The Doylestown Algorithm: A Test to Improve the Performance of AFP in the Detection of Hepatocellular Carcinoma

Mengjun Wang[1], Karthik Devarajan[2], Amit G. Singal[3], Jorge A. Marrero[3], Jianliang Dai[4], Ziding Feng[4], Jo Ann S. Rinaudo[5], Sudhir Srivastava[5], Alison Evans[6], Hie-Won Hann[7], Yinzhi Lai[7], Hushan Yang[7], Timothy M. Block[8], and Anand Mehta[1]

Abstract

Biomarkers for the early diagnosis of hepatocellular carcinoma (HCC) are needed to decrease mortality from this cancer. However, as new biomarkers have been slow to be brought to clinical practice, we have developed a diagnostic algorithm that utilizes commonly used clinical measurements in those at risk of developing HCC. Briefly, as α-fetoprotein (AFP) is routinely used, an algorithm that incorporated AFP values along with four other clinical factors was developed. Discovery analysis was performed on electronic data from patients who had liver disease (cirrhosis) alone or HCC in the background of cirrhosis. The discovery set consisted of 360 patients from two independent locations. A logistic regression algorithm was developed that incorporated log-transformed AFP values with age, gender, alkaline phosphatase, and alanine aminotransferase levels. We define this as the Doylestown algorithm. In the discovery set, the Doylestown algorithm improved the overall performance of AFP by 10%. In subsequent external validation in over 2,700 patients from three independent sites, the Doylestown algorithm improved detection of HCC as compared with AFP alone by 4% to 20%. In addition, at a fixed specificity of 95%, the Doylestown algorithm improved the detection of HCC as compared with AFP alone by 2% to 20%. In conclusion, the Doylestown algorithm consolidates clinical laboratory values, with age and gender, which are each individually associated with HCC risk, into a single value that can be used for HCC risk assessment. As such, it should be applicable and useful to the medical community that manages those at risk for developing HCC. *Cancer Prev Res; 9(2); 172–9. ©2015 AACR.*

Introduction

Hepatocellular carcinoma (HCC) is the second leading cause of cancer-related death worldwide and the leading cause of death in patients with cirrhosis (1). Globally, hepatitis B virus (HBV) infection is the leading cause of HCC, whereas most HCC cases in the United States are related to hepatitis C virus (HCV) infection (2, 3).

Prognosis for HCC patients is related to tumor stage at time of diagnosis, with higher rates of curative treatment and better overall survival among those with early-stage tumors. Therefore,

HCC surveillance has been recommended in at-risk patients using ultrasonography, with or without serum levels of the oncofetal glycoprotein, α-fetoprotein (AFP; refs. 4, 5). However, there has been extensive debate about the utility of AFP given its suboptimal sensitivity and specificity (6–8). Thus, there has been a great desire to identify new molecules that could be used as biomarkers for HCC (8–16). We have previously utilized novel biostatistical methods to develop algorithms using biomarkers and basic clinical information that can improve early HCC detection (17, 18). Although highly accurate, these algorithms included experimental biomarkers that are years away from being widely available. Therefore, in the current study, we evaluated the performance of an algorithm using just AFP and clinical information and compared it to the performance of AFP alone for early HCC detection.

Materials and Methods

Study populations

Clinical data from nested case–control studies from the University of Michigan (UM) and the HALT-C study (see below) were used as a discovery set to develop the Doylestown algorithm. For the UM cohort, patients with cirrhosis were enrolled from UM Liver Clinics between September 2001 and August 2004, with the full protocol described in detail elsewhere (19). Diagnosis of cirrhosis was based on liver histology or clinical, laboratory, and imaging evidence of hepatic decompensation or portal hypertension. Patients with a liver mass on ultrasound or elevated serum AFP were required to have an MRI without evidence of HCC

[1]Drexel University College of Medicine, Philadelphia, Pennsylvania. 19102. [2]Fox Chase Cancer Center, Philadelphia, Pennsylvania. [3]Division of Digestive and Liver Diseases, University of Texas Southwestern, Dallas, Texas. [4]Department of Biostatistics, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas. [5]Cancer Biomarkers Research Group, Division of Cancer Prevention, NCI, Bethesda, Maryland. [6]Drexel University School of Public Health, Philadelphia, Pennsylvania. [7]Thomas Jefferson University, Philadelphia, Pennsylvania. [8]The Baruch Blumberg Institute, Doylestown, Pennsylvania.

Note: Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerpreres.aacrjournals.org/).

Corresponding Author: Anand Mehta, Drexel University College of Medicine, Room 18307 New College Building, 245 North 15th Street, Philadelphia, PA 19102. Phone: 215-762-7586; Fax: 215-489-4920; E-mail: anand.mehta@drexelmed.edu

doi: 10.1158/1940-6207.CAPR-15-0186

AACR

within 3 months prior to enrollment or 6 months after enrollment. For this study, patients who developed HCC during follow-up were used as cases and age- (±10 years) and gender-matched patients with cirrhosis served as controls. The diagnosis of HCC was made by histopathology, including all T1 lesions, or by two imaging modalities MRI or CT, showing a vascular enhancing mass > 2 cm with delayed washout. Cirrhosis controls were followed for a median of 12 months (range, 7–18 months) after enrollment to confirm absence of HCC. A 20-mL blood sample was drawn from each subject, spun, aliquoted, and serum stored at −80°C until testing. Blood samples from HCC patients were drawn prior to initiation of HCC-directed treatment. AFP was tested using commercially available immunoassays utilizing enhanced chemiluminescence at the UM Hospital Clinical Diagnostic Laboratory. The UM's Institutional Review Board approved the study protocol. Patient information is provided in Supplementary Table S1.

