

Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis

A. SINGAL*, M. L. VOLK*, A. WALJEE*, R. SALGIA*, P. HIGGINS*, M. A. M. ROGERS†
& J. A. MARRERO*

*Division of Gastroenterology and

†Division of General Medicine,
Department of Internal Medicine,
University of Michigan, Ann Arbor,
MI, USA

Correspondence to:

Dr J. A. Marrero, Division of
Gastroenterology, Department of
Internal Medicine, University of
Michigan, 3912 Taubman Center, SPC
5362, Ann Arbor, MI 48109, USA.
E-mail: jmarrero@umich.edu

Publication data

Submitted 8 January 2009

First decision 30 January 2009

Resubmitted 16 February 2009

Resubmitted 1 April 2009

Accepted 2 April 2009

Epub Accepted Article 8 April 2009

SUMMARY

Background

A majority of studies investigating the accuracy of ultrasound for detecting hepatocellular carcinoma (HCC) do not reflect how this test is used for surveillance vs. diagnosis.

Aim

To determine the performance characteristics of surveillance with ultrasound for the detection of HCC, particularly early HCC as defined by the Milan criteria.

Methods

A systematic literature review using the MEDLINE and SCOPUS databases yielded six studies that evaluated the accuracy of ultrasound for HCC at any stage and 13 studies that were specific to early HCC.

Results

Surveillance ultrasound detected the majority of tumours before they presented clinically, with a pooled sensitivity of 94%. However, ultrasound was less effective for detecting early HCC with a sensitivity of 63%. Alpha-fetoprotein provided no additional benefit to ultrasound. Meta-regression analysis demonstrated a significantly higher sensitivity for early HCC with ultrasound every 6 months than with annual surveillance. Current studies have limitations such as verification bias and are of suboptimal quality.

Conclusions

Surveillance with ultrasound demonstrates limited sensitivity for early HCC, although this may be improved by testing at 6-month intervals. Currently available evidence evaluating surveillance ultrasound has significant limitations and future studies are necessary to determine optimal surveillance methods for early HCC.

Aliment Pharmacol Ther 30, 37–47

INTRODUCTION

Hepatocellular carcinoma (HCC) is the fifth most common tumour worldwide, with an increasing incidence in Europe and the US.¹⁻³ It is currently the third leading cause of cancer-related deaths worldwide, resulting in over 500 000 deaths/year.²⁻⁴ Cirrhosis, particularly when related to viral hepatitis, is the most notable risk factor for HCC and is found in nearly 80-90% of cases.^{1, 5} Despite advances in technology and available treatments, there has been little improvement in survival with a 5-year survival of 5% in 1996 compared to 4% in 1985.^{6, 7}

The stage of disease at the time of diagnosis largely determines the effectiveness of treatment. The treatment of advanced HCC continues to be primarily palliative, with curative options only available for early HCC. In patients with preserved hepatic function, no evidence of portal hypertension, and single asymptomatic tumours <5 cm in diameter, surgical resection has provided 5-year survival rates of 70%.⁸ Similarly, liver transplantation for tumours meeting the Milan criteria (one nodule <5 cm or three nodules each <3 cm in diameter) has a 5-year survival rate of nearly 74%.⁸⁻¹⁰ In patients with early-stage disease who are not amenable to resection or transplantation, radiofrequency ablation has demonstrated 5-year survival rates of 37%.⁸ These survival rates are in stark contrast to the average survival of <1 year reported for advanced HCC.¹¹ Unfortunately, less than 30% of patients are diagnosed early enough to meet criteria for resection or transplantation.¹²

Surveillance strives to detect HCC at an early stage when it is amenable to curative therapy to reduce mortality.¹³ Currently, surveillance for HCC is widely accepted among high-risk populations, most notably patients with cirrhosis. Current guidelines from the American Association for the Study of Liver Disease (AASLD) and the European Association for the Study of the Liver recommend surveillance of cirrhotic patients with ultrasound with or without alpha-feto-protein (AFP) every 6-12 months. Although these tests have been extensively studied for the purpose of diagnosis,^{14, 15} the results of those studies do not reflect the performance of these tests in clinical practice. Few trials have prospectively evaluated the utility of ultrasound and AFP as a surveillance test. Given the lack of a randomized trial of HCC surveillance among patients with cirrhosis, a meta-analysis is needed to estimate more precisely the accuracy of ultrasound

and AFP as surveillance tests for HCC. The aim of our study was to determine the pooled sensitivity, specificity and diagnostic odds ratio (OR) of ultrasound and AFP for the detection of HCC, particularly early HCC, during surveillance.

METHODS

Literature search

We searched the MEDLINE and SCOPUS databases from database inception through 1 July 2007 with the following keyword combinations: hepatocellular carcinoma AND screening, hepatocellular carcinoma AND surveillance, hepatocellular carcinoma AND cirrhosis or hepatocellular carcinoma AND ultrasound. Manual searching of the reference lists from applicable studies was performed to identify any studies through 1 July 2007 that may have been missed by the electronic search.

Study selection

Two investigators (A.S. and R.S.) independently reviewed the publications titles identified by the search strategy. If the applicability of an article could not be determined by title or abstract alone, the full text was reviewed. The articles were independently checked for possible inclusion and any disagreements were resolved through consensus with a third reviewer (M.V.).

Study inclusion and exclusion criteria

Studies were included for analysis if they (i) utilized ultrasound, with or without concomitant AFP, for HCC surveillance in cirrhotic patients; (ii) performed the tests prospectively in a series of patients and (iii) reported the number of discovered HCC, number of early HCC and number of missed lesions.

