Comparison of a multitarget blood test to ultrasound and alpha-fetoprotein for hepatocellular carcinoma surveillance: Results of a network meta-analysis

Amit G. Singal1 | Benjamin Haaland2 | Neehar D. Parikh3 | A. Burak Ozbay4 | Carol Kirshner5 | Shubham Chakankar5 | Kyle Porter4 | Jagpreet Chhatwal6 | Turgay Ayer7,8

1Division of Digestive and Liver Diseases, University of Texas Southwestern Medical Center, Dallas, Texas, USA
2University of Utah School of Medicine and Huntsman Cancer Institute, Salt Lake City, Utah, USA
3Division of Gastroenterology and Hepatology, University of Michigan Medical School, Ann Arbor, Michigan, USA
4Exact Sciences Corporation, Madison, Wisconsin, USA
5Value Analytics Labs, Boston, Massachusetts, USA
6Institute for Technology Assessment, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA
7Georgia Institute of Technology, Atlanta, Georgia, USA
8Emory Medical School, Atlanta, Georgia, USA

Correspondence
Amit G. Singal, Division of Digestive and Liver Diseases, UT Southwestern Medical Center, 5959 Harry Hines Blvd, POB 1, Suite 420, Dallas TX 75390–8887, USA. Email: amit.singal@utsouthwestern.edu

Abstract
Ultrasound-based surveillance has suboptimal sensitivity for early detection of hepatocellular carcinoma (HCC) in patients with cirrhosis. There are several emerging alternatives, including a novel multitarget HCC blood test (Mt-HBT). We compared performance of mt-HBT against ultrasound with or without alpha-fetoprotein (AFP) for early HCC detection in patients with cirrhosis. Per the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines, two reviewers searched PubMed, Cochrane, Embase, and clinicaltrials.gov databases from January 1990 through December 2020 to identify studies reporting sensitivity and/or specificity of ultrasound and AFP for overall and early stage HCC detection in patients with cirrhosis. Mt-HBT diagnostic performance was derived from a clinical validation study. A network meta-analysis model was built for comparative assessment, and pooled estimates of sensitivity at a fixed specificity were estimated based on Bayesian binormal receiver operating characteristic models for each modality. Forty-one studies (comprising 62,517 patients with cirrhosis) met inclusion criteria. Ultrasound-alone sensitivity was 51.6% (95% credible interval [CrI], 43.3%–60.5%) for early stage HCC detection, which increased with the addition of AFP to 74.1% (95% CrI, 62.6%–82.4%); however, this was offset by decreased specificity (87.9% vs. 83.9%, respectively). With specificity fixed at 90%, mt-HBT sensitivity for early stage HCC detection was higher than ultrasound alone (18.2%; 95% CrI, 0.2%–37.7%) and similar to ultrasound and AFP (−3.3%; 95% CrI, −22.3%–17.4%). Pairwise posterior probabilities suggested a preference for mt-HBT over ultrasound alone in 97.4% of cases but only 36.3% of cases versus ultrasound with AFP. Conclusion: A blood-based mt-HBT has higher sensitivity than ultrasound alone for early stage HCC detection but similar sensitivity compared to ultrasound and AFP. Mt-HBT could
INTRODUCTION

Hepatocellular carcinoma (HCC) is the third leading cause of cancer death worldwide and a rising cause of cancer-related death in the United States. Despite improved treatment options over time, the average 5-year survival for HCC has remained below 20%. The prognosis for patients with HCC depends on tumor stage at diagnosis, with curative options available for early stage HCC, affording 5-year survival exceeding 70%. In contrast, patients with late-stage HCC have a median survival of only 1–2 years.

Accordingly, professional society practice guidelines recommend ultrasound-based surveillance, with or without alpha-fetoprotein (AFP), every 6 months in patients who are at risk, including those with cirrhosis from any etiology. Several cohort studies have demonstrated that HCC surveillance is associated with significant improvements in early HCC detection, curative treatment receipt, and overall survival in patients with cirrhosis. Although ultrasound with AFP has high sensitivity for any stage HCC detection, it misses over one third of HCC at an early stage, with a pooled sensitivity of only 63% in a recent systematic review and meta-analysis. Further, ultrasound and AFP often yield false-positive or indeterminate results, resulting in potential physical, financial, and psychological harm. Finally, patient and provider barriers to HCC surveillance result in suboptimal adherence in clinical practice, with less than 30% of patients with cirrhosis undergoing surveillance. These notable limitations of current surveillance strategies highlight an unmet need for a simpler efficacious method for HCC surveillance.

A novel, multitarget, HCC blood test (mt-HBT) composed of three methylation markers (homeobox A1 [HOXA1], testis-specific Y-encoded-like protein 5 [TSPYL5], and beta-1,3-galactosyltransferase 6 [B3GALT6]), AFP, and patient sex was recently shown to have high accuracy for early stage HCC detection (82% sensitivity for early stage HCC and 87% specificity) in a phase II biomarker validation study. These data highlight the potential of mt-HBT to provide an efficacious alternative to ultrasound-based surveillance, particularly given the relative ease of implementing blood-based biomarker strategies; however, there is no head-to-head comparison between these two tests.

The objective of this study was to conduct a systematic literature review of studies for the sensitivity and specificity of ultrasound with or without AFP for HCC surveillance and compare test performance with mt-HBT by using a network meta-analysis.

MATERIALS AND METHODS

The study was reported in accordance to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Network Meta-Analysis (PRISMA-NMA) guidelines for a complete systematic literature review and network meta-analysis.