HALT-C cohort. The clinical values from the UM data set were combined with data from a selected set of patients from the HALT-C study to develop the Doylestown algorithm (see below). The design of the HALT-C study, including inclusion criteria, as well as cirrhosis and/or HCC diagnostic criteria are described in a recent publication in great detail (20). For our study, 151 individuals (49 HCC cases and 102 HCV non-HCC controls) were examined (21–24). As this was a longitudinal study, for the HCC cases, data the time closest to HCC diagnosis were used. This was generally 0 to 3 months prior to HCC detection as described in ref. 20. More information is found in the main HALT-C publication (20). The study was performed in compliance with and after approval from the respective institutional review boards of all sites. Patient information is provided in Supplementary Table S2.

Early detection research network (EDRN) cohort. The first validation cohort consisted of 870 patients (432 HCC cases and 438 non-HCC cirrhosis controls) enrolled in the NCI EDRN study (25). The description below is taken from the recent publication describing this cohort (24). Briefly, cases included consecutive adult patients with HCC seen between February 2005 and August 2007 at seven medical centers in the United States (25). The study was performed in compliance with and after approval from the respective institutional review boards of all sites. A complete blood count, a liver panel, and AFP level were obtained at the local clinical center at each visit using standard procedures and methods. Full information regarding this group is found in ref. 25. Patient information is provided in Supplementary Table S3. Briefly, the cirrhosis controls were younger than those with early HCC ($P < 0.0001$), and there was a male predominance in all groups and a predominance of white ethnicity in cirrhotic controls and HCC cases. The majority of cases and controls had a viral etiology of their liver disease, with HCV in 61% controls and 51% HCC cases, of which 58% had early-stage HCC (BCLC stage 0 or BCLC stage A). HBV was the underlying etiology of liver disease in 5% cirrhosis controls and 16% of the HCC cases, of which 16% were early stage (BCLC stage 0 and BCLC stage A). Early stage was defined by a single lesion between 2 and 5 cm or ≤3 lesions each ≤3 cm, without portal vein thrombosis or extrahepatic metastasis (25).

Data from the Thomas Jefferson University. The second validation study used data from Thomas Jefferson University (TJU), consisting of 699 patients (113 HBV-related HCC and 586 HBV-positive controls). The patients were identified from an existing clinic-based patient cohort, which has been described in detail elsewhere (26). Briefly, this set contained Asian American patients who had HCC induced by chronic HBV infection (excluding all other etiologies) or HBV-infected patients without HCC (excluding coinfection with HCV). Thus, both cases and controls were treated according to American association for the study of liver disease (AASLD) guidelines for their HBV and thus were DNA negative. Patients without complete records of the analyzed variables [i.e., age, gender, AFP, alanine aminotransferase (ALT), and ALP] were excluded. Serum levels of AFP, ALT, and alkaline phosphatase (ALK) were determined using commercially available kits at the Thomas Jefferson Hospital or other clinical diagnostic laboratories. Patient information is provided in Supplementary Table S4.

University of Texas Southwestern (UTSW) cohort. The third validation cohort used data from UTSW and the Parkland Health and Hospital System, consisting of 1,229 patients (425 HCC cases and 804 cirrhosis controls). Patient recruitment has been previously described in detail (27). In brief, patients with HCC were identified using ICD9 codes and lists of patients seen in a multidisciplinary HCC clinic, with all cases adjudicated to confirm they met AASLD criteria. Patients with cirrhosis were identified using ICD9 codes and adjudicated to confirm the presence of cirrhosis on imaging. All control patients were required to have 6 months of follow-up to confirm absence of HCC. Serum AFP and labs were determined using commercially available immunoassays at UTSW. Patient data collection and the study protocol were approved by the Institutional Review Board at the UT Southwestern Medical Center. Patient information is provided in Supplementary Table S5.

Statistical methods

Data sets. As stated, a data set utilizing samples from the UM (Supplementary Table S1) and HALT-C (Supplementary Table S2) was used for feature selection and algorithm development. This approach was adopted to increase the statistical learning space and to ensure the development of robust algorithms. Patients without complete records of the analyzed variables (i.e., age, gender, AFP, ALT, and ALP) were excluded.

We applied univariate logistic regression to check the association of each predictor with HCC. We also applied multivariate logistic regression to check the association of each predictor with HCC or cirrhosis alone adjusting for the effects of remaining predictors, shown in Supplementary Table S6. More information on feature selection and analysis is provided in the Supplementary Materials and Methods.

Building the Doylestown algorithm. We applied logistic regression with the subsets of predictors. There were 21 subset features that were selected from the feature selection algorithms and market basket analysis. In addition, we added a full predictors subset and AFP-alone subset to be conference subsets, and from this, we built 23 logistic regression algorithms. To judge the fitness of each regression, we derived AIC, R^2, Dxy, likelihood ratio test, Pearson goodness-of-fit, log-likelihood, deviation statistic, and the area under ROC curve (AUROC) of apparent validation (28). These results are shown in Supplementary Table S7.