Prospective studies performed among a noncirrhotic cohort, such as patients with chronic hepatitis, were excluded from the meta-analysis. If the study cohort included both patients with cirrhosis and chronic hepatitis, only data regarding cirrhotic patients were included. Studies that evaluated surveillance techniques other than ultrasound, e.g. computed tomography (CT) scan or new biomarkers, were excluded. If multiple techniques were used for surveillance, only lesions discovered by ultrasound and/or AFP were

recorded as true positives; HCC nodules seen only by other techniques were counted as missed lesions. Studies using sequential test combinations, such as ultrasound testing in patients based on AFP levels, were excluded; information bias from the initial study could have unpredictable effects on the ultrasound operating characteristics. Studies evaluating ultrasound for screening instead of surveillance were not included in the analysis. Screening was defined as the one-time application of the test to detect a previously undiagnosed lesion, whereas surveillance was defined as the repeated use of the test at a set interval over time. Studies that failed to detail the number of false-negative results, i.e. patients with missed lesions, were excluded given that lack of this information precluded sensitivity calculations. Additional exclusion criteria included non-English language, nonhuman data, lack of original data and incomplete reports including meeting abstracts. If duplicate publications used the same cohort of patients, the data from the most recent manuscript were included.

Data extraction

Two reviewers (A.S. and R.S.) independently reviewed and extracted the required information from eligible studies using standardized forms. A third investigator (M.V.) was available to resolve any discrepancies between the two sets of extracted data. The data extraction form included the following study design items: geographical location and date of study, characteristics and size of study cohort, inclusion and exclusion criteria, surveillance methods, surveillance interval, duration of follow-up and 'gold-standard' methods for confirmation of HCC. In addition, the extraction form recorded the following primary data: number of HCC discovered during surveillance (true positives), number of false positives, number of missed lesions (false negatives) and number of true negatives. The method of tumour detection, i.e. ultrasound or AFP, was recorded for each tumour. We recorded the proportion of HCCs discovered at an early stage as defined by Milan criteria: one nodule <5 cm or three nodules each <3 cm in diameter, without gross vascular invasion. Some studies were excluded if they otherwise defined early-stage disease (e.g. unifocal lesion <3 cm) and there were insufficient data to determine the number of patients meeting Milan criteria.

Two independent reviewers (A.S. and R.S.) assessed the study quality by a modified checklist based upon

the Quality Assessment Tool for Diagnostic Accuracy (QUADAS) guidelines¹⁶ with discrepancies resolved by a consensus reviewer (M.V.).

Statistical analysis

The first aim of this study was to determine the sensitivity and specificity of surveillance ultrasound to detect HCC at any stage. The second aim was to determine the sensitivity of surveillance ultrasound to detect early HCC and if there is any additional benefit of concurrently checking AFP. For each individual study, per-patient sensitivity, per-patient specificity and diagnostic ORs with 95% confidence intervals were calculated. Pooled estimates of each calculation were then computed using STATA 10 (StataCorp, College Station, TX, USA). Estimates of effect were pooled using the DerSimonian and Laird method for a random effects model.

The heterogeneity of diagnostic test parameters was initially evaluated graphically by examination of forest plots and then statistically by the chi-squared test of heterogeneity and the inconsistency index (I^2). A chi-squared P -value <0.05 or I^2 values >50% are consistent with the possibility of substantial heterogeneity.^{17, 18} Sensitivity analysis, in which one study is removed at a time from the model, was performed to determine if there was possible undue influence of a single study.¹⁹ Among the studies assessing surveillance for early HCC, publication bias was initially evaluated graphically by funnel plot analysis and then statistically using Begg's test.²⁰ A symmetric funnel plot would help rule out the possibility of small studies that were not published due to unfavourable results. A summary receiver operator characteristics curve (SROC curve) was constructed to illustrate the distribution of sensitivities and specificities.²¹⁻²³ The area under the curve (AUC) was computed, with perfect tests having an AUC of 1 and poor tests having an AUC close to 0.5.²⁴

Subset analysis was planned for the detection of early HCC for predefined subsets of studies based on (i) use of concurrent surveillance tests; (ii) length of surveillance interval; (iii) location of study; (iv) date of study; (v) percentage of viral hepatitis patients; (vi) percentage of Child's A cirrhotics; (vii) incidence of HCC and (viii) length of follow-up. Meta-regression, using the Knapp-Hartung method for variance estimation,²⁵ was performed to investigate possible sources of heterogeneity in sensitivity measures across the

studies. Models were then refitted using a Monte-Carlo permutation test with 10 000 replications and extended to assess multiplicity.²⁶

RESULTS

Studies

Upon review of the 8826 titles identified by the search strategies, 192 abstracts were further examined. Sixty-three publications underwent full-text review to determine their eligibility for the meta-analysis and 48 were excluded. Thirteen studies were excluded because they did not use ultrasound, two articles used ultrasound but not as a surveillance tool, 10 studies were not conducted among patients with cirrhosis, nine studies were retrospective, eight studies were excluded for lack of original data and eight studies had insufficient data for extraction. The remaining 13 studies were selected after meeting all applicable inclusion criteria (Figure 1). Six studies detailed the number of false-positive and false-negative lesions and were selected for the first part of the analysis in which the sensitivity and specificity of ultrasound to detect HCC at any stage were assessed (Table 1). All 13 studies were used in the second part of the analysis in which the sensitivity of ultrasound with and without AFP to

detect early HCC was assessed (Table 2). Only six of the 13 studies reported false positives for the detection of early HCC, thus limiting accurate evaluation of specificity with regard to early HCC. There was excellent agreement between the two reviewers for both parts of the analysis ($\kappa = 1.0$).