Systematic literature review

An electronic-based search was performed using medical subject headings terms, text words, title abstracts, and keywords in PubMed/MEDLINE, Cochrane, and Embase database libraries, with different combinations to identify the pool of studies published between January 1, 1990, and December 8, 2020, evaluating sensitivity and specificity of HCC surveillance modalities in patients with cirrhosis (Table S2). After duplicate studies were removed, the remaining publications underwent title/abstract and full-text screening. Both title/abstract and full-text screening were performed by two independent investigators, and a third investigator resolved any disagreements or discrepancies.

Study selection and inclusion/exclusion criteria

Potential eligible study designs included both randomized controlled trials (with no restriction on randomization procedure) and observational studies (case–control and cohort studies). The study population included patients with cirrhosis, with no restrictions on age, who were undergoing HCC surveillance with any of the following modalities: ultrasound alone, AFP alone, ultrasound and AFP, and mt-HBT. Studies not reported in English, nonhuman data, conference abstracts, review articles, case report studies with less than five patients, opinion letters, and studies focusing on imaging for diagnostic purposes only (rather than surveillance) were excluded. Studies with sequential test combinations (imaging based on AFP levels) were excluded because of information bias from the initial study. Studies with <50% of patients with cirrhosis or those that did not report the proportion of patients with cirrhosis were also excluded. If the same cohort of patients was used in duplicate publications, data from the most recent study were included.
Data extraction

After a set of articles that met inclusion and exclusion criteria was identified through full-text review (Table S1), the data were extracted from individual articles by independent investigators and validated by the scientific team. Data from included studies were extracted following the structure highlighted in Table S3. The extracted outcomes data included sensitivity and specificity of HCC surveillance modalities and overall and early stage HCC; effect sizes for test comparative effectiveness with credible intervals (CrIs); and number of HCC cases (Tables S3 and S4). Data for early stage HCC detection were extracted if studies used either Milan criteria or the Barcelona Clinic Liver Cancer staging system.

Quality/bias assessment

Quality/bias assessment of included studies was performed using a modified checklist based on the National Institutes of Health study quality assessment tool for observational studies.

Network meta-analysis

Because comparisons of ultrasound versus AFP, ultrasound alone versus ultrasound plus AFP, and AFP versus mt-HBT are present in the literature, we conducted an NMA to estimate indirect comparisons of ultrasound versus mt-HBT and ultrasound plus AFP versus mt-HBT. The NMA was conducted based on a Bayesian binormal receiver operating characteristic (ROC) model, jointly for sensitivity and specificity of each modality. The NMA model included terms for within-study variation in sensitivity and specificity estimates, across-study averages in sensitivity and specificity for each modality, between-study variation in sensitivity and specificity for each modality, and correlations between the study effects for distinct modalities. The within-study variation terms captured sampling variability due to finite within-study sample sizes, while the between-study variation terms captured the study-to-study variations in performance of each modality. The binormal ROC model provides a link between the diversity of operating specificities for a particular diagnostic modality and the corresponding sensitivities and provides a means for estimating the sensitivity at a fixed-target specificity.

The overall structure of the NMA for comparison of HCC surveillance modalities is illustrated in Figure S1. The primary comparative assessments of current surveillance modalities to mt-HBT were conducted based on average sensitivity at a fixed 90% operating specificity. The threshold of 90% specificity was established based on recommendations in biomarker validation. Estimates and contrasts were accompanied by 95% CrIs, 95% predictive intervals, and pairwise posterior probabilities that each modality is superior to the other. In addition, the surface under cumulative ranking curve (SUCRA) for each modality was estimated to identify hierarchy between modalities and quantify acceptability of the HCC surveillance tests. Prior distributions were constructed un informatively but in a manner to rule out implausible estimates. Given the limited range of operating specificities, particularly for ultrasound-based diagnostics that identify presence/absence of lesions, meta-analyzed average sensitivities and specificities are reported in secondary analyses. Study-to-study heterogeneity in modality performance was assessed by stratification on study-level characteristics and predictive intervals, which are constructed to capture the plausible range of study-specific performance and comparative performance of each modality in a new unobserved study setting. In addition, we generated separate forest plots for sensitivity and specificity, showing point estimates along with credible and predictive intervals. Each estimate was annotated with the corresponding probability that the modality is the best. Finally, we generated Pareto frontier, which provides a visual summary of multiobjectives in multiobjective optimization (for joint posterior distribution of sensitivity and specificity); graphics were generated as graphical summaries.

Subgroup analyses

We additionally conducted subgroup analyses stratified by geography (North America vs. other), study design (prospective vs. retrospective), study period (study period ending after vs. during or before 2006), and proportion of viral hepatitis (≤75% vs. >75% of patients). Local clinical validation study for the mt-HBT modality (which had a prospective study design, had completed data collection in 2020, and included 48.5% of patients with viral etiology), estimates for mt-HBT and corresponding comparisons were only included for one side of each stratification.

RESULTS

Study selection

Of 8598 studies identified through a systematic search, 815 duplicates were removed and 7783 underwent title and abstract screening. There were 119 studies that went on to full-text review, of which 41 (comprising 62,517 patients with cirrhosis) met all inclusion and exclusion criteria and were suitable for data extraction (Figure 1). A majority (71%) of the studies reported data on more than one surveillance modality, such as both ultrasound alone and ultrasound plus AFP. Overall, 34 studies (n = 13,544
patients) reported performance characteristics for ultrasound alone and 14 studies (n = 7140 patients) evaluated the performance of ultrasound with AFP. We extracted data from these 41 studies and compared them against the published clinical validation study of mt-HBT.

