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*  

¹Drexel University College of Medicine, 245 North 15th Street, Philadelphia, PA 19102.  
²Fox Chase Cancer Center, Philadelphia, PA, 18901.  
³Division of Digestive and Liver Diseases, University of Texas Southwestern, 5959 Harry Hines Blvd POB I Suite 420B, Dallas TX 75201 – 8887.  
⁴Department of Biostatistics, Division of Quantitative Sciences, The University of Texas, MD Anderson Cancer Center, Houston, TX.  
⁵Cancer Biomarkers Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.  
⁶Drexel University School of Public Health, Nesbitt Hall, 3215 Market St., Philadelphia, PA, 19104  
⁷Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA, 19102.  
⁸The Baruch Blumberg Institute, 3805 Old Easton Road, Doylestown, PA, 18902.  

*To whom Correspondence should be addressed: Anand Mehta, Drexel University College of Medicine, Room 18307 New College Building, 245 N. 15th Street, Philadelphia, PA 19102.  
Tel: (215-762-7586).  
Email: anand.mehta@drexelmed.edu.  

Common abbreviations:  
HCC, hepatocellular carcinoma;  
AFP alpha-fetoprotein  
AUROC, Area Under the Receiver Operating Characteristic Curve  

Running title: An algorithm to detect HCC  

Conflict of Interest: The authors have nothing to declare and have no conflict of interest regarding the work described.  

Financial Support: This work was supported by grants R01 CA120206 (A. Mehta), U01 CA168856 (A. Mehta) and P30 CA 06927 (K. Devarajan) from the National Cancer Institute (NCI), the Hepatitis B Foundation, and an appropriation from The Commonwealth of Pennsylvania (T.M Block).  

Word count: 3562  
Abstract word count: 249  
Number of tables: 1  
Number of figures: 4
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 alpha fetoprotein (AFP) is routinely used, an algorithm that incorporated AFP values along with 4 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 3 independent sites the Doylestown algorithm improved detection of HCC as compared to AFP alone by 4-20%. In addition, at a fixed specificity of 95%, the Doylestown algorithm improved the detection of HCC as compared to AFP alone by 2-20%. 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, alpha-fetoprotein (AFP)(4, 5). However, there has been extensive debate about the utility of AFP given its suboptimal sensitivity and specificity(6, 7)(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 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 University of Michigan 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 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 (magnetic resonance imaging [MRI], or computed tomography [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 University of Michigan Hospital Clinical Diagnostic Laboratory. The University of Michigan’s Institutional Review Board