Ultrasound for detecting HCC at any stage

Six studies detailed the number of false-positive and false-negative lesions and were selected for the first part of the analysis in which the sensitivity and specificity of ultrasound to detect HCC at any stage were assessed. The included studies had significant heterogeneity ($\chi^2 = 12.8$, $P = 0.02$, $I^2 = 60.8\%$) and hence meta-analysis was not initially possible. Inspection of forest plots suggested that the Caturelli study²⁷ was an important outlier. The calculated OR (=38 239) from this study was significantly higher than that of other included studies and is inconsistent with what is seen in clinical practice. This could have been related in part to the higher rate of HCC in this study population, suggesting that the study population is different from that in the other studies. We performed sensitivity analysis and found that omission of the Caturelli study had a large effect on the overall estimate of the relative risk. After omission of this study, there was

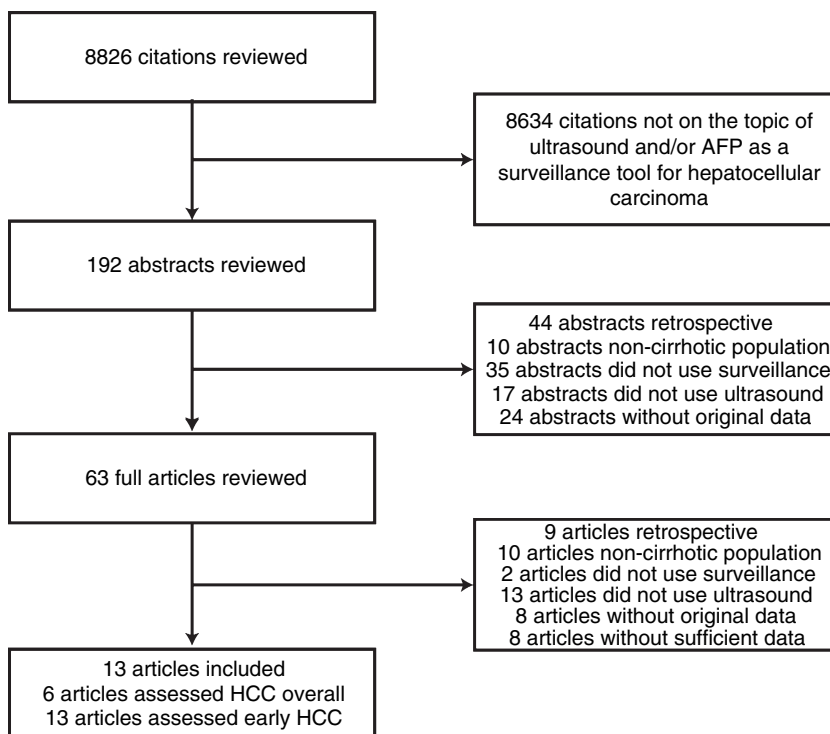


Figure 1. Map of the literature search and selection process. HCC, hepatocellular carcinoma; AFP, alpha-fetoprotein.

Table 1. Studies evaluating ultrasound for the detection of hepatocellular carcinoma at any stage

First author and study	Country	No. pts	Incidence of HCC (%)	Mean follow-up (months)	Surveillance interval (months)	HCC criteria
Kobayashi ³³	Japan	95	8.4	50.4	2–3	Histology or angiography
Pateron ³⁴	France	118	11.9	35.8	6	Histology or CT
Bolondi ⁴²	Italy	313	19.5	56	6	Histology or CT
Caturelli ²⁷	Italy	1827	14.7	43.1	4	Histology or AFP >200 with mass on imaging
Sangiovanni ⁴³	Italy	417	26.8	148	6–12	Histology, CT or angiography
Sangiovanni ⁴⁴	Italy	214	31.8	114	6–12	Histology or AFP >200 with mass on imaging

Pts, patients; HCC, hepatocellular carcinoma; AFP, alpha-fetoprotein.

Table 2. Studies evaluating ultrasound for the detection of early hepatocellular carcinoma

First author and study	Country	No. pts	Incidence of early HCC (%)	Mean follow-up (months)	Surveillance interval (months)	Criteria used to establish diagnosis of HCC
Kobayashi ³³	Japan	95	4.2	50.4	2–3	Histology or angiography
Arrigoni ²⁸	Italy	164	6.7	32.5	4–12	Histology or arteriography
Oka ³¹	Japan	140	19.2	41.1	2–3	CT or arteriography
Pateron ³⁴	France	118	2.5	36	6	Histology or CT
Cottone ²⁹	Italy	147	17.7	24	6	Histology
Zoli ³²	Italy	164	18.9	7–77	3–6	Biopsy or CT
Tradati ⁴⁵	Italy	40	5	48	12	Histology or AFP >400 with mass on imaging
Henrion ³⁰	Belgium	94	4.3	34	3–6	Histology, CT or MRI
Bolondi ⁴²	Italy	313	16	56	6	Histology or CT
Caturelli ²⁷	Italy	1599	16.5	43.1	4	Histology or AFP >200 with mass on imaging
Santagostino ⁴⁶	Italy	66	3	72	6–12	Histology, AFP > 400 with mass on ultrasound, CT or MRI
Sangiovanni ⁴³	Italy	417	13.2	148	6–12	Histology, CT or angiography
Sangiovanni ⁴⁴	Italy	214	15.9	114	6–12	Histology or AFP >200 with mass on imaging

Pts, patients; HCC, hepatocellular carcinoma; AFP, alpha-fetoprotein; CT, computed tomography; MRI, magnetic resonance imaging.

significant reduction in the heterogeneity ($\chi^2 = 5.8$, $P = 0.22$, $I^2 = 30.9\%$).

Repeat analysis after exclusion of the Caturelli study demonstrated a pooled sensitivity of 94% (95% CI: 83–98), a pooled specificity of 94% (95% CI: 89–97) and a pooled diagnostic OR of 232.7 (95% CI: 105.9–511.2) (Table 3, Figure 2). Using these pooled estimates for sensitivity and specificity, 82.1% of patients with a positive ultrasound would have HCC. Sensitivity analysis with the remaining five studies did not show any significant change in the relative risk with removal of

any other studies. There did not appear to be any publication bias by Beggs test ($P = 0.71$) or funnel plot analysis. SROC analysis demonstrated an AUC of 0.98 (95% CI: 0.96–0.99) suggesting high diagnostic accuracy (Figure 2).