**Study characteristics**

Characteristics of the studies included in NMA are detailed in Table 1. Of the 41 studies, 21 were retrospective studies, 18 were prospective, one was a case–control study, and one was a randomized controlled trial. Fourteen studies were conducted in North America, 10 in Europe, 12 in Asia, and the remaining five in other regions (Argentina, Australia, or Egypt). The mean age of participants ranged from 41 to 67 years, and the proportion of men ranged from 35% to 99%.

**Test sensitivities**

The sensitivities of AFP alone, ultrasound alone, ultrasound with AFP, and mt-HBT at the observed specificities are reported in Table 2. As expected, the sensitivity of AFP alone for early stage HCC detection was poor at 49.1% (95% CrI, 40.7%–56.1%) with specificity of 87.9% (95% CrI, 83.4%–92.5%). Ultrasound alone had a sensitivity of only 51.6% (95% CrI, 43.3%–60.5%) for early stage HCC detection, which was increased to 74.1% with the addition of AFP (95% CrI, 62.6%–82.4%); however, this was offset by a small decrease in specificity (87.9% vs. 83.9%). When specificity was fixed at 90%, the sensitivities of ultrasound alone and ultrasound with AFP for early stage HCC detection were 50.9% (95% CrI, 42.6%–59.1%) and 72.4% (95% CrI, 60.8%–80.5%), respectively (Table 3).

Compared with ultrasound alone, the sensitivity of mt-HBT did not significantly differ for any stage HCC (difference 6.6%; 95% CrI, −11.2%–23.2%) or early stage HCC (18.2%; 95% CrI, −0.2%–37.7%) when specificity was fixed at 90% (Table 4). Similarly, mt-HBT sensitivity did not significantly differ from ultrasound and AFP for detection of any stage HCC (5.5%; 95% CrI, −13.7%–21.1%) or early stage HCC (−3.3%; 95% CrI, −22.3%–17.4%). SUCRA values for any stage and early stage HCC surveillance showed preference for mt-HBT over ultrasound alone; however, cumulative
| Study | Type of study | Location | Study period | Total enrolled | With cirrhosis | With HCC | Male % | Mean age | Surveillance modality |
|-------|--------------|----------|--------------|----------------|----------------|----------|--------|----------|----------------------|
| Atiq et al. [34] | retrospective | USA | Jul 2010–Jul 2013 | 680 | 680 | 78 | 64.7 | 54.3 | US & AFP |
| Caviglia et al. [12] | retrospective | Italy | Nov 2012–Jan 2018 | 349 | 200 | 86 | 67 | 61 | AFP |
| Chalasani et al. [9] | retrospective | USA | Jan 1994–Dec 1997 | 285 | 285 | 27 | 56 | 49 | US & AFP |
| Chang et al. [13] | retrospective | Taiwan | Jul 2002–Jul 2010 | 5664 | 1597 | 363 | 64.9 | 57 | US, AFP, & US+AFP |
| Chun et al. [35] | retrospective | Korea | Mar 2012–Aug 2013 | 49,381 | 48,122 | 24 | 91.4 | 60 | AFP |
| Curley et al. [14] | retrospective | USA | Jan 1994–Dec 1997 | 416 | 140 | 36 | 58.8 | 54.8 | US, AFP, & US+AFP |
| D’Onofrio et al. [36] | retrospective | Italy | 2003–2006 | 128 | 128 | 57 | 62.5 | 46 | US & AFP |
| Eltabbakh et al. [37] | prospective | Egypt | Jun 2008–Jan 2011 | 1286 | 1286 | 102 | 35.4 | 51.5 | US & US+AFP |
| Frey et al. [38] | retrospective | Switzerland | 2011–2012 | 285 | 248 | 9 | 80 | 55 | US |
| Gambarin-Galwan et al. [10] | retrospective | USA | 1 year (dates NA) | 106 | 105 | 19 | 68.4 | 58.7 | US, AFP, & US+AFP |
| Gopal et al. [15] | retrospective | USA | Jan 2005–Jun 2012 | 1128 | 1128 | 452 | 78 | 55 | AFP |
| Henrion [39] | retrospective | Belgium | 1995 | 141 | 141 | 6 | 63.1 | NA | US, AFP, & US+AFP |
| Ishii et al. [40] | prospective | Japan | Nov 1992–Mar 1994 | 918 | 299 | 29 | 78 | 66 | AFP |
| Jasirwan et al. [16] | retrospective | Indonesia | 2015–2017 | 330 | 66 | 132 | 71.2 | NA | AFP |
| Kim et al. [41] | prospective | Korea | Jan 2007–Jun 2010 | 392 | 392 | 64 | 63.5 | 53.6 | US, AFP, & US+AFP |
| Lersritwimanmaen and Nimanong [17] | retrospective | Thailand | Jan 2012–Dec 2014 | 452 | 350 | 9 | 63.5 | 59.2 | US & US+AFP |
| Lok et al. [18] | case control | USA | 3.5 years (dates NA) | 1031 | 116 | 39 | 71 | 50 | US & AFP |
| Oka et al. [42] | prospective | Japan | Jan 1983–Dec 1988 | 140 | 140 | 40 | 70 | 54.9 | US & US+AFP |
| Parangaba-Vezozzo et al. [43] | prospective | Canada | Aug 1998–Aug 2008 | 884 | 884 | 72 | 58.3 | 55 | US & US+AFP |
| Pateron et al. [44] | prospective | France | Jan 1986–Dec 1987 | 322 | 118 | 14 | 57.6 | 55 | US & AFP |
| Paul et al. [45] | prospective | India | Apr 2001–Nov 2014 | 301 | 301 | 107 | 88.3 | 45.1 | US |
| Phulpoto et al. 2012 [46] | prospective | Pakistan | Apr 2008–Nov 2010 | 301 | 194 | 9 | 70 | 45.1 | US |
| Pinero et al. 2014 [11] | retrospective | Argentina | Jun 2005–Dec 2011 | 643 | 643 | 56 | 70 | 59 | US |
| Pocha et al. [47] | prospective | USA | Jun 2002–Feb 2011 | 163 | 163 | 9 | 99.4 | 59.3 | US |
| Qian et al. [48] | retrospective | Australia | Oct 1998–Aug 2005 | 268 | 268 | 22 | 69.4 | 57.1 | US, AFP, & US+AFP |
| Ren et al. [49] | prospective | USA | Jun 1995–Jun 2005 | 6089 | 201 | 264 | 71 | 50 | US & AFP |
| Sangiovanni et al. [50] | prospective | Italy | 1985–1986 | 447 | 447 | 112 | 79 | 61 | US & AFP |
| Sato et al. [51] | retrospective | Japan | 1994–2004 | 1431 | 124 | 243 | 57 | 67 | US & AFP |
| Singal et al. [52] | prospective | USA | Jan 2004–Sep 2006 (2 years, 9 months) | 446 | 446 | 41 | 58.6 | 52.8 | US, AFP, & US+AFP |
| Sinn et al. [19] | retrospective | South Korea | Jan 2008–Dec 2010 | 132 | 94 | 97 | 76.5 | 57.7 | US, AFP, & US+AFP |

