Published in final edited form as:

Clin Gastroenterol Hepatol. 2016 March ; 14(3): 469–475.e2. doi:10.1016/j.cgh.2015.07.049.

Improved detection of hepatocellular carcinoma using a longitudinal α -fetoprotein screening algorithm

Nabihah Tayob, PhD^a, Anna S. F. Lok, MD^b, Kim-Anh Do, PhD^a, and Ziding Feng, PhD^a

^a The University of Texas MD Anderson Cancer Center, Department of Biostatistics, Houston, TX.

^b Division of Gastroenterology and Hepatology, University of Michigan, Ann Arbor, MI.

Abstract

Background and Aims—Hepatocellular carcinoma (HCC) has limited treatment options when diagnosed at advanced stages; therefore early detection is critical to reduce mortality. There is disagreement about the value of α -Fetoprotein (AFP) in HCC surveillance. We aim to improve the sensitivity of AFP in HCC surveillance using an algorithm that incorporates screening history to define patient-specific thresholds for positive screen.

Methods—De-identified data from the Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) Trial, which enrolled 1050 patients with hepatitis C and advanced fibrosis or cirrhosis prospectively followed every 3–6 months, were analyzed. AFP was assayed at each visit and ultrasonography was performed every 6–12 months. A panel adjudicated the diagnosis of HCC. A parametric empirical Bayes (PEB) screening algorithm, which incorporates screening history, was compared to a single threshold (ST) approach for interpreting AFP results.

Results—During a median follow-up of 80 months, 88 patients (48/427 with cirrhosis and 40/621 with advanced fibrosis) were diagnosed with HCC. PEB improved the sensitivity of AFP for detecting all HCC from 60.4% to 77.1% (p-value<0.0005) in patients with cirrhosis and from 72.5% to 87.5% (p-value=0.0015) in patients with advanced fibrosis, when the false positive rate among all screenings was set at 10%. PEB algorithm detected HCC 1.7–1.9 years earlier in the cirrhosis group and 1.4–1.7 years earlier in the advanced fibrosis group, compared to ST approach.

Conclusions—PEB increases the sensitivity of AFP testing and detects HCC earlier among hepatitis C patients with advanced fibrosis or cirrhosis. These data should prompt a reevaluation of how AFP is used in combination with ultrasound in HCC surveillance.

Corresponding Author: Nabihah Tayob, Address: The University of Texas MD Anderson Cancer Center, Department of Biostatistics, 1400 Pressler Street, Floor 4, Houston, TX 77025, ntayob@mdanderson.org, Tel: 713-792-8172.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of interest declared: Anna Lok has no conflicts to declare other than that she was one of the HALT-C investigators. The other authors have no conflicts of interest to declare.

Author Contributions: ZF originated the study concepts. NT, AL, KAD and ZF all contributed to the development of study design. NT analyzed the data, implemented the algorithm, generated tables and figures. All authors reviewed the data analysis and the interpretation of the results. NT wrote the initial manuscript draft, AL revised the manuscript with a focus on medical interpretation, and all authors contributed to the content and final approval of the manuscript.

Keywords

early detection; hepatitis C; cirrhosis; surveillance

Introduction

Liver cancer is the second most common cause of cancer-related deaths worldwide and hepatocellular carcinoma (HCC) accounts for most of the liver cancer cases (1). The five-year survival of patients with HCC is less than 12% (2). The most important risk factor for HCC is cirrhosis (3). One of the barriers to reducing HCC mortality is the late detection of most cases and the lack of treatment options for individuals with advanced stage disease. Very early stage HCC is defined as a single nodule ≤ 2 cm in size; it is asymptomatic and can only be detected through surveillance programs. Early stage HCC is defined as a single nodule ≤ 5 cm in size, or up to three nodules all of which are <3 cm in size. Patients with very early or early stage HCC have multiple treatment options, including surgical resection, liver transplantation and loco-regional therapies with five-year survival between 50-75% (3). Therefore early detection of HCC is critical for reducing HCC mortality.