Ultrasound for detecting early HCC

All 13 studies were used in the second part of the analysis in which the sensitivity of ultrasound with and without AFP to detect early HCC was assessed.

Table 3. Performance characteristics of ultrasound for the detection of hepatocellular carcinoma at any stage

First author and study	No. pts	No. HCC	Sensitivity (95% CI)	Specificity (95% CI)	Odds ratio (95% CI)
Kobayashi ³³	95	8	0.75 (0.35–0.97)	0.98 (0.92–1.0)	128 (15–1000)
Pateron ³⁴	118	14	0.79 (0.49–0.95)	0.96 (0.90–0.99)	91 (18–463)
Bolondi ⁴²	313	61	0.93 (0.84–0.98)	0.95 (0.91–0.97)	262 (82–833)
Caturelli ²⁷	1599	269	1.0 (0.99–1.0)	0.99 (0.98–0.99)	38 240†
Sangiovanni ⁴³	417	112	0.98 (0.94–1.0)	0.85 (0.81–0.89)	317 (76–1000)
Sangiovanni ⁴⁴	214	68	0.99 (0.92–1.0)	0.90 (0.84–0.95)	632 (81–1000)
Pooled estimates	2715	412	0.95 (0.89–0.98)	0.91 (0.76–0.97)	191.6 (89.2–411.8)
Pooled estimates*	1116	143	0.94 (0.83–0.98)	0.94 (0.89–0.97)	232.7 (105.9–511.2)

Pts, patients; HCC, hepatocellular carcinoma; AFP, alpha-fetoprotein.

* Pooled estimates after exclusion of outlier studies (Caturelli *et al.*²⁷).

† Confidence interval could not be determined given high upper limit.

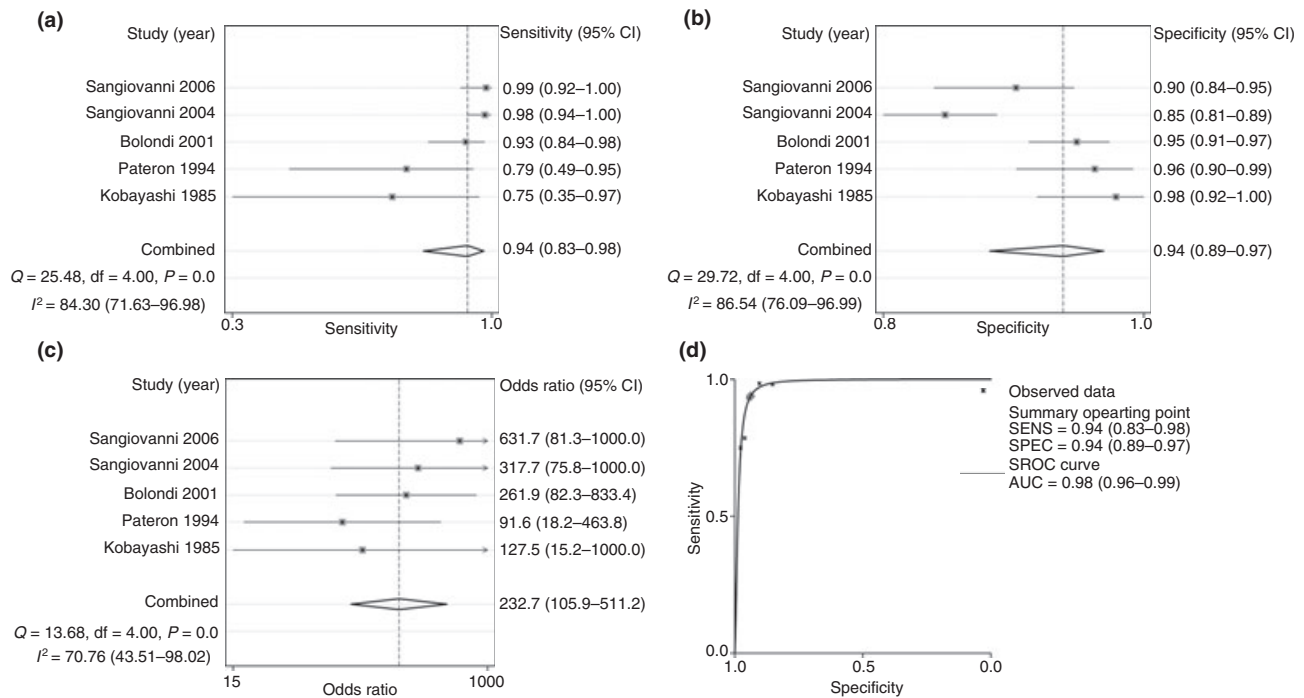


Figure 2. Performance characteristics of ultrasound for the detection of hepatocellular carcinoma (HCC) at any stage: (a) forest plot for the sensitivity of ultrasound to detect HCC at any stage; (b) forest plot for the specificity of ultrasound to detect HCC at any stage; (c) forest plot for the odds ratio of ultrasound to detect HCC at any stage; (d) summary receiver operator curve plot for the detection of HCC at any stage by ultrasound. Q , chi-squared test of heterogeneity; I^2 , inconsistency index; AUC , area under the curve.