To avoid overfitting, we applied leave-one-out cross-validation, bootstrap validation, and 3-fold cross-validation to validate the
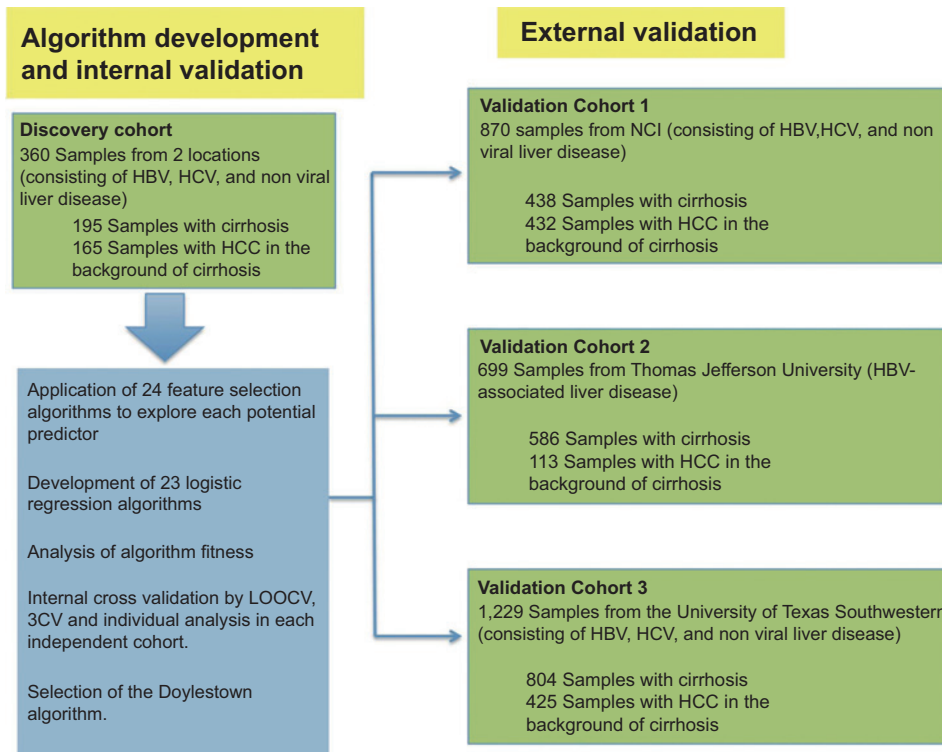
Algorithm development and internal validation

Discovery cohort
360 Samples from 2 locations (consisting of HBV, HCV, and non viral liver disease)

195 Samples with cirrhosis
165 Samples with HCC in the background of cirrhosis

Application of 24 feature selection algorithms to explore each potential predictor

Development of 23 logistic regression algorithms

Analysis of algorithm fitness

Internal cross validation by LOOCV, 3CV and individual analysis in each independent cohort.

Selection of the Doylestown algorithm.

External validation

Validation Cohort 1
870 samples from NCI (consisting of HBV,HCV, and non viral liver disease)

438 Samples with cirrhosis
432 Samples with HCC in the background of cirrhosis

Validation Cohort 2
699 Samples from Thomas Jefferson University (HBV-associated liver disease)

586 Samples with cirrhosis
113 Samples with HCC in the background of cirrhosis

Validation Cohort 3
1,229 Samples from the University of Texas Southwestern (consisting of HBV, HCV, and non viral liver disease)

804 Samples with cirrhosis
425 Samples with HCC in the background of cirrhosis

Figure 1.
Study design. Model development utilized 360 samples with HBV, HCV, and nonviral liver disease. After model development and internal validation, external validation was performed by independent analysis of the Doylestown algorithm in three sample sets consisting of over 2,700 patient samples. Samples consisted of those with HBV, HCV, and nonviral liver disease.

10 candidate models (Supplementary Tables S8–10). Based upon the performance of the cross-validation and the properties of the calibration, the model with logAFP, age, gender, ALK, and ALT was selected for further development. We refer to this as the Doylestown algorithm (Supplementary Table S11). More information on the models and methods used for algorithm development is provided in the Supplementary Materials and Methods. Other models such as conditional inference tree or classification and regression tree were tried as well, but this performance was inferior to that obtained with AFP in a logistic regression analysis (shown in Supplementary Table S12).

External validation. For external validation, the Doylestown algorithm was sent as an equation (as shown in Fig. 2A) to our collaborators. All selection of patients, application of the algorithm, and data analysis were performed at the specific external validation sites.

Results

Model development and performance in training set

In our previous efforts to develop noninvasive tests for the early detection of HCC, we had utilized a combination of novel protein and glycomic markers with AFP to detect HCC in the background of liver cirrhosis (17, 18). However, it was noticed that the performance of AFP alone was improved through inclusion of factors such as age or gender in the algorithm. Thus, we examined the performance of an algorithm that contained AFP values along with several clinical variables but excluded our novel biomarkers and compared this to the performance obtained with AFP alone. The study design is shown in Fig. 1. Supplementary Table S6 shows the 10 clinical factors analyzed, of which 5 were found to be

associated with HCC, which included age, gender, ALK, ALT and log-transformed AFP values. The logistic regression equation is presented in Fig. 2A. Before external validation, we tested the algorithm in the two data sets independently to determine how the algorithm improved the performance of AFP. Briefly, in just the 209 patients from the UM patient set, the mean value of AFP was 11.8 ng/mL (SD, 34.6) in patients with cirrhosis and 9657.6 ng/mL (SD, 3975.3) in patients with HCC. In this initial analysis, the AUROC of AFP was increased from 0.8398 (95%CI, 0.7870–0.8926) with AFP alone to 0.9388 (95% CI, 0.9103–0.9674) with the Doylestown algorithm (Fig 2B). Importantly, when only those patients that had early-stage cancer were examined, the AUROC was increased from 0.7983 (95% CI, 0.7251–0.8715) for AFP alone to 0.9491 (95% CI, 0.9138–0.9843) for the Doylestown algorithm.

When this algorithm was utilized on just the 151 samples from the HALT-C study, as shown in Fig. 2C, the performance of AFP was increased from 0.8153 (95%CI, 0.7430–0.8875) with AFP alone to 0.8533 (95%CI, 0.7912–0.9153) for the Doylestown algorithm. If only early cancers ($n = 39$) were used, the performance of AFP was 0.8026 (95%CI, 0.7192–0.8860) and 0.8339 (95%CI, 0.7627–0.9052) for the Doylestown algorithm. Although the increase was smaller (5%) than observed with the UM data set, this difference was statistically significant ($P < 0.0001$).