The American Association for the Study of Liver Diseases (AASLD) practice guidelines recommend ultrasonography surveillance for HCC in patients with cirrhosis (3) but there is disagreement about the benefit of surveillance as there has been little evidence of improved survival in randomized clinical trials (4). However HCC surveillance using ultrasound and/or α -fetoprotein (AFP) testing is widely applied in practice. Strategies that increase the sensitivity of existing surveillance tests will improve the likelihood of early detection of HCC and may provide stronger support for HCC surveillance.

Ultrasonography is reported to have sensitivity of 65-80% and greater than 90% specificity. In the United States, the majority of surveillance ultrasounds are performed at local hospitals with variable quality because ultrasonography is operator dependent, not sensitive in detecting early lesions, difficult to perform in obese patients, and difficult to interpret in the setting of a nodular cirrhotic liver (3). In places with limited resources, surveillance with ultrasonography is not feasible (5). A blood-based surveillance test with high sensitivity would complement ultrasonography as a blood test is more standardized and provides a feasible alternative in limited resource settings.

The reported sensitivity of serum concentration of AFP in both diagnostic and screening settings and across different study designs ranges between 41-100% and specificity between 70-95% (6,7). Most studies evaluating AFP in HCC screening have focused on single AFP measurements at a standard cut-off value for all patients. Lee et al. (8) found that the standard deviation and rate of increase of AFP has improved prognostic accuracy of HCC screening, versus using only the most recent AFP level. While their approach required at least five prior measurements of AFP, it provided evidence that trends in AFP have prognostic value. McIntosh and Urban (9) and Skates et al (10) have proposed screening strategies that use serial biomarker values. These methods have shown potentially large gains in sensitivity and specificity compared to approaches based on a single threshold in

other cancer types, such as CA125 in ovarian cancer (11), and do not require a minimum number of prior screenings.

We hypothesize that a more sensitive AFP algorithm, which incorporates serial values while maintaining a high level of specificity, could increase the early detection of HCC by identifying patients who require further imaging with computed tomography (CT) or magnetic resonance imaging (MRI) despite the lack of suspicious lesions on ultrasonography. The Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) Trial has provided valuable biomarker data to study HCC screening (8,12,13). We evaluated the performance of a serial AFP algorithm to detect HCC in patients enrolled in the HALT-C Trial to determine the potential gains resulting from the incorporation of a patient's screening history versus using only the current AFP value.

Methods

HALT-C Trial Design

The HALT-C Trial enrolled patients with chronic hepatitis C in a randomized controlled trial in which the patients had to have at least stage 3 fibrosis (bridging fibrosis or cirrhosis) by the Ishak scoring system (range 0-6) and a history of failure to respond to previous interferon-based therapy. All patients had radiological imaging to exclude HCC prior to enrollment. The trial aimed to determine if long-term low-dose pegylated interferon therapy was a safe and efficacious treatment in preventing fibrosis progression and clinical outcome, including HCC, and found no reduction in the incidence of HCC compared to no treatment (14). The patients were followed for a median of 80 months and a maximum of 109 months. Patient visits were scheduled every three months during the first 42 months and every six months thereafter. At each visit, patients were evaluated clinically and had local laboratory tests including AFP. Patients also had an ultrasound of the liver at 6, 18, 30 and 42 months post-randomization and every six months thereafter. Patients with elevated AFP or new lesions on ultrasound were further evaluated with CT or MRI. Diagnosis of HCC was based on histology and in the absence of histology, by imaging with or without AFP. A panel of investigators adjudicated all cases of HCC.

There were two subgroups of patients in the HALT-C Trial, those with advanced fibrosis (Ishak score 3 or 4) and those with cirrhosis (Ishak score ≥ 5). Both subgroups of patients were at risk of developing HCC, but the risks differed and the performance of AFP screening may also differ; therefore we examined the screening algorithm separately for each subgroup. Cases were defined as patients with confirmed HCC per the HALT-C adjudication panel. Controls were defined as patients who were not diagnosed with HCC up to the last follow-up visit. Screening visits for controls during the last 12 months of follow-up were excluded from data analyses to ascertain that controls did not have HCC.