Only six of the 13 studies reported false positives for the detection of early HCC, thus limiting accurate evaluation of specificity with regard to early HCC. The 13 studies evaluating early HCC had a significant degree of heterogeneity ($\chi^2 = 195.8$, $P < 0.001$,

$I^2 = 94\%$). Inspection of the forest plots confirmed a large variation in pooled estimates with six studies^{27–32} having ORs >300 (Figure 3). Sensitivity analysis suggested that the studies by Caturelli and Sangiovanni were important outliers with large effects on the overall

estimate of the OR. Omission of the Caturelli study resulted in substantial improvement in the heterogeneity ($\chi^2 = 84.6$, $P < 0.001$, $I^2 = 87\%$). The OR (=3849) from the Caturelli study was significantly higher than that of other included studies, suggesting marked differences in the underlying patient population. Although repeat sensitivity analysis suggested that the two studies by Sangiovanni continued to be outliers, omission of these studies did not result in

significant improvement in the degree of heterogeneity ($\chi^2 = 42.4$, $P < 0.001$, $I^2 = 79\%$) so they were not excluded. After exclusion of the Caturelli study, repeat analysis demonstrated a pooled sensitivity of 63% (95% CI: 49–76) (Figure 3, Table 4). Although the study by Oka *et al.*³¹ was the only study to include prevalent tumours, i.e. HCC diagnosed within the first 6 months, this study was not an outlier on sensitivity analysis or inspection of forest plots. The relatively small number

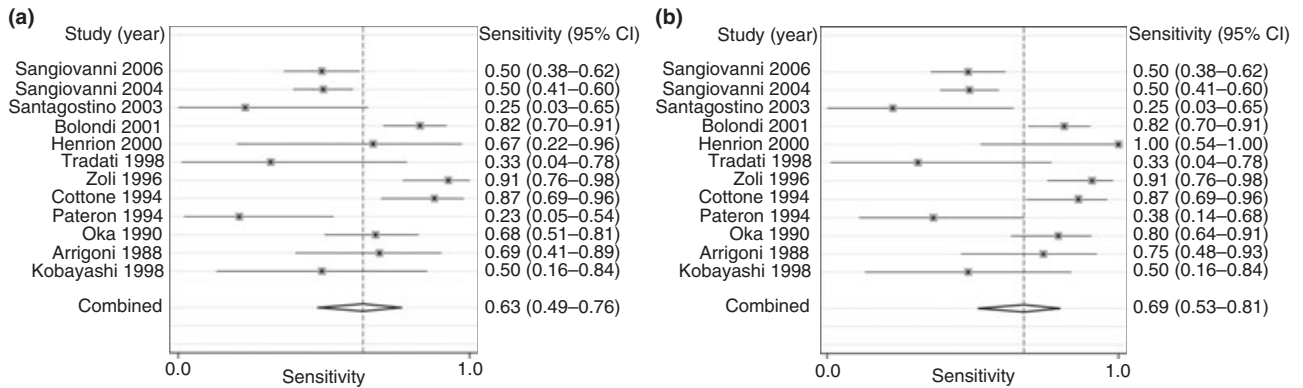


Figure 3. Sensitivity of ultrasound with and without afp for the detection of early-stage hepatocellular carcinoma (HCC): (a) forest plot for the sensitivity of ultrasound to detect early HCC; (b) forest plot for the sensitivity of ultrasound with AFP to detect early HCC. Q, chi-squared test of heterogeneity; I², inconsistency index.

Table 4. Sensitivity of ultrasound with or without alpha-fetoprotein for the detection of early-stage HCC

First author and study	No. pts	No. HCC	No. early HCC	Ultrasound sensitivity (95% CI)	Ultrasound with AFP sensitivity (95% CI)
Kobayashi ³³	95	8	6	0.5 (0.16–0.84)	0.50 (0.16–0.84)
Arrigoni ²⁸	164	16	12	0.69 (0.41–0.89)	0.75 (0.48–0.93)
Oka ³¹	140	40	33	0.68 (0.51–0.81)	0.8 (0.64–0.91)
Pateron ³⁴	118	14	5	0.23 (0.05–0.54)	0.38 (0.14–0.68)
Cottone ²⁹	147	5	4	0.87 (0.69–0.96)	0.87 (0.69–0.96)
Zoli ³²	164	34	32	0.91 (0.76–0.98)	0.91 (0.76–0.98)
Tradati ⁴⁵	40	6	2	0.33 (0.04–0.78)	0.33 (0.04–0.78)
Henrion ³⁰	94	6	6	0.67 (0.22–0.96)	1.0 (0.54–1.0)
Bolondi ⁴²	313	61	50	0.82 (0.70–0.91)	0.82 (0.70–0.91)
Caturelli ²⁷	1599	269	264	0.98 (0.96–0.99)	0.99 (0.96–0.99)
Santagostino ⁴⁶	66	8	2	0.25 (0.03–0.65)	0.25 (0.03–0.65)
Sangiovanni ⁴³	417	112	55	0.50 (0.41–0.60)	0.50 (0.41–0.60)
Sangiovanni ⁴⁴	214	68	34	0.50 (0.38–0.62)	0.50 (0.38–0.62)
Pooled estimates	3113	498	434	0.69 (0.50–0.83)	0.74 (0.56–0.87)
Pooled estimates*	1514	229	170	0.63 (0.49–0.76)	0.69 (0.53–0.81)

Pts, patients; HCC, hepatocellular carcinoma; AFP, alpha-fetoprotein.
 * Pooled estimates after exclusion of outlier studies (Caturelli *et al.*²⁷).

of prevalent HCC ($n = 2$) in this study was unlikely to have a large statistical impact and this study was not excluded. There did not appear to be any publication bias by Begg's test ($P = 0.75$) or funnel plot analysis.

Possible causes for heterogeneity in the sensitivity of ultrasound for the detection of early HCC were then evaluated by meta-regression. The use of concurrent tests, e.g. CT scan, accounted for a significant degree of variation in sensitivity across the included studies ($P = 0.002$). The pooled sensitivity for the two studies with concurrent tests^{33, 34} was 33.3% (95% CI: 7.7–58.9), while the pooled sensitivity for studies without concurrent tests was 64.3% (95% CI: 52.2–76.5). Differences in the interval between surveillance examinations also explained heterogeneity in sensitivity between studies ($P = 0.001$). Studies with surveillance intervals of <6 months had a pooled sensitivity of 70.1% (95% CI: 55.6–84.6), while the studies with surveillance intervals between 6 and 12 months had a pooled sensitivity of 50.1% (95% CI: 40.0–59.2) (Figure 4). Both the use of concurrent tests and length of surveillance interval remained statistically significant after testing for multiplicity using a Monte-Carlo permutation test with 10 000 replications. The proportion of patients with Child's A cirrhosis ($P = 0.44$) and duration of follow-up from enrolment ($P = 0.26$) were not statistically significant causes of heterogeneity. There was no significant difference in the sensitivity of ultrasound between studies conducted in Europe

and those conducted in Asian countries ($P = 0.98$). Similarly, there were no differences between studies conducted before and after 1992 ($P = 0.91$), suggesting that advances in technology did not play a major role.

