with CT or the consequences of finding radiological lesions elsewhere such as “incidentalomas”. We feel the researchers also did not point out to patients the best available evidence on screening for HCC comes from the randomized controlled trial by Zhang et al36 which, despite its limitations, did not find evidence that screening reduced all-cause mortality or HCC related mortality (when adjusted for clustering). We also feel the potential for negative psychosocial consequences with screening could have been discussed in greater detail. The provision of this information which is relevant when discussing the disadvantages of screening may have changed the outcome of the feasibility study.

HCC is an important health problem globally with no curative treatment for those diagnosed with intermediate or late-stage tumors. If a curative treatment for HCC emerges then the question of whether screening should be performed deserves a large well-conducted randomized controlled trial to evaluate its effect on all-cause mortality, HCC related mortality, adverse events, quality of life and cost-effectiveness.

CONCLUSIONS

There is no evidence of effect from randomized controlled trials that screening for HCC reduces all-cause mortality, HCC related mortality, or results in more HCCs being detected. There is a need for a high quality randomized controlled trial which should include adverse events and quality of life as outcomes.

Funding: None.

Authorship contributions: JOC performed the systematic review design, search, data extraction, analysis, and manuscript writing. SR performed data extraction, analysis, and manuscript writing.

Competing interests: The authors completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available upon request from the corresponding author), and declare no conflicts of interest.

Correspondence to:
Dr. James O’Connell, MB BCh BAO, MSc
Department of International Health and Tropical Medicine,
Royal College of Surgeons in Ireland
Beaux Lane House
Lower Mercer Street
Dublin 2
Dublin
Ireland
jamesoconnell@rcsi.com

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY-4.0). View this license’s legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.
REFERENCES


49. Stahl-Timmins W. *Graphical Overview for Evidence Reviews (GOIER)*. School of Medicine, University of Exeter; 2011.


SUPPLEMENTARY MATERIALS

Online Supplementary Document