(Continues)
ranking preferences were similar between mt-HBT and ultrasound with AFP for detection of any stage HCC (81.8% vs. 65.2%, respectively) and early stage HCC (77.6% vs. 87.9%, respectively) (Table 4). Pairwise posterior probabilities suggested a preference for mt-HBT over ultrasound and AFP in 70.3% of cases for any stage HCC detection and 36.3% of cases for early stage HCC detection (Table 3).

There was variation in performance of modalities from study to study in both absolute and relative terms, as demonstrated by the predictive intervals in Figure 2. However, even in the presence of this variation, the comparative analysis showed that mt-HBT would be statistically significantly superior to AFP alone for both any stage and early stage HCC detection (posterior predictive probability mt-HBT superior to AFP alone: 97.2% in any stage and 90.4% in early stage HCC). Further, there was moderate evidence that mt-HBT would be superior to ultrasound alone in early stage HCC surveillance (posterior predictive probability mt-HBT superior to ultrasound alone, 86.4% in early stage) (Figure 3).

Subgroup analyses

North American sites only

Results for the subgroup of studies conducted in North America are included in Table 5. Specifically, we report the sensitivities of AFP alone, ultrasound alone, ultrasound with AFP, and mt-HBT, at the observed specificities in Table 5A, at fixed specificities in Table 5B, and sensitivity comparisons at 90% fixed specificity in Table 5C. The sensitivity of AFP alone for early stage HCC detection was 50.0% (95% CrI, 38.8%–58.8%) with specificity of 88.7% (95% CrI, 84.7%–92.7%). Ultrasound alone had a poor sensitivity of only 44.0% (95% CrI, 27.6%–58.4%) for early stage HCC detection, which was substantially lower than the pooled estimate of ultrasound performance across all studies globally. The addition of AFP to ultrasound increased early stage sensitivity to 63.5% (95% CrI, 42.8%–83.2%); however, this was offset by a decrease in specificity (90.7% vs. 84.2%). When specificity was fixed at 90%, the sensitivities of ultrasound alone and ultrasound with AFP for early stage HCC detection were 68.6% (95% CrI, 59.5%–76.4%) and 80.0% (95% CrI, 64.1%–96.4%), respectively (Table 5B).

Compared with ultrasound alone, the sensitivity of mt-HBT provided a clinically and statistically significant improvement (difference 25.2%; 95% CrI, 8.4%–44.1%) for early stage HCC when specificity was fixed at 90% (Table 5C). On the other hand, mt-HBT sensitivity did not significantly differ from ultrasound with AFP for detection of any stage HCC (0.4%; 95% CrI, –20.2%–19.5%) or early stage HCC (12.3%; 95% CrI, –11.3%–34.6%).
SUCRA values showed preference for mt-HBT over ultrasound alone for detection of early stage HCC (95.1% vs 16%, respectively) and any stage HCC (83.2% vs. 38.2%, respectively). In addition, SUCRA values for early stage HCC favored mt-HBT over ultrasound with AFP (95.1% vs. 62.8%, respectively); however, cumulative ranking preferences were similar between mt-HBT and ultrasound with AFP for detection of any stage HCC (83.2% vs. 78.5%, respectively).

Other subgroup analysis

Results from subgroup analyses stratified by study design, study period, and proportion of viral hepatitis are presented in Tables S1–S6. Overall, the stratified results were qualitatively similar to the overall results, although there was an indication that mt-HBT performance relative to other modalities may be stronger in populations with less viral hepatitis, and the estimated performance of AFP alone, ultrasound alone, and ultrasound with AFP may be worse for retrospective studies.

Quality assessment

Quality/bias assessment of included studies is provided in Table S5. Although most studies had appropriate patient selection, three studies [9–11] only enrolled patients listed for liver transplantation and nine studies [12–21]...
examined cohorts diagnosed with HCC, potentially introducing selection bias and overestimating index test performance with a dearth of false-negative results. Several studies excluded patients with limited life expectancy due to comorbidities or Child C cirrhosis, but this was felt to be appropriate and not introduce selection bias given surveillance is not recommended in patients with Child C cirrhosis outside of candidates for liver transplantation. Most studies used standard reference tests to confirm HCC diagnosis, including computed tomography, magnetic resonance imaging (MRI), and biopsy; however, in most studies, it was unclear if the reference standard was interpreted without knowledge of the index test, potentially overestimating surveillance test performance. Finally, most studies failed to perform reference tests in those with negative surveillance tests, introducing verification bias with likely overestimation of surveillance test performance.