Performance in independent external validation sets

The potential of this algorithm was further tested through blinded external validation in three sample sets consisting of over 2,700 patients, which allowed for greater analysis of the algorithm in those with early HCC. Validation cohort 1 (Fig. 1) consisted of a large multicenter case–control study collected by
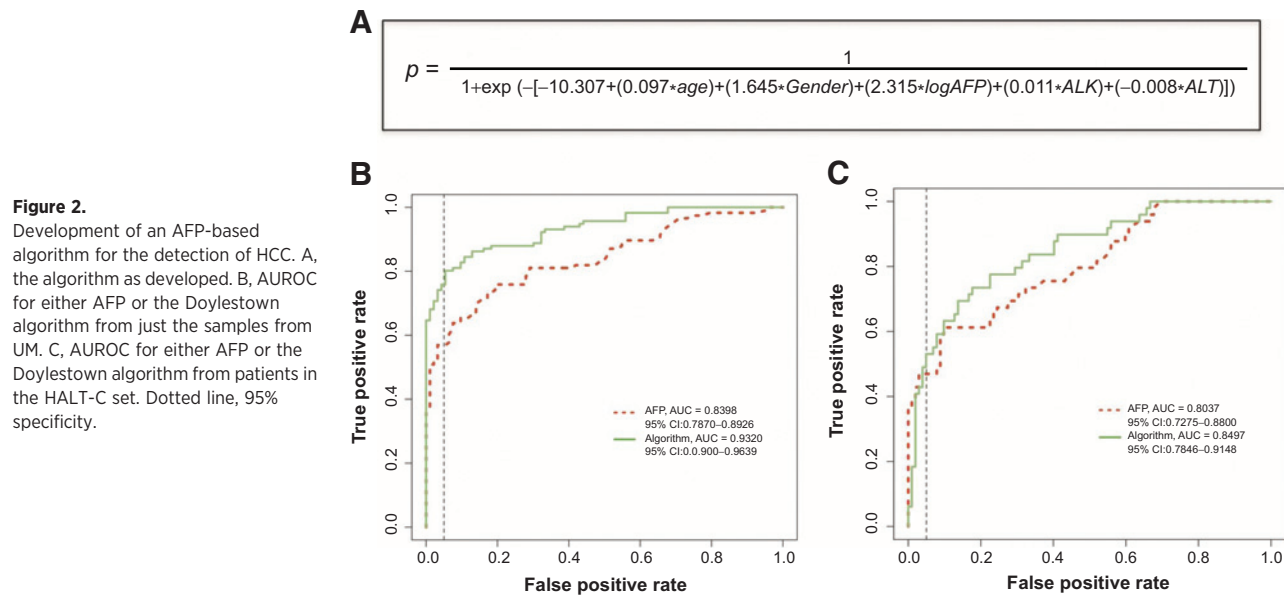
A

$$p = \frac{1}{1+\exp\left(-[-10.307+(0.097*age)+(1.645*Gender)+(2.315*logAFP)+(0.011*ALK)+(-0.008*ALT)]\right)}$$

Figure 2.
Development of an AFP-based algorithm for the detection of HCC. A, the algorithm as developed. B, AUROC for either AFP or the Doylestown algorithm from just the samples from UM. C, AUROC for either AFP or the Doylestown algorithm from patients in the HALT-C set. Dotted line, 95% specificity.

B

AFP, AUC = 0.8398
95% CI:0.7870–0.8926
Algorithm, AUC = 0.9320
95% CI:0.0.900–0.9639

C

AFP, AUC = 0.8037
95% CI:0.7275–0.8800
Algorithm, AUC = 0.8497
95% CI:0.7846–0.9148

the EDRN of the NCI. This case–control study consists of 870 patients, 438 patients with liver cirrhosis and 432 patients with HCC (Supplementary Table S3). In this study, AFP had an AUROC of 0.8109 in the detection of all HCC. Consistent with the derivation cohort, the Doylestown algorithm increased the AUROC to 0.8409 (Fig. 3A). This increase was statistically significant ($P < 0.0001$). In addition, if only patients with early cancers ($n = 225$) were examined, as Fig. 3B shows, a similar increase in performance was seen (0.7856–0.8104). Importantly, in this group of patients with early-stage HCC, at a fixed specificity of 95%, the sensitivity was increased from 31% for AFP alone to 43% with the Doylestown algorithm. Thus, consistent with the previous analysis in a case–control study, the application of the Doylestown algorithm could increase the detection of HCC and, importantly, was able to increase the detection of early tumors in potentially a clinically meaningful way, without any detrimental impact on specificity.

As the EDRN validation set primarily consisted of patients with HCV-associated cirrhosis, we wanted to ensure that a similar performance could be obtained in HBV-associated liver disease (validation cohort 2). This is important given patients with chronic HBV comprise the largest at-risk group worldwide, with high particularly high rates in Asia and Africa. Although antiviral therapy significantly reduces the incidence of liver cancer in these patients (~50% reduction), the risk remains very high, almost 20- to 30-fold higher than the normal population (29). Therefore, these patients will continue to require surveillance for HCC. Thus, the second external patient cohort examined was from TJU and consisted of those with HBV-associated liver disease and treated for their infection following AASLD guidelines and were DNA negative at the time of the study. This set comprised of 699 patients, 113 that had HBV-associated early HCC and 586 of which had chronic HBV infections (Supplementary Table S4). In this group, AFP had a mean value of 20.0 ng/mL (SD, 72.6) in the control group and 1568.3 ng/mL (SD, 6626.7) in the HCC group. As shown in Fig. 3C, similar to the other studies performed, the AUROC of AFP alone was 0.8257

(95% CI, 0.7877–0.8637) when differentiating HCC from the HBV disease group. Consistent with the previous data, the Doylestown algorithm significantly increased the AUROC 7% to 0.8920 (95% CI, 0.8633–0.9206). Again, this difference was statistically significant ($P < 0.0001$) and highlights the ability of this algorithm to improve the performance of AFP over a wide range of diseases and conditions.