Longitudinal algorithm for HCC screening with AFP

The traditional approach to evaluating biomarkers is a single-threshold (ST) screening method that ignores the prior screening history of the patient and indicates a positive screening if the biomarker value exceeds a fixed threshold.

McIntosh and Urban (9) proposed comparing the biomarker level at each screening to a threshold that is individually tailored to that patient. The threshold is based on the parametric empirical Bayes (PEB) estimate, a weighted average of the mean biomarker level in the control population and the average of prior screening values for that patient. This approach combines known patient information with a model for the expected behavior of the biomarker in the control population to evaluate if the patient has a positive screen. The parameters of the model are estimated using data from control patients only and data from cases are not used to develop the algorithm. This avoids bias in estimating sensitivity when the same patient cohort is used to develop the screening algorithm and evaluate its screening performance.

If a patient has no screening history, the PEB and ST rules are equivalent. As the patient accumulates history, the PEB rule depends more on that patient's screening history. An advantage of the PEB algorithm is the ability to learn from prior false positive screening. Patients with stable biomarker trajectories that are consistently higher than average are identified and no longer indicated to have positive screening after a few false positives. The ST rule does not have this ability and will indicate positive screening in patients with biomarker values that are persistently higher than the fixed threshold value even though these values are stable.

For this study, we estimated the false positive rate (FPR) at the screening level, defined as the proportion of positive results among all the screenings conducted in the control group, because each false positive result leads to further testing that can be expensive and may lead to complications and anxiety. The screening-level specificity is defined as $1 - \text{FPR}$. We estimated the true positive rate (TPR) or sensitivity at the patient level, which is defined as the proportion of HCC cases with at least one positive screening during the pre-diagnostic period.

A bootstrap procedure was used to assess the statistical significance of the difference between the PEB algorithm and the ST method. We implemented both screening algorithms in 2000 bootstrap datasets, each of which was constructed by randomly sampling patients from the cohort with replacement. The bootstrap p-values are the proportion of the bootstrap datasets where the difference between the PEB algorithm and ST method is less than or equal to 0.

Results

Data description

Of the 1050 patients in the HALT-C Trial, 1048 patients with HCC outcome data and serial AFP measurements were included in this analysis. Based on the baseline biopsy, 427 were diagnosed with cirrhosis and 621 with advanced fibrosis. Among patients with cirrhosis at baseline biopsy, 361 had more than 12 months of follow-up from enrollment and were not diagnosed with HCC during a median follow-up period of 78 months (range 15-109 months) and 48 had confirmed HCC during study follow-up. Among patients with advanced fibrosis, 558 had more than 12 months of follow-up from enrollment and were not diagnosed with HCC during a median follow-up period of 83 months (range 12-109 months) and 40 had

confirmed HCC during study follow-up. Only 4.3% of patients not diagnosed with HCC were excluded from our analysis because they had less than 12 months of follow-up from enrollment to ensure they are free of HCC; results were essentially the same had they been included. Figure 1 illustrates the subgroups of randomized patients in the HALT-C Trial used in our analysis.

Evaluation of screening algorithm

The screening algorithms were applied to $\log(\text{AFP}+0.01)$ to ensure normality of AFP values. The receiver operating characteristic (ROC) curve displays the patient-level TPR against the screening-level FPR (Figure 2). We observed that the PEB algorithm detected more HCC cases than the ST method in both screening subgroups, with a greater increase in the area under the ROC curve (AUC) within the cirrhosis subgroup (0.84 to 0.93) compared to the advanced fibrosis (0.91 to 0.96) subgroup. In both subgroups, the increase in the AUC is statistically significantly (bootstrap p-value <0.0005).