**DISCUSSION**

Our results reinforce that ultrasound alone has a poor sensitivity for early stage HCC detection, with a pooled...
sensitivity of only 51.6%. Early stage HCC detection is significantly increased to 74.1% when ultrasound is used in combination with AFP, although this improved sensitivity was offset with decreased specificity. The poor effectiveness of ultrasound-based surveillance has underscored the importance of novel imaging and blood-based surveillance strategies, including mt-HBT, which was recently validated in a phase II case–control biomarker study. Herein, we found that mt-HBT had similar sensitivity as ultrasound with or without AFP for any stage and early stage HCC detection when specificity was fixed at 90%.

Given increasing recognition of the limitations of ultrasound-based surveillance, there has been increasing interest in alternative imaging and blood-based surveillance strategies. Promising imaging-based strategies include abbreviated MRI, which has demonstrated high sensitivity and specificity in several small case–control studies. There have also been several proposed blood-based biomarkers with promising early results. GALAD is a biomarker panel comprised of sex (gender), age, AFP-L3%, AFP, and des-γ-carboxyprothrombin (DCP); it has demonstrated high sensitivity and specificity for early stage HCC detection in large case–control studies and small cohort studies, with ongoing phase III evaluation in the Early Detection Research Network Hepatocellular Carcinoma Early Detection Strategy Study. The mt-HBT panel was also shown to have high sensitivity and specificity for early stage HCC detection in a large national case–control phase II study, with comparable performance to that of GALAD. A conjoint survey study among three academic health systems suggested that patients would strongly prefer these emerging MRI-based and biomarker-based strategies to that of ultrasound with or without AFP.

Of three available surveillance strategies—ultrasound alone, ultrasound with AFP, and mt-HBT—we found ultrasound with AFP and mt-HBT are preferred, with the highest performance for early stage HCC detection and both achieving sensitivities exceeding 70%. It is important to consider that there is often a gap between test efficacy and effectiveness, given differences in operator expertise, patient characteristics, and test utilization. This gap is particularly notable for ultrasound, which is known to be operator dependent and have worse visualization in patients with obesity and those with nonviral etiologies of cirrhosis. Studies have also demonstrated site-to-site variation in ultrasound performance, potentially driven by differences in ultrasound volume and protocols. These issues with ultrasound were highlighted in our study by the lower sensitivity and specificity of ultrasound among studies conducted in North America compared to those conducted elsewhere. This gap is not surprising given ultrasound in the United

### TABLE 5

Any stage and early stage HCC sensitivity estimates by modality and comparisons of modalities

| Surveillance modality | Any stage sensitivity estimate (95% CrI) | Early stage sensitivity estimate (95% CrI) | Specificity estimate (95% CrI) |
|-----------------------|----------------------------------------|------------------------------------------|-------------------------------|
| AFP alone             | 57.4 (48.1–65.5)                       | 50.0 (38.8–58.8)                        | 88.7 (84.7–92.7)              |
| US alone              | 67.6 (58.5–76.2)                       | 44.0 (27.6–58.4)                        | 90.7 (83.6–96.3)              |
| US+AFP                | 83.9 (68.5–96.7)                       | 63.5 (42.8–83.2)                        | 84.2 (74.1–93.9)              |
| mt-HBT                | 79.6 (63.0–92.9)                       | 69.9 (46.9–84.9)                        | 89.6 (78.8–97.9)              |

| Surveillance modality | Any stage HCC Estimate (95% CrI) | SUCRA | Early stage HCC Estimate (95% CrI) | SUCRA |
|-----------------------|---------------------------------|-------|----------------------------------|-------|
| AFP alone             | 55.1 (49.2–60.7)                | <1%   | 48.6 (34.9–56.1)                 | 26.2% |
| US alone              | 68.6 (59.5–76.4)                | 38.2% | 45.7 (31.9–57.8)                 | 16.0% |
| US+AFP                | 80.0 (64.1–96.4)                | 78.5% | 58.5 (38.5–78.8)                 | 62.8% |
| mt-HBT                | 80.4 (68.1–92.3)                | 83.2% | 70.9 (57.0–83.6)                 | 95.1% |

| Modality comparison | Any stage HCC Difference in sensitivity estimate (95% CrI) | Posterior probability mt-HBT superior | Early stage HCC Difference in sensitivity estimate (95% CrI) | Posterior probability mt-HBT superior |
|---------------------|-----------------------------------------------------------|--------------------------------------|-----------------------------------------------------------|--------------------------------------|
| mt-HBT vs. AFP      | 25.3 (14.7–36.3)                                          | >99%                                 | 22.3 (10.1–34.1)                                          | >99%                                 |
| mt-HBT vs. US       | 11.8 (–1.3–24.3)                                          | 96.2%                                | 25.2 (8.8–44.1)                                           | >99%                                 |
| mt-HBT vs. US+AFP   | 0.4 (–20.2–19.5)                                          | 53.3%                                | 12.3 (–11.3–34.6)                                         | 85.5%                                |

Abbreviations: AFP, alpha-fetoprotein; CrI, credible interval; HCC, hepatocellular carcinoma; mt-HBT, multitarget hepatocellular carcinoma blood test; SUCRA, surface under cumulative ranking curve; US, ultrasound; US+AFP, ultrasound plus alpha-fetoprotein.