Validation cohort 3 was from the UTSW and consisted of 1,229 patients—804 with a background of liver cirrhosis and 425 with HCC (Supplementary Table S5). AFP had a mean value of 12 ng/mL (SD, 39.0) in the control cirrhotic group and 23,681 ng/mL (SD, 116,731) in the HCC group. As Fig. 4A shows, unlike the other patient groups examined, AFP alone had an AUROC of 0.877 in the differentiation of cirrhosis from HCC. The Doylestown algorithm did not change this and had an AUROC of 0.876, which was not statistically significant ($P = 0.9328$). When only patients with early-stage HCC were examined ($n = 139$), AFP alone had an AUROC of 0.7898. Surprisingly, the Doylestown algorithm did not alter this and had an AUROC of 0.7709.

However, in our analysis of this sample set (see Supplementary Table S5), it was noticed that many patients had very high levels of AFP, with a mean level over 23,000 ng/mL and 154 patients with AFP values >1,000 ng/mL, all of which had HCC. Thus, when AFP is already elevated to such a high level, this algorithm appears to have limited impact. In addition, a large proportion of patients ($n = 763$) had AFP <10 ng/mL (the mean level of AFP in this group, 4.00 ng/mL; SD, 2.17). Not surprisingly, in these patients, the Doylestown algorithm did not alter the detection of HCC (AUROC of 0.6313 for AFP and 0.6417 for the Doylestown algorithm). In contrast, the Doylestown algorithm had the greatest benefit for those with AFP in the range of 10 to 100 ng/mL, where the AUROC was increased from 0.579 for AFP alone to 0.700 for the Doylestown algorithm. As Fig. 4B shows, when patients were broken down into specific groups based on the AFP level, the Doylestown algorithm increased the AUROC in almost all groups, from those with AFP between 10 and 100 ng/mL all the way to AFP levels between 10 and 10,0000 ng/mL. As expected, at
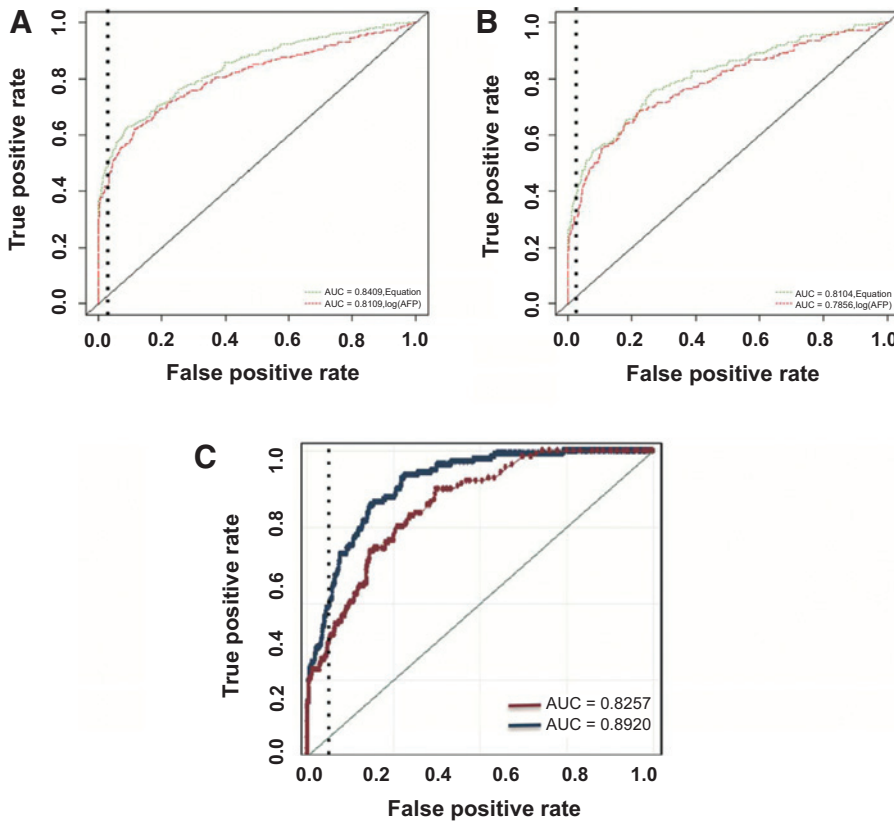
Figure 3.
Validation of the Doylestown algorithm in the NCI EDRN sample set and in an HBV-infected sample set from TJU. A, AUROC of AFP or the Doylestown algorithm in the NCI EDRN validation set in all patients. B, AUROC of AFP or the Doylestown algorithm in the NCI EDRN validation set in patients with early-stage HCC. C, AUROC of AFP versus the Doylestown algorithm in the TJU validation group. Dotted line, 95% specificity.

higher levels of AFP, the AUROC increase was less, and no further increases were observed when patients with AFP >100,000 ng/mL were included. An examination of patients with early-stage HCC was done to see the performance of the algorithm in this subgroup of patients. In patients with early-stage HCC AFP had an AUROC of 0.578 and the Doylestown increased this to 0.629. In contrast, if only patients with late-stage HCC and AFP between 10 and 100 ng/mL were examined, AFP had an AUROC of 0.578 and the Doylestown algorithm increased this to 0.756. To see if similar increases in HCC detection were observed in the other sets, we reevaluated the performance of the Doylestown algorithm in the TJU patients only with AFP in the zone of 10 to 100 ng/mL (all of these patients had early HCC). Consistent with the results shown in Fig. 4C, the AUROC was increased in the TJU group from 0.5308 for AFP alone to 0.7940 for the Doylestown algorithm (n = 104 controls and 40 cases; Fig. 4D). Consistent with this, an examination of the EDRN set reveled that the AUROC increased from 0.6439 to 0.7591 in patients with AFP between 10 and 100 ng/mL (n = 106 controls and 202 cases; Fig. 4E) when patients with all stages of HCC were examined. When only patients with early-stage cancer (n = 109) who had AFP between 10 and 100 ng/mL were examined the AUROC also increased from 0.641 to 0.773 (see Fig. 4F).

In the discovery set, a similar increase was observed. For example, in the UM set, a similar result was seen, where AFP's AUROC in those with AFP between 10 and 100 ng/mL was 0.6636 and this was increased to 0.9110 with the Doylestown algorithm. And in the HALT-C data set, the AUROC of AFP in patients with

AFP between 10 and 100 ng/mL was 0.6583 and this was increased by the Doylestown algorithm to 0.7077.