Various cut-off values for AFP have been used to trigger further evaluation of patients undergoing HCC surveillance. Previous studies in surveillance populations using the most commonly accepted cutoff value of 20ng/ml have shown specificity of 89.8% (15) and 90.0% (16). In this study, using a threshold of 20 ng/ml, the ST method had 77.1% screening-level specificity in the cirrhosis subgroup and 88.1% in the advanced fibrosis subgroup. To ensure a fair comparison between the PEB and ST methods, we identified thresholds that meet fixed screening-level specificities and calculated the patient-level sensitivity.

Screening implemented in a high-risk population can accommodate lower specificity than one used in the general population. We compared the two methods at screening-level specificities 80%, 85% or 90%. The corresponding AFP thresholds for the ST method were 22.3, 29.0 and 42.6 ng/ml, respectively in the cirrhosis subgroup, and 14.0, 16.6 and 22.9ng/ml, respectively in the advanced fibrosis subgroup. At all levels of specificity, the PEB algorithm had higher TPR within each screening subgroup (Table 1). Both methods had superior performance in the advanced fibrosis subgroup compared to the cirrhosis subgroup.

The aim of incorporating the screening history is not only to improve HCC detection overall but to detect HCC earlier when it is more amenable to cure. We compared the time of the first positive PEB test to the first positive ST test in the HCC cases (Table 2). The McNemar test was applied to evaluate the performance of the two methods, with respect to earlier detection of HCC. Across a range of specificities of 80-90%, the PEB algorithm detected HCC on average 1.7-1.9 years earlier in patients with cirrhosis and 1.4-1.7 years earlier in patients with advanced fibrosis.

We examined the timing of a positive screening more closely since a positive screening more than two years prior to clinical diagnosis is unlikely to lead to visualization of the tumor by imaging and more likely reflects AFP as a risk factor. Figures S1 and S2 in Supplementary Materials show PEB and ST screening algorithms with 90% screening-level

specificity within each subgroup. It can be seen that many HCC cases had positive screenings several years prior to diagnosis.

To evaluate the performance of the PEB algorithm in the context of early detection of HCC, we focused on screenings within two years of diagnosis. Figure 3 shows the TPR at the patient level 0-T months prior to diagnosis, where T ranged from 6 to 24 months. The entire screening history was used to estimate the PEB threshold, but only positive screenings 0-T months prior to diagnosis are considered when calculating the TPR. Screening-level specificity was fixed at 90%. Figure 3 shows that the PEB algorithm has a higher TPR than the ST method in both subgroups for all HCC patients and for those HCC patients with early stage tumors. The results for very early stage tumors are based on a total of 12 cirrhosis patients and between 8 and 9 advanced fibrosis patients. The PEB algorithm has a higher TPR than the ST method for most T.

In the HALT-C Trial protocol, AFP was tested every three months during the first 42 months and every six months thereafter. Current guidelines for HCC screening recommend six-month surveillance (17). We evaluated the performance of a six-month surveillance schedule by removing AFP values at month 3, 9, 15, etc. during the first 42 months, thereby mimicking a six-month surveillance approach. Table 3 shows the TPR for the six-month surveillance was lower, compared to using all AFP results (Table 1). The PEB algorithm had a higher TPR than the ST method within each subgroup.

Discussion

Using data from the HALT-C Trial, a large prospective trial involving more than 1000 patients with advanced fibrosis or cirrhosis secondary to chronic hepatitis C infection who were followed for a median of 80 months, we observed statistically significant gains in the sensitivity of AFP when serial biomarker measurements were incorporated into the screening algorithm. In this study, we found that when specificity at the screening-level was set at 90%, i.e. a false positive rate of 10%, the PEB algorithm increased the sensitivity of AFP screening from 60.4% to 77.1% (p-value=0.0005) for patients with cirrhosis and from 72.5% to 87.5% (p-value=0.0020) for patients with advanced fibrosis, compared to the ST method.