*Analysis restricted to North American sites.*
States is conducted by technicians with frozen images subsequently interpreted by radiologists, whereas ultrasound is typically conducted and interpreted concurrently by clinicians in Europe and Asia [29]. Therefore, the need for alternative strategies, such as biomarkers, may vary by site and be driven by local performance of ultrasound-based surveillance rather than pooled global estimates. Further, issues with ultrasound-based surveillance will likely become more prevalent as cirrhosis etiologies shift from being viral hepatitis predominant to an increasing proportion related to alcohol and non-alcoholic steatohepatitis-related cirrhosis. [29] In contrast, the performance of mt-HBT appears to be consistent across liver disease etiologies and obesity classes. [30] Therefore, it will be important to continue monitoring ultrasound performance and reassess the relative benefit of biomarker-based strategies, such as mt-HBT, over time. Finally, patients and providers both report barriers to HCC surveillance, resulting in underuse of surveillance in clinical practice. [31,32] Several observed barriers are specific to imaging-based surveillance, such as scheduling and transportation for a separate radiology appointment, so adherence to blood-based surveillance strategies may be higher. [33] Therefore, despite similar efficacy, blood-based surveillance strategies may have higher effectiveness for early stage HCC detection than ultrasound-based surveillance.

Results of our meta-analysis should be interpreted considering its limitations. First, network meta-analyses are often used in the absence of direct head-to-head comparisons to provide indirect comparisons; however, these results would not obviate the need for prospective validation studies. This is particularly important, as the performance of biomarkers are often overestimated in phase II case–control biomarker studies while phase III and IV validation studies typically provide more accurate estimates. Accordingly, mt-HBT is currently undergoing prospective evaluation in the Performance of a Multitarget HCC Test in Subjects With Increased Risk (ALTUS) study, although these results are not anticipated for years and our data provide important insights in the interim. Second, several included studies had potential for verification bias and dependence between surveillance and diagnostic tests, which could have led to overestimation of surveillance test performance. Third, the systematic review and network meta-analysis included both randomized and observational study designs, and there is a possibility of systematic differences between these study designs. Further, differences in patient characteristics and study settings across studies may contribute to observed heterogeneity and inconsistency, particularly given that ultrasound is operator dependent with site-to-site variation in performance and test accuracy can vary by patient characteristics. Fourth, results of the study also combined test performance across phase II and phase III biomarker studies, with the former potentially overestimating test performance compared to a phase III study. This is particularly important as the mt-HBT validation study was a phase II case–control study, although a prospective validation study is ongoing. [33] Additionally, the studies included in this NMA did not allow for assessment of coherence between direct and indirect comparisons to mt-HBT because the clinical validation study of mt-HBT was compared to AFP. Finally, the binormal ROC model has limitations in its ability to model all shapes of ROC curves and could introduce a bias, particularly if used to estimate sensitivity at a specificity far from the observed specificity. Here, our target specificity of 90% is near the operating specificities of all modalities so any impact would be slight.

In summary, we found that ultrasound alone had poor sensitivity for early stage HCC detection, although this was significantly increased when used in combination with AFP. The mt-HBT achieved similar test performance for early stage HCC detection as ultrasound with AFP, suggesting that it may be another potential emerging tool for HCC surveillance.

**AUTHOR CONTRIBUTIONS**

Study concept and design: Amit Singal, Turgay Ayer. Acquisition of data: Carol Kirshner, Shubham Chakankar, Kyle Porter, Turgay Ayer. Analysis and interpretation of data: Amit Singal, Benjamin Haaland, Neehar Parikh, Burak Ozbay, Carol Kirshner, Kyle Porter, Jagpreet Chhatwal, Turgay Ayer. Statistical analysis: Benjamin Haaland. Study supervision: Turgay Ayer. Drafting of the manuscript: Amit Singal, Turgay Ayer. Critical revision of the manuscript for important intellectual content: all authors. Guarantor of this article: Amit Singal.

**CONFLICTS OF INTEREST**

Amit Singal has served as a consultant on advisory boards for Exact Sciences, FujiFilm Medical Sciences, Glycotest, Roche, Bayer, and GRAIL. Neehar Parikh has served as a consultant or on advisory boards for Exact Sciences, Eli Lilly, Elsai, Genentech, Freenome, Fujifilm Medical Sciences, and Exelixis; he received grants from Glycotest. Burak Ozbay is a full-time employee and shareholder of Exact Sciences. Jagpreet Chhatwal served as a consultant to Bayer, NovoNordisk, and Flatiron Health and is a partner and shareholder of Value Analytics Labs. Turgay Ayer is a partner and shareholder of Value Analytics Labs. Benjamin Haaland consults for the National Kidney Foundation, Value Analytics Health, Avidity Partners Management, and Ally Bridge. Kyle Porter is a shareholder of Exact Sciences. Carol Kirshner has nothing to report.