AFP and ultrasound for detecting early HCC

Finally, we explored the additional benefit of AFP in conjunction with ultrasound for the detection of early HCC and found that the pooled sensitivity increased to 69% (95% CI: 53–81%; $P = 0.65$). The forest plot of the sensitivity of ultrasound and AFP for detecting early HCC is shown in Figure 3. A wide range of AFP cut-offs (15–400 ng/mL) were used to diagnose HCC in the included studies, although the cut-off level did not appear to affect the utility of AFP ($P = 0.95$).

Quality assessment

Using the QUADAS¹⁶ checklist for methodological quality, we found that 12 of the 13 included studies were limited by verification bias. Only Kobayashi *et al.*³³ had reference tests, CT scan and infusion hepatic angiography in every patient regardless of ultrasound results. Additionally, none of the studies followed patients for an additional period of time to confirm that patients without HCC at the end of the study did not have undetected tumours. In each of the other 12 studies, no additional tests were performed in all cirrhotic patients to confirm the absence of HCC.

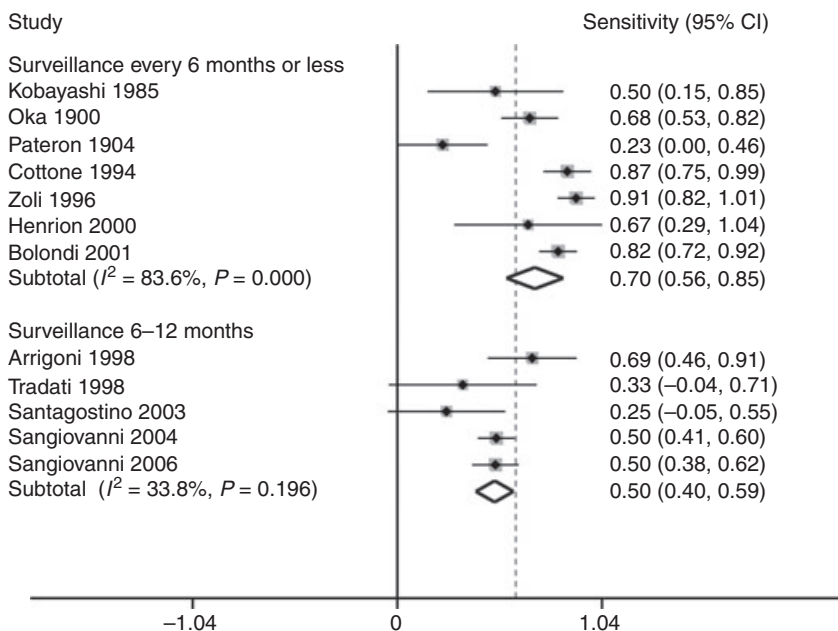


Figure 4. Meta-regression using surveillance intervals for the sensitivity of ultrasound to detect early hepatocellular carcinoma (HCC). Studies with surveillance at least every 6 months had a significantly higher sensitivity for early-stage HCC than studies performing surveillance on an annual basis ($P = 0.001$).

Kobayashi was also the only study in which reviewers of the reference standard were clearly blinded to results of the index ultrasound. The other 12 studies relied primarily on ultrasound-guided biopsy to confirm the diagnosis of HCC.

DISCUSSION

Our study is the first meta-analysis to evaluate ultrasound with or without AFP as a surveillance tool for early HCC in cirrhotic patients. We demonstrated that surveillance programmes with ultrasound are highly accurate for HCC at any stage, with a pooled sensitivity of 94% and a pooled specificity of 94%. However, the detection of early HCC is of greater importance for surveillance to be successful. Our study demonstrated that ultrasound only has a pooled sensitivity of 63% for those with early HCC. Meta-regression analysis demonstrated a significantly higher sensitivity for early HCC with an ultrasound every 6 months than with annual surveillance ($P = 0.001$).

Although a systematic review has been previously performed on the efficacy of ultrasound for the diagnosis of HCC,¹⁵ there are several significant differences from our meta-analysis. First, we only included studies that used ultrasound as a surveillance tool in a prospective manner, whereas previous studies assessed ultrasound as a one-time diagnostic tool. This is an important distinction given that ultrasound is most commonly used as a surveillance tool in clinical practice. Second, our analysis specifically evaluated the sensitivity of ultrasound for early HCC. Once again, this is clinically relevant as curative measures are only available for early HCC, making detection of tumours at this stage essential during surveillance.

As commonly discovered in meta-analyses of diagnostic tests,³⁵ we found a moderate degree of heterogeneity for the sensitivity of surveillance ultrasound to detect HCC. This heterogeneity can be related to variability in diagnostic thresholds, study populations, test equipment or methods, study quality or a combination of these factors.³⁶ We were unable to explore some possible aetiologies for heterogeneity including differences in body habitus, operator skill or experience and inter-operator reliability due to limited available information. In our meta-analysis, we found the use of concurrent tests with ultrasound was able to explain a significant degree of heterogeneity in the pooled sensitivity estimate for early HCC. The pooled sensitivity for the two studies with concomitant tests was 33.3%,

which was significantly lower than the sensitivity of 64.3% in the studies without concurrent tests ($P = 0.002$). In these latter studies, it is likely that some tumours were never detected by ultrasound and AFP during the follow-up period. Therefore, these reported sensitivities may be falsely high and our pooled sensitivity of 63% for early HCC may overestimate ultrasound's true performance characteristics during surveillance.