One reason why the PEB algorithm, which takes into account each patient's past AFP values, is superior to the ST method is that injury to the liver and cirrhosis in the absence of HCC can increase AFP levels. The PEB algorithm adjusts for the high baseline AFP values in patients with persistently elevated but stable AFP, thereby decreasing the FPR. On the other hand, not all patients with cirrhosis have elevated AFP levels and not all HCC tumors secrete AFP. For patients with low baseline AFP values, the PEB algorithm identifies HCC with small increases in AFP values even though they remain below the cut-point thus increasing the TPR. Elevated baseline AFP values in the absence of HCC (background noise) are more common in patients with cirrhosis than in those with earlier stage liver disease (13), and may explain why both PEB and ST had higher sensitivity in the advanced fibrosis subgroup, though the magnitudes of the improvement by PEB were similar.

An important goal of HCC screening is to detect tumors at an early stage when they are amenable to cure. A positive screening that is too early, when the tumors cannot be visualized by imaging, may not be an indicator of early stage HCC but rather a risk factor for HCC. Indeed, AFP has been found to be a predictor of HCC development in many studies of cirrhosis patients (18). An early positive screening may help to identify patients at high risk of HCC in the near future prompting more frequent surveillance or further evaluation with CT or MRI, which may then lead to earlier diagnosis of the tumor. When we focused on the period within 24 months prior to HCC diagnosis, we found that the sensitivity of the PEB algorithm was between 10.6-14.3% higher in patients with cirrhosis and between 5.1-10.5% higher in those with advanced fibrosis, compared to the ST method.

Six-month intervals have been recommended for HCC surveillance by the AASLD based on tumor doubling times. A multi-center European study involving 1278 patients with cirrhosis, randomized to three-month versus six-month ultrasound surveillance, did not find a difference in survival or in the detection of tumors <2cm or <3cm (19). In our study, the sensitivity of the PEB algorithm was higher when we included all AFP values in the HALT-C Trial, which were assessed every three months in the initial 42 months and every six months thereafter, compared to a six-month screening interval. Our finding of an increased sensitivity of three-month versus six-month AFP screening using the PEB algorithm should be validated before clinical adoption in HCC surveillance programs that utilize AFP.

A limitation of our analysis is that all the patients in the HALT-C Trial had hepatitis C. Further studies including patients with other etiologies of liver disease are needed to determine whether our findings can be generalized. Our study included a subgroup of patients who did not have cirrhosis at baseline. Patients without cirrhosis are usually not recommended to undergo HCC surveillance but all the non-cirrhotic patients in the HALT-C Trial had advanced fibrosis on baseline biopsy and the incidence of HCC in this group was only slightly lower than the subgroup with cirrhosis at baseline (20). Because patients with cirrhosis, who are targets for HCC surveillance, tend to have higher AFP levels even in the absence of HCC, the performance of AFP screening may be different when patients without cirrhosis are included. In this study, we analyzed the two subgroups with and without cirrhosis on baseline biopsy separately and showed that the TPR from the PEB algorithm was higher than that of the ST method in both subgroups. We did not test the PEB algorithm in an external cohort because the PEB algorithm used only data from the controls to calculate the model parameters, thus our analysis is sufficient to validate the detection rate of the PEB algorithm.

In conclusion, we showed that in patients with chronic hepatitis C and advanced fibrosis or cirrhosis, the PEB algorithm, which takes into account serial AFP measurements of the individual patient, has a significantly higher sensitivity in detection of HCC compared to the ST method without sacrificing on specificity. Our findings should prompt a re-evaluation of the role (or lack thereof) of serial AFP in the surveillance of HCC. The use of ultrasonography and serial AFP in combination for HCC surveillance should be validated in a future study. In addition, the availability of biomarker tests with improved sensitivity while retaining a high level of specificity may provide stronger support for HCC surveillance, which is widely practiced but controversial.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant support: Kim-Anh Do and Ziding Feng are partially supported by a Cancer Center Support Grant (NCI Grant P30CA016672). Ziding Feng is partially supported by EDNRN grant (NCI Grant U24086368). Nabihah Tayob is partially supported by start-up and incentive funds to Kim-Anh Do and Ziding Feng.