**ORCID**

Amit G. Singal DOI https://orcid.org/0000-0002-1172-3971
Neehar D. Parikh DOI https://orcid.org/0000-0002-5874-9933
Jagpreet Chhatwal DOI https://orcid.org/0000-0001-8741-4430
REFERENCES
1. Rahib L, Wehner MR, Matrisian LM, Nead KT. Estimated projection of US cancer incidence and death to 2040. JAMA Netw Open. 2021;4:e214708.
2. Singal AG, Zhang E, Narasimman M, Rich NE, Waijek AK, Hoshida Y, et al. HCC surveillance improves early detection, curative treatment receipt, and survival in patients with cirrhosis: a meta-analysis. J Hepatol. 2022;77:128–39.e1.
3. Tzartzeva K, Obi J, Rich NE, Parikh ND, Morad 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.e1.
4. Singal AG, Patibandla S, Obi J, Fullington H, Parikh ND, Yopp AC, et al. Benefits and harms of hepatocellular carcinoma surveillance in a prospective cohort of patients with cirrhosis. Clin Gastroenterol Hepatol. 2021;19:1925–32.e1.
5. Wolf E, Rich NE, Marrero JA, Parikh ND, Singal AG. Use of hepatocellular carcinoma surveillance in patients with cirrhosis: a systematic review and meta-analysis. Hepatology. 2021;73:713–25.
6. Chalasani NP, Porter K, Bhattacharya A, Book AJ, Neis BM, Xiong KM, et al. Validation of a novel multtarget blood test shows high sensitivity to detect early stage hepatocellular carcinoma. Clin Gastroenterol Hepatol. 2022;20:173–82.e7.
7. Feng Z, Pepe MS. Adding rigor to biomarker evaluations: EDRN experience. Cancer Epidemiol Biomarkers Prev. 2020;29:2575–82.
8. Singal AG, Hoshida Y, Pinato DJ, Marrero J, Nault JC, Paradis V, et al. International Liver Cancer Association (ILCA) white paper on biomarker development for hepatocellular carcinoma. Gastroenterology. 2021;160:2572–84.
9. Chalasani N, Horlander JC, Said A, Hoen H, Kopecky KK, Stockberger SM, et al. Screening for hepatocellular carcinoma in patients with advanced cirrhosis. Am J Gastroenterol. 1999;94:2988–93.
10. Gambarin-Gelwan M, Wolf DC, Shapiro R, Schwartz ME, Min AD. Sensitivity of currently available screening tests in detecting hepatocellular carcinoma in cirrhotic patients undergoing liver transplantation. Am J Gastroenterol. 2000;95:1535–8.
11. Piñero F, Marciano S, Anders M, Orozco F, Zerega A, Cabrera G, et al. Screening for liver cancer during transplant waiting list: a multicenter study from South America. Eur J Gastroenterol Hepatol. 2015;27:355–60.
12. Caviglia GP, Ciruolo M, Abate ML, Carucci P, Rolle E, Rosso C, et al. Alpha-fetoprotein, protein induced by vitamin K absence or antagonist II and glypican-3 for the detection and prediction of hepatocellular carcinoma in patients with cirrhosis of viral etiology. Cancers (Basel). 2020;12:3218.
13. Chang TS, Wu YC, Tung SY, Wei KL, Hsieh YY, Huang HC, et al. Alpha-fetoprotein measurement benefits hepatocellular carcinoma surveillance in patients with cirrhosis. Am J Gastroenterol. 2015;110:836–44. Erratum in: Am J Gastroenterol. 2016;111:1668.
14. Curley SA, Izzo F, Gallipoli A, de Bellis M, Cremona F, Parisi V. Identification and screening of 416 patients with chronic hepatitis at high risk to develop hepatocellular cancer. Ann Surg. 1995;222:375–80.
15. Gopal P, Yopp AC, Waijek 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. Clin Gastroenterol Hepatol. 2014;12:870–7.
16. Jasirwan COM, Fahira A, Siregar L, Loho I. The alpha-fetoprotein serum is still reliable as a biomarker for the surveillance of hepatocellular carcinoma in Indonesia. BMC Gastroenterol. 2020;20:215.
17. Lersriwimanmaen P, Nimamong S. Hepatocellular carcinoma surveillance: benefit of serum alpha-fetoprotein in real-world practice. Eurasian J Hepatogastroenterol. 2018;8:83–7.
18. Lok AS, Sterling RK, Everhart JE, Wright EC, Hoefs JC, Di Bisceglie AM, et al. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. Gastroenterology. 2010;138:493–502.
19. Sinn DH, Yi J, Choi MS, Kim YJ, Gwak GY, Lee JH, et al. Serum alpha-fetoprotein may have a significant role in the surveillance of hepatocellular carcinoma in hepatitis B endemic areas. Hepatogastroenterology. 2015;62:327–32.
20. Trevisani F, Cantarini MC, Labate AM, De Notaris S, Rapaccini G, Farinati F, et al.; Italian Liver Cancer (ITALICA) Group. Surveillance for hepatocellular carcinoma in elderly Italian patients with cirrhosis: effects on cancer staging and patient survival. Am J Gastroenterol. 2004;99:1470–6. Erratum in: Am J Gastroenterol. 2004;99:2074.
21. Zapata E, Zubiaurre L, Castiella A, Salvador P, Garcia-Bengoechea M, Esandi P, et al. Are hepatocellular carcinoma surveillance programs effective at improving the therapeutic options. Rev Esp Enferm Dig. 2010;102:484–8.
22. Gupta P, Soundararajan R, Patel A, Kumar-M P, Sharma V, Kalra N. Abbreviated MRI for hepatocellular carcinoma screening: a systematic review and meta-analysis. J Hepatol. 2021;75:108–19.
23. Parikh ND, Mehta AS, Singal AG, Block T, Marrero JA, Lok AS. Biomarkers for the early detection of hepatocellular carcinoma. Cancer Epidemiol Biomarkers Prev. 2020;29:2495–503.
24. Singal AG, Tayob N, Mehta A, Marrero JA, El-Serag H, Jin Q, et al. GALAD demonstrates high sensitivity for HCC surveillance in a cohort of patients with cirrhosis. Hepatology. 2022;75:541–9.
25. Woolen SA, Singal AG, Davenport MS, Troost JP, Khalatbari S, Mittal S, et al. Patient preferences for hepatocellular carcinoma surveillance parameters. Clin Gastroenterol Hepatol. 2022;20:204–15.e6.
26. Schoenberger H, Chong N, Fetzer DT, Rich NE, Yokoo T, Khatri G, et al. Dynamic changes in ultrasound quality for hepatocellular carcinoma screening in patients with cirrhosis. Clin Gastroenterol Hepatol. 2022;20:1561–9.e4.
27. Singal AG, Nehra M, Adams-Huet B, Yopp AC, Tiro JA, Marrero JA, et al. Detection of hepatocellular carcinoma at advanced stages among patients in the HALT-C trial: where did surveillance fail? Am J Gastroenterol. 2013;108:425–32.
28. Fetzer DT, Browning T, Xi Y, Yokoo T, Singal AG. Associations of ultrasound LI-RADS visualization score with examination-, sonographer-, and radiologist-factors: retrospective assessment in over 10,000 examinations. AJR Am J Roentgenol. 2022;218:1010–20.
29. Chong N, Schoenberger H, Yekkaluri S, Fetzer DT, Rich NE, Yokoo T, et al. Association between ultrasound quality and test performance for HCC surveillance in patients with cirrhosis: a retrospective cohort study. Aliment Pharmacol Ther. 2022;55:683–90.
30. Chalasani N, Porter K, Book A, Xiong K, Ramasubramanian TS, Kielar-Grevstad E, et al. The multi-target hepatocellular carcinoma blood test provides high sensitivity for detecting early-stage hepatocellular carcinoma across important patient subgroups. Gastroenterology. 2022;160:2572–84.
31. Singal AG, Tayob N, Mehta A, Marrero JA, El-Serag H, Jin Q, et al. GALAD demonstrates high sensitivity for HCC surveillance in a cohort of patients with cirrhosis. Clin Gastroenterol Hepatol. 2022;20:1561–9.e4.
34. Atiq Q, Tiro J, Yopp AC, Muffler A, Marrero JA, Parikh ND, et al. An assessment of benefits and harms of hepatocellular carcinoma surveillance in patients with cirrhosis. Hepatology. 2017;65:1196–205.