Discussion

In this article, we demonstrated the usefulness of incorporating biomarkers and relevant clinical variables into a statistical model for predicting the incidence of HCC. Specifically, we investigated the predictive performance of AFP alone or after the inclusion of clinical factors, such as age, gender, and serum ALK, and ALT levels. As shown, the inclusion of these clinical variables increased the AUROC of AFP 4% to 12% and had equal benefit regardless of tumor size or the etiology of liver disease. It is also important to note that the inclusion of these factors did not have a detrimental impact on the specificity of AFP. For example, in the HALT-C control group, no patient who had an AFP of <20 ng/mL was misclassified by the Doylestown algorithm (i.e., no increase in false positives). In contrast, of the 20 patients within the HALT-C control group who had AFP values greater than 20 ng/mL and were misclassified by AFP as having HCC, the Doylestown algorithm correctly reclassified 12 of these (60%). Additionally, as Table 1 shows, at a fixed specificity of 95%, the Doylestown algorithm improved the sensitivity in all the studies performed—even in cases where AFP already performed strongly. Thus, we strongly believe that this algorithm could be used as a simple replacement for AFP with immediate clinical benefit and, more importantly, without any harm to the patients.

Figure 4.
Validation of the Doylestown algorithm in the UTSW patient set. A, AUROC of AFP or the Doylestown algorithm in the UTSW sample set. B, AUROC of either AFP alone or the Doylestown algorithm in patients with varying ranges of AFP. For graph, the Y axis is the AUC for either AFP or the Doylestown algorithm in the specified patients. For the X axis, group 0 are patients with AFP<10; group 1 are patients with 10<AFP≤100; group 2 are patients with 10<AFP≤200; group 3 are patients with 10<AFP≤300; group 4 are patients with 10<AFP≤400; group 5 are patients with 10<AFP≤500; group 6 are patients with 10<AFP≤600; group 7 are patients with 10<AFP≤700; group 8 are patients with 10<AFP≤800; group 9 are patients with 10<AFP<900; group 10 are patients with 10<AFP≤1,000; group 11 are patients with 10<AFP< = 10,000; group 12 are patients with AFP≤100,000; group 13 are all patients. In all cases, AFP values are ng/mL. C, AUROC of the Doylestown algorithm and AFP in the UTSW set only in patients with AFP between 10 and 100 ng/mL. D, AUROC of the Doylestown algorithm and AFP in the TJU set only in patients with AFP between 10 and 100 ng/mL. E, AUROC of the Doylestown algorithm and AFP in the EDRN set only in patients with AFP between 10 and 100 ng/mL. F, AUROC of the Doylestown algorithm and AFP in the EDRN set only in patients with early-stage HCC and AFP between 10 and 100 ng/mL. Dotted line, 95% specificity.
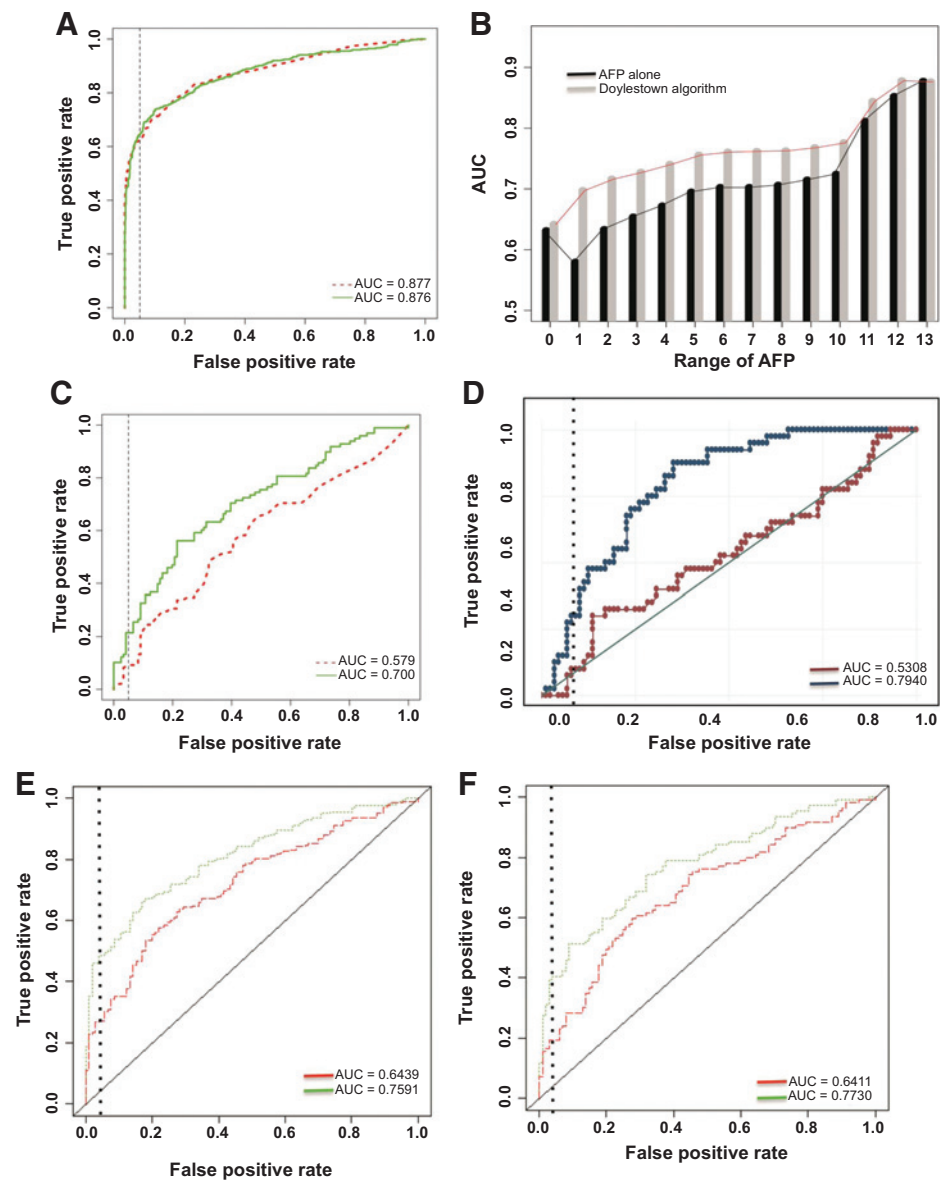
Several recent reports have described similar algorithms that contain many of the same factors presented here. Most notably, El-Serag and colleagues have recently described an algorithm to predict HCC in patients with HCV and cirrhosis (30). While our system contains many of the same factors (AFP, age, ALT), it was developed and tested in individuals with HCV, HBV, and patients with nonviral liver disease and thus expanded upon the work presented by El-Serag, and colleagues, which only examined HCV patients. Our analysis also included both internal and external validation from multiple sources, which was different from the El-Serag study. However, both of these studies clearly indicate that improvements to AFP can be attained through the inclusion of clinical variables into a simple algorithm to increase the detection of HCC.

