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

Mengjun Wang1, Karthik Devarajan2, Amit G. Singal3, Jorge A. Marrero3, Jianliang Dai4, Ziding Feng4, Jo Ann S. Rinaudo5, Sudhir Srivastava5, Alison Evans6, Hie-Won Hann7, Yinzhi Lai7, Hushan Yang7, Timothy M. Block8, and Anand Mehta1

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.

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. 

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

Corresponding Author: Anand Mehta, Drexel University College of Medicine, Room 16307 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
©2015 American Association for Cancer Research.
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 UTSW 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, R², 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
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 Doylest-
town 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 algo-
rithm was sent as an equation (as shown in Fig. 2A) to our
collaborators. All selection of patients, application of the algo-
rithm, 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 AUC 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 AUC 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 perfor-
manace 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
An Algorithm to Detect HCC

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.

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 that 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 0.876, 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 10,000 ng/mL. As expected, at
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 revealed 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.
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. It is possible that this imparted some potential selection bias in the external validation. That is, only patients with the required clinical factors were used in the external validation. 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. The true flexibility will only be determined when the clinical community uses the model.

The 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. The true flexibility will only be determined when the clinical community uses the model.
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 Arbusa BioPharma, has ownership interest (including patents) in Glycotech, and is a consultant/advisory board member for Glycotech. 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

Grant Support

This work was supported by grants R01 CA120206 (A. Mehta), U01 CA168856 (A. Mehta), and P30 CA06927 (K. Devarajan) from the NCI, the Hepatitis B Foundation, and an appropriation from The Commonwealth of Pennsylvania (T.M. Block).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 6, 2015; revised November 24, 2015; accepted December 11, 2015; published OnlineFirst December 28, 2015.

References

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;53:1066–1. author reply 1–2


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

Mengjun Wang, Karthik Devarajan, Amit G. Singal, et al.