35. Chun S, Rhie SY, Ki CS, Kim JE, Park HD. Evaluation of alpha-fetoprotein as a screening marker for hepatocellular carcinoma in hepatitis prevalent areas. Ann Hepatol. 2015;14:882–8.

36. D’Onofrio M, Faccioli N, Zamboni G, Malagò R, Caffarri S, Fattovich G, et al. Focal liver lesions in cirrhosis: value of contrast-enhanced ultrasonography compared with Doppler ultrasound and alpha-fetoprotein levels. Radiol Med. 2008;113:978–91.

37. Eltabbakh M, Zaghla H, Abdel-Razek W, Elshinnawy H, Ezzat SB. Surveillance for hepatocellular carcinoma: real-life performance in a hepatology outpatient clinic. Swiss Med Wkly. 2015;145:w14200.

38. Henrion J. Surveillance for hepatocellular carcinoma…a hotly debated issue. Acta Gastroenterol Belg. 2004;67:255–64.

39. Ishii M, Gama H, Chida N, Ueno Y, Shinzawa H, Takagi T, et al. Simultaneous measurements of serum alpha-fetoprotein and protein induced by vitamin K absence for detecting hepatocellular carcinoma. South Tohoku District Study Group. Am J Gastroenterol. 2000;95:1036–40.

40. Kim MN, Kim BK, Kim SU, Park JY, Ahn SH, Han HK, et al. Longitudinal assessment of alpha-fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis. Scand J Gastroenterol. 2019;54:1283–90.

41. Oka H, Kurioka N, Kim K, Kanno T, Kuroki T, Mizoguchi Y, et al. Prospective study of early detection of hepatocellular carcinoma in patients with cirrhosis. Ann Hepatol. 2014;13:386–93.

42. Pateron D, Ganne N, Trinchet JC, Aurousseau MH, Mal F, Meicler C, et al. Prospective study of screening for hepatocellular carcinoma in Caucausian patients with cirrhosis. J Hepatol. 1994;20:65–71.

43. Paul SB, Sreenivas V, Gulati MS, Madan K, Gupta AK, Mukhopadhyay S, et al. Incidence of hepatocellular carcinoma among Indian patients with cirrhosis of liver: an experience from a tertiary care center in northern India. Indian J Gastroenterol. 2007;26:274–8.

44. Phulpoto JA, Shah IA, Bhatti Z. Prevalence of hepatocellular carcinoma in cirrhotic patients of Northern Sindh attending liver clinics at Ghulam Mohammad Mahar Medical College Hospitals Sukkur and Khairpur. J Liaquat Uni Med Health Sci. 2012;11:29–33.

45. Poch A, Dieperink E, McMaken KA, Knott A, Thurs P, Ho SB. Surveillance for hepatocellular cancer with ultrasonography vs. computed tomography -- a randomised study. Aliment Pharmacol Ther. 2013;38:303–12.

46. Qian MY, Yuwei JR, Angus P, Schellemann T, Johnson L, Gow P. Efficacy and cost of a hepatocellular carcinoma screening program at an Australian teaching hospital. J Gastroenterol Hepatol. 2010;25:951–6.

47. Ren FY, Piao XX, Jin AL. Efficacy of ultrasonography and alpha-fetoprotein on early detection of hepatocellular carcinoma. World J Gastroenterol. 2006;12:4656–9.