One concern with the algorithm performance is the potential variation in the clinical testing of these factors. The performance of tests such as ALK, ALT, and AFP can vary when performed in one laboratory to another. This interassay performance variation could theoretically impact the ability of an algorithm to correctly classify a patient. In model development, given a fixed age and gender, assay variations in all three continuous variables of up to 15% can occur without misclassification, with greater variation tolerated in individual markers. However, the true flexibility will only be determined when the clinical community uses the model.

The data presented in this article also have several limitations that will have to be addressed in future studies. The first is potential selection bias in the external validation. That is, only patients with the required clinical factors were used in the external validation. It is possible that this imparted some selection bias that may have affected the results. In addition, this study was done with clinical information collected either at the time of HCC detection or close to it. Thus, a longitudinal study will have to be performed to truly determine how this

Wang et al.

Table 1. Sensitivity of AFP or the Doylestown algorithm at a fixed specificity of 95%

Sample source	Samples, N	HCC samples	Control samples	AFP sensitivity at 95% specificity	Doylestown algorithm sensitivity at 95% specificity	P
UM[a]	209	116	93	55%	75%	<0.0001
UM early-stage HCC only[b]	153	60	93	48%	88%	<0.0001
HALT-C[c]	151	49	102	43%	55%	<0.0001
HALT-C—Early-stage HCC only[d]	141	39	102	36%	41%	<0.0001
EDRN[e]	870	432	438	42%	53%	<0.0001
EDRN—Early-stage only[f]	656	224	432	31%	43%	<0.0001
TJU[g]	699	113	586	38%	58%	<0.0001
UTSW[h]	1229	425	804	61%	63%	0.9378
UTSW—Early-stage only[i]	943	139	804	40%	35%	>0.05
UTSW only 10–100 ng/mL of AFP[j]	229	98	121	10%	25%	<0.0001
EDRN only 10–100 ng/mL of AFP[j]	308	202	106	24%	47%	<0.0001
TJU only 10–100 ng/mL of AFP[j]	128	40	88	12%	27%	<0.0001

[a]The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the UM training set.
[b]Sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the UM training set in only those with early-stage HCC.
[c]Sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the HALT-C training set.
[d]Sensitivity of either AFP or the Doylestown algorithm at 95% specificity in only those with early-stage HCC.
[e]The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the EDRN validation set.
[f]The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the EDRN validation set in only those with early-stage HCC
[g]The sensitivity of either AFP or the Doylestown algorithm at 90% specificity in the TJU set.
[h]The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in whole University of Texas set.
[i]The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in patients with early-stage HCC.
[j]The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in patients with AFP between 10 and 100 ng/mL in each of the respective validation cohorts.

algorithm would be used in the management of patients at risk of developing HCC.

Disclosure of Potential Conflicts of Interest

T. Block reports receiving commercial research grant from Arbutus Bio-Pharma, has ownership interest (including patents) in Glycotest, and is consultant/advisory board member for Glycotest. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: M. Wang, K. Devarajan, S. Srivastava, A. Evans, H.-W. Hann, T.M. Block, A. Mehta
Development of methodology: M. Wang, K. Devarajan, S. Srivastava, A. Mehta
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.G. Singal, J.A. Marrero, Z. Feng, H.-W. Hann, H. Yang, A. Mehta
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Wang, K. Devarajan, A.G. Singal, J.A. Marrero, J. Dai, Z. Feng, A. Evans, Y. Lai, H. Yang, T.M. Block, A. Mehta

Writing, review, and/or revision of the manuscript: M. Wang, K. Devarajan, A.G. Singal, J.A. Marrero, Z. Feng, J.A.S. Rinaudo, S. Srivastava, A. Evans, H.-W. Hann, H. Yang, T.M. Block, A. Mehta
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Z. Feng, S. Srivastava, A. Mehta
Study supervision: K. Devarajan, Z. Feng, J.A.S. Rinaudo, A. Mehta

References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer 2013; Available from: http://globocan.iarc.fr
2. Block TM, Mehta AS, Fimmel CJ, Jordan R. Molecular viral oncology of hepatocellular carcinoma. Oncogene 2003;22:5093–107.
3. Welzel TM, Graubard BI, Quraishi S, Zeuzem S, Davila JA, El-Serag HB, et al. Population-attributable fractions of risk factors for hepatocellular carcinoma in the United States. Am J Gastroenterol 2013;108:1314–21.
4. Chen DS, Sung JL, Sheu JC, Lai MY, How SW, Hsu HC, et al. Serum alpha-fetoprotein in the early stage of human hepatocellular carcinoma. Gastroenterology 1984;86:1404–9.
5. Marrero JA, El-Serag HB. Alpha-fetoprotein should be included in the hepatocellular carcinoma surveillance guidelines of the American Association for the Study of Liver Diseases. Hepatology (Baltimore, Md. 2011;53:1060–1; author reply 1–2.
6. Cheema AW, Hirschtritt T, Van Thiel DH. Markedly elevated alpha-fetoprotein levels without hepatocellular carcinoma. Hepatogastroenterology 2004;51:1676–8.