Abbreviations

AFP	α -Fetoprotein
AASLD	American Association for the Study of Liver Diseases
AUC	Area under the receiver operating characteristic curve
CT	Computed tomography
FPR	False positive rate
HCC	Hepatocellular carcinoma
HALT-C	Hepatitis C Antiviral Long-term Treatment against Cirrhosis
MRI	Magnetic resonance imaging
PEB	Parametric empirical Bayes
ROC	Receiver operating characteristic
ST	Single threshold
TPR	True positive rate

References

1. Ferlay J, Soerjomataram I, Ervik M, et al. Cancer incidence and mortality worldwide: IARC CancerBase No. 11 [Internet]. GLOBOCAN. 2012 v1.0 2013.
2. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med*. 2011; 365(12):1118–1127. [PubMed: 21992124]
3. Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology*. 2005; 42(5):1208–1236. [PubMed: 16250051]
4. Kansagara D, Papak J, Pasha AS, et al. Screening for Hepatocellular Carcinoma in Chronic Liver Disease: A Systematic Review. *Ann Intern Med*. 2014; 161(4):261–269. [PubMed: 24934699]
5. McMahon BJ, Bulkow L, Harpster A, et al. Screening for Hepatocellular Carcinoma in Alaska Natives Infected With Chronic Hepatitis B: A 16-Year Population-Based Study. *Hepatology*. 2000; 32(4):842–846. [PubMed: 11003632]
6. Gupta S, Bent S, Kohlwes J. Test characteristics of α -fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C: A systematic review and critical analysis. *Ann Intern Med*. 2003; 139(1):46–50. [PubMed: 12834318]
7. Gebo KA, Chander G, Jenckes MW, et al. Screening tests for hepatocellular carcinoma in patients with chronic hepatitis C: A systematic review. *Hepatology*. 2002; 36(5B):s84–s92. [PubMed: 12407580]

8. Lee E, Edward S, Singal AG, Lavieri MS, Volk M. Improving screening for hepatocellular carcinoma by incorporating data on levels of α -fetoprotein, over time. *Clin Gastroenterol Hepatol*. 2013; 11(4):437–440. [PubMed: 23247324]
9. McIntosh MW, Urban N. A parametrical empirical Bayes method for cancer screening using longitudinal observations of a biomarker. *Biostatistics*. 2003; 4(1):27–40. [PubMed: 12925328]
10. Skates SJ, Pauler DK, Jacobs IJ. Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. *JASA*. 2001; 96(454):429–439.
11. Drescher CW, Shah C, Thorpe J, et al. Longitudinal screening algorithm that incorporates change over time in CA125 levels identifies ovarian cancer earlier than a single-threshold rule. *J Clin Oncol*. 2013; 31(3):387–392. [PubMed: 23248253]
12. Lok AS, Sterling RK, Everhart JE, et al. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterol*. 2010; 138(2):493–502.
13. Sterling RK, Wright EC, Morgan TR, et al. Frequency of elevated hepatocellular carcinoma (HCC) biomarkers in patients with advanced hepatitis C. *Am J Gastroenterol*. 2012; 107(1):64–74. [PubMed: 21931376]
14. Lok AS, Everhart JE, Wright EC, et al. Maintenance peginterferon therapy and other factors associated with hepatocellular carcinoma in patients with advanced hepatitis C. *Gastroenterol*. 2011; 140(3):840–849.
15. Gopal P, Yopp AC, Waljee AK, et al. Factors that affect accuracy of α fetoprotein test in detection of hepatocellular carcinoma in patients with cirrhosis. *Clin Gastroenterol Hepatol*. 2014; 12(5):870–877. [PubMed: 24095974]
16. Marrero JA, Feng Z, Wang Y, et al. α -Fetoprotein, des- γ -carboxyprothrombin, and lectin-bound α -fetoprotein in early hepatocellular carcinoma. *Gastroenterol*. 2009; 137(1):110–118.
17. Bruix J, Sherman M. Management of hepatocellular carcinoma: An update. *Hepatology*. 2011; 53(3):1020–1022. [PubMed: 21374666]
18. Oka H, Tamori A, Kuroki T, Kobayashi K, Yamamoto S. Prospective study of α fetoprotein in cirrhotic patients monitored for development of hepatocellular carcinoma. *Hepatology*. 1994; 19(1):61–66. [PubMed: 7506227]
19. Trinchet JC, Chaffaut C, Bourcier V, et al. Ultrasonographic surveillance of hepatocellular carcinoma in cirrhosis: A randomized trial comparing 3- and 6-month periodicities. *Hepatology*. 2011; 54(6):1987–1997. [PubMed: 22144108]
20. Lok AS, Seeff LB, Morgan TR, et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease. *Gastroenterol*. 2009; 136(1):138–148.