There has been considerable debate regarding the additional benefit of AFP to ultrasound during surveillance as well as the optimal surveillance interval.³⁷ We demonstrated that the addition of AFP to ultrasound does not substantially improve the sensitivity of surveillance for early HCC, independent of the cut-off level used. Although the pooled sensitivity for early HCC minimally increased from 63% to 69%, this was not statistically significant ($P = 0.65$). This finding is consistent with the AASLD practice guidelines, which suggest that AFP is not an adequate screening test, but has a role in the diagnosis of HCC when >200 ng/mL in the setting of a mass on imaging.³⁸

Conversely, variation in surveillance intervals resulted in significant differences in sensitivity for early HCC. The pooled sensitivity of the studies with surveillance at least every 6 months had a pooled sensitivity of 70.1%, which was significantly better than the sensitivity of 50.1% in studies performing surveillance on an annual basis ($P = 0.001$). Our meta-analysis suggests that surveillance with an ultrasound every 6 months is currently the best interval for detecting early HCC among patients with cirrhosis.

While some studies have proposed that CT or magnetic resonance imaging may be more sensitive as alternative imaging studies for the detection of HCC, they have not been adequately studied as surveillance tests or with regard to early HCC.¹⁵ Additionally, the increased cost and potential adverse effects such as radiation exposure limit their utility in surveillance.³⁹ There have been promising tumour biomarkers, including des-gamma carboxy-prothrombin and the lens culinaris-agglutinin reactive fraction of AFP (APF-L3%), but there is insufficient evidence for their use in clinical practice.^{40, 41} Overall, more studies are needed to find novel surveillance tests to improve the detection of HCC at stages where curative interventions can be applied.

Although the included studies are the best data currently available, the primary limitations of our meta-analysis are the biases observed in these studies. On

quality assessment, verification bias was a significant concern in all but one study. Only Kobayashi *et al.*³³ performed concurrent imaging that can serve as a reference standard in all patients, whereas all other studies performed a 'gold standard' reference test only in patients with a positive ultrasound or AFP. Similarly, none of the studies followed patients for an additional period of time to confirm that patients without HCC at the end of the study did not have any undetected tumours. In these studies, the calculated sensitivity for ultrasound may have been subsequently falsely elevated. These limitations are important and suggest that the sensitivity of ultrasound for early HCC is 63% at best and may in fact be significantly worse.

Other than the biases of the individual studies, another limitation of our meta-analysis is that we only evaluated surveillance in patients with cirrhosis and our results may not be generalizable to other populations undergoing HCC surveillance, such as patients with hepatitis B. Additionally, most of the studies were conducted in experienced liver centres in Europe and Asia; the performance of ultrasound may be worse in an American cohort in which obesity can further limit its sensitivity and many ultrasounds are performed outside high-volume medical centres by technicians instead of radiologists. Finally, our meta-analysis only evaluated the efficacy of ultrasound to detect early HCC given the lack of prospective trials evaluating the

effect of surveillance on outcomes such as overall survival. Despite these limitations, this was the first meta-analysis evaluating ultrasound and AFP as surveillance tools for early-stage HCC, rather than single-application diagnostic tests. More importantly, this is the first compilation of studies specifically evaluating the efficacy of surveillance to detect early HCC.

In summary, ultrasound demonstrates a limited sensitivity of 63% but is currently the best surveillance tool for early-stage HCC among patients with cirrhosis. The addition of AFP to ultrasound is of minimal benefit, whereas performing ultrasound every 6 months instead of annually significantly improves sensitivity for early HCC to 70%. Unfortunately, existing studies suffer from significant limitations that include verification bias, unknown efficacy of ultrasonography in less-experienced centres and questionable generalizability of these results to American patients with cirrhosis. Further studies should be performed to overcome these limitations as well as determine if the addition of novel biomarkers can help improve the detection of early HCC.

ACKNOWLEDGEMENTS

Declaration of personal interests: None. *Declaration of funding interests:* This study was funded in part by grant number DK064909 (JAM) and DK077707 (JAM).

REFERENCES

- 1 Bosch FX, Ribes J, Cleries R, Diaz M. Epidemiology of hepatocellular carcinoma. *Clin Liver Dis* 2005; 9: 191–211 v.
- 2 El-Serag HB. Hepatocellular carcinoma: recent trends in the United States. *Gastroenterology* 2004; 127(Suppl. 1): S27–34.
- 3 Espey DK, Wu XC, Swan J, *et al.* Annual report to the nation on the status of cancer, 1975–2004, featuring cancer in American Indians and Alaska Natives. *Cancer* 2007; 110: 2119–52.
- 4 El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 2007; 132: 2557–76.
- 5 Fattovich G, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology* 2004; 127(Suppl. 1): S35–50.
- 6 Capocaccia R, Sant M, Berrino F, Simonetti A, Santi V, Trevisani F. Hepatocellular carcinoma: trends of incidence and survival in Europe and the United States at the end of the 20th century. *Am J Gastroenterol* 2007; 102: 1661–70; quiz 1660, 1671.
- 7 El-Serag HB, Mason AC, Key C. Trends in survival of patients with hepatocellular carcinoma between 1977 and 1996 in the United States. *Hepatology* 2001; 33: 62–5.
- 8 Llovet JM, Bruix J. Early diagnosis and treatment of hepatocellular carcinoma. *Baillieres Best Pract Res Clin Gastroenterol* 2000; 14: 991–1008.
- 9 Ioannou GN, Perkins JD, Carithers RL Jr. Liver transplantation for hepatocellular carcinoma: impact of the MELD allocation system and predictors of survival. *Gastroenterology* 2008; 134: 1342–51.
- 10 Mazzaferro V, Regalia E, Doci R, *et al.* Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med* 1996; 334: 693–9.
- 11 Llovet JM, Bustamante J, Castells A, *et al.* Natural history of untreated nonsurgical hepatocellular carcinoma: rationale for the design and evaluation of therapeutic trials. *Hepatology* 1999; 29: 62–7.
- 12 Marrero JA, Fontana RJ, Barrat A, *et al.* Prognosis of hepatocellular carcinoma: comparison of 7 staging systems in an American cohort. *Hepatology* 2005; 41: 707–16.
- 13 Meissner HI, Smith RA, Rimer BK, *et al.* Promoting cancer screening: learning from experience. *Cancer* 2004; 101 (Suppl. 5): 1107–17.
- 14 Gupta S, Bent S, Kohlwes J. Test characteristics of alpha-fetoprotein for detecting

- hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *Ann Intern Med* 2003; 139: 46–50.
- 15 Colli A, Fraquelli M, Casazza G, *et al.* Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. *Am J Gastroenterol* 2006; 101: 513–23.
 - 16 Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3: 25.
 - 17 Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539–58.
 - 18 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Br Med J* 2003; 327: 557–60.
 - 19 Copas JB, Shi JQ. A sensitivity analysis for publication bias in systematic reviews. *Stat Methods Med Res* 2001; 10: 251–65.
 - 20 Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001; 20: 641–54.
 - 21 Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann Thorac Surg* 2005; 79: 16–20.
 - 22 Rosman AS, Korsten MA. Application of summary receiver operating characteristics (sROC) analysis to diagnostic clinical testing. *Adv Med Sci* 2007; 52: 76–82.
 - 23 Walter SD. Properties of the summary receiver operating characteristic (sROC) curve for diagnostic test data. *Stat Med* 2002; 21: 1237–56.
 - 24 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36.
 - 25 Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003; 22: 2693–710.
 - 26 Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004; 23: 1663–82.
 - 27 Caturelli E, Bartolucci F, Biasini E, *et al.* Diagnosis of liver nodules observed in chronic liver disease patients during ultrasound screening for early detection of hepatocellular carcinoma. *Am J Gastroenterol* 2002; 97: 397–405.
 - 28 Arrigoni A, Andriulli A, Gindro T, Piantino P, Capussotti L, Rizzetto M. Pattern analysis of serum alpha-fetoprotein in the early diagnosis of hepatocellular carcinoma in liver cirrhosis. *Int J Biol Markers* 1988; 3: 172–6.
 - 29 Cottone M, Turri M, Caltagirone M, *et al.* Screening for hepatocellular carcinoma in patients with Child's A cirrhosis: an 8-year prospective study by ultrasound and alphafetoprotein. *J Hepatol* 1994; 21: 1029–34.
 - 30 Henrion J, Libon E, De Maeght S, *et al.* Surveillance for hepatocellular carcinoma: compliance and results according to the aetiology of cirrhosis in a cohort of 141 patients. *Acta Gastroenterol Belg* 2000; 63: 5–9.
 - 31 Oka H, Kurioka N, Kim K, *et al.* Prospective study of early detection of hepatocellular carcinoma in patients with cirrhosis. *Hepatology* 1990; 12: 680–7.
 - 32 Zoli M, Magalotti D, Bianchi G, Gueli C, Marchesini G, Pisi E. Efficacy of a surveillance program for early detection of hepatocellular carcinoma. *Cancer* 1996; 78: 977–85.
 - 33 Kobayashi K, Sugimoto T, Makino H, *et al.* Screening methods for early detection of hepatocellular carcinoma. *Hepatology* 1985; 5: 1100–5.
 - 34 Pateron D, Ganne N, Trinchet JC, *et al.* Prospective study of screening for hepatocellular carcinoma in Caucasian patients with cirrhosis. *J Hepatol* 1994; 20: 65–71.
 - 35 Pai M, McCulloch M, Enanoria W, Colford JM Jr. Systematic reviews of diagnostic test evaluations: what's behind the scenes? *ACP J Club* 2004; 141: A11–3.
 - 36 Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995; 48: 119–30; discussion 131–2.
 - 37 Di Bisceglie AM. Issues in screening and surveillance for hepatocellular carcinoma. *Gastroenterology* 2004; 127(Suppl. 1): S104–7.
 - 38 Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology* 2005; 42: 1208–36.
 - 39 Brenner DJ, Hall EJ. Computed tomography – an increasing source of radiation exposure. *N Engl J Med* 2007; 357: 2277–84.
 - 40 Marrero JA. Screening tests for hepatocellular carcinoma. *Clin Liver Dis* 2005; 9: 235–51, vi.
 - 41 Marrero JA, Su GL, Wei W, *et al.* Desgamma carboxyprothrombin can differentiate hepatocellular carcinoma from non-malignant chronic liver disease in american patients. *Hepatology* 2003; 37: 1114–21.
 - 42 Bolondi L, Sofia S, Siringo S, *et al.* Surveillance programme of cirrhotic patients for early diagnosis and treatment of hepatocellular carcinoma: a cost effectiveness analysis. *Gut* 2001; 48: 251–9.
 - 43 Sangiovanni A, Del Ninno E, Fasani P, *et al.* Increased survival of cirrhotic patients with a hepatocellular carcinoma detected during surveillance. *Gastroenterology* 2004; 126: 1005–14.
 - 44 Sangiovanni A, Prati GM, Fasani P, *et al.* The natural history of compensated cirrhosis due to hepatitis C virus: a 17-year cohort study of 214 patients. *Hepatology* 2006; 43: 1303–10.
 - 45 Tradati F, Colombo M, Mannucci PM, *et al.* A prospective multicenter study of hepatocellular carcinoma in italian hemophiliacs with chronic hepatitis C. The Study Group of the Association of Italian Hemophilia Centers. *Blood* 1998; 91: 1173–7.
 - 46 Santagostino E, Colombo M, Rivi M, *et al.* A 6-month versus a 12-month surveillance for hepatocellular carcinoma in 559 hemophiliacs infected with the hepatitis C virus. *Blood* 2003; 102: 78–82.