7. Di Bisceglie AM, Hoofnagle JH. Elevations in serum alpha-fetoprotein levels in patients with chronic hepatitis B. Cancer 1989;64:2117–20.
8. Marrero JA, Romano PR, Nikolaeva O, Steel L, Mehta A, Fimmel CJ, et al. GP73, a resident Golgi glycoprotein, is a novel serum marker for hepatocellular carcinoma. J Hepatol 2005;43:1007–12.
9. Hann HW, Wang M, Hafner J, Long RE, Kim SH, Ahn M, et al. Analysis of GP73 in patients with HCC as a function of anti-cancer treatment. Cancer Biomark 2010;7:269–73.
10. Comunale MA, Rodemich-Betesh L, Hafner J, Wang M, Norton P, Di Bisceglie AM, et al. Linkage specific fucosylation of alpha-1-antitrypsin in liver cirrhosis and cancer patients: implications for a biomarker of hepatocellular carcinoma. PLoS ONE 2010;5:e12419.
11. Wang M, Long RE, Comunale MA, Junaidi O, Marrero J, Di Bisceglie AM, et al. Novel fucosylated biomarkers for the early detection of hepatocellular carcinoma. Cancer Epidemiol Biomarkers Prev 2009;18:1914–21.
12. Comunale MA, Wang M, Hafner J, Krakover J, Rodemich L, Kopenhaver B, et al. Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. J Proteome Res 2009;8: 595–602.

13. Mehta AS, Long RE, Comunale MA, Wang M, Rodemich L, Krakover J, et al. Increased levels of galactose-deficient anti-Gal immunoglobulin G in the sera of hepatitis C virus-infected individuals with fibrosis and cirrhosis. J Virol 2008;82:1259–70.
14. Comunale MA, Lowman M, Long RE, Krakover J, Philip R, Seeholzer S, et al. Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. J Proteome Res 2006;5:308–15.
15. Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, Fimmel C, et al. Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. Proc Natl Acad Sci U S A 2005;102:779–84.
16. Comunale MA, Mattu TS, Lowman MA, Evans AA, London WT, Semmes OJ, et al. Comparative proteomic analysis of de-N-glycosylated serum from hepatitis B carriers reveals polypeptides that correlate with disease status. Proteomics 2004;4:826–38.
17. Wang M, Mehta A, Block TM, Marrero J, Di Bisceglie AM, Devarajan K. A comparison of statistical methods for the detection of hepatocellular carcinoma based on serum biomarkers and clinical variables. BMC Med Genom 2013;6:S9.
18. Wang M, Block TM, Marrero J, Di Bisceglie AM, Devarajan K, Mehta A. Improved biomarker performance for the detection of hepatocellular carcinoma by inclusion of clinical parameters. Proceedings (IEEE Int Conf Bioinformatics Biomed) 2012;2012.
19. Singal AG, Conjeevaram HS, Volk ML, Fu S, Fontana RJ, Askari F, et al. Effectiveness of hepatocellular carcinoma surveillance in patients with cirrhosis. Cancer Epidemiol Biomarkers Prev 2012;21:793–9.
20. Lok AS, Seeff LB, Morgan TR, Di Bisceglie AM, Sterling RK, Curto TM, et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease. Gastroenterology 2009;136:138–48.
21. Di Bisceglie AM, Sterling RK, Chung RT, Everhart JE, Dienstag JL, Bonkovsky HL, et al. Serum alpha-fetoprotein levels in patients with advanced hepatitis C: results from the HALT-C Trial. J Hepatol 2005; 43:434–41.
22. Hoefs JC, Shiffman ML, Goodman ZD, Kleiner DE, Dienstag JL, Stoddard AM. Rate of progression of hepatic fibrosis in patients with chronic hepatitis C: results from the HALT-C Trial. Gastroenterology 2011;141: 900–8 e1-2.
23. Haque M, Yoshida EM. Hepatitis C antiviral long-term treatment against cirrhosis (HALT-C) trial. Ann Hepatol 2009;8:78–9.
24. Goodman ZD, Stoddard AM, Bonkovsky HL, Fontana RJ, Ghany MG, Morgan TR, et al. Fibrosis progression in chronic hepatitis C: morphometric image analysis in the HALT-C trial. Hepatology (Baltimore, Md. 2009;50: 1738–49.
25. Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, et al. Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. Gastroenterology 2009;137:110–8.
26. Hann H-W, Wan S, Myers RE, Hann RS, Xing J, Chen B, et al. Comprehensive analysis of common serum liver enzymes as prospective predictors of hepatocellular carcinoma in HBV patients. PloS One 2012;7:e47687.
27. Gopal P, Yopp AC, Waljee AK, Chiang J, Nehra M, Kandunoori P, et al. Factors that affect accuracy of alpha-fetoprotein test in detection of hepatocellular carcinoma in patients with cirrhosis. Clin Gastroenterol Hepatol 2014;12:870–7.
28. Steyerberg E. Clinical prediction models: Springer; 2009.
29. Cho JY, Paik YH, Sohn W, Cho HC, Gwak GY, Choi MS, et al. Patients with chronic hepatitis B treated with oral antiviral therapy retain a higher risk for HCC compared with patients with inactive stage disease. Gut 2014;63: 1943–50.
30. El-Serag HB, Kanwal F, Davila JA, Kramer J, Richardson P. A new laboratory-based algorithm to predict development of hepatocellular carcinoma in patients with hepatitis C and cirrhosis. Gastroenterology 2014;146:1249–55e1.