Figure 1. Standards for Reporting of Diagnostic accuracy (STARD) flow diagram. HCC, hepatocellular carcinoma.
(A) Cirrhosis (B) Advanced Fibrosis

Figure 2.

Receiver operating characteristic (ROC) curve of patient-level true positive rate against screening-level false positive rate for the parametric empirical Bayes (PEB) algorithm (black line) and single threshold (ST) method (gray line). The grey dashed line corresponds to the line of no discrimination. AUC, Area under the ROC curve.

(A) Cirrhosis (B) Advanced Fibrosis

Figure 3.

True positive rate (TPR) at patient-level 0-T months prior to diagnosis of hepatocellular carcinoma, in all patients (solid line), those with early stage disease (dashed line) and those with very early stage disease (dotted line), for the parametric empirical Bayes (PEB) algorithm (black line) and single threshold (ST) method (gray line) when screening-level specificity is 90%. N, N-ES and N-VES are the number of patients with at least one screening in the 0-T months prior to diagnosis of hepatocellular carcinoma in all patients, those with early stage disease and those with very early stage disease respectively.

Table 1

Patient-level true positive rate (TPR) for hepatocellular carcinoma using the parametric empirical Bayes (PEB) algorithm and the single threshold (ST) method.

Screening-level specificity	TPR (Cirrhosis [†] , N=48)		TPR (Advanced fibrosis [‡] , N=40)	
	PEB	ST	PEB	ST
80%	85.4	72.9	95.0	85.0
85%	81.2	68.8	92.5	77.5
90%	77.1	60.4	87.5	72.5

[†]P-values from bootstrap distribution were 0.0035, 0.0110, and <0.0005

[‡]P-values from bootstrap distribution were 0.0545, 0.0100, and 0.0015; for screening-level specificity of 80%, 85%, and 90%, respectively.

Table 2

Number of hepatocellular carcinoma cases for which the first positive test was obtained by either parametric empirical Bayes (PEB), single threshold (ST) or at the same time for both approaches; and number for which no positive tests were obtained by either PEB or ST method.

Screening-level specificity	Cirrhosis [†] (N=48)				Advanced fibrosis [‡] (N=40)			
	PEB first	Same time	ST first	Neither positive	PEB first	Same time	ST first	Neither positive
80%	18	23	0	7	12	23	4	1
85%	19	20	0	9	13	22	2	3
90%	21	16	0	11	15	20	0	5

[†]P-values from McNemar test all <0.001

[‡]P-values from McNemar test were 0.08, 0.009 and <0.001; for screening-level specificity of 80%, 85%, and 90%, respectively.

Table 3

Patient-level true positive rate (TPR) for hepatocellular carcinoma using a 6-month surveillance interval. PEB, parametric empirical Bayes; ST, single threshold.

Screening-level specificity	Cirrhosis		Advanced fibrosis	
	TPR [†] (N=48)		TPR [‡] (N=40)	
	PEB	ST	PEB	ST
80%	83.3	70.8	95.0	77.5
85%	81.2	66.7	90.0	75.0
90%	68.8	56.2	85.0	70.0

[†]P-values from Bootstrap distribution were 0.0085, 0.0125, and 0.0075

[‡]P-values from Bootstrap distribution were 0.0120, 0.0045, and <0.0005; for screening-level specificity of 80%, 85%, and 90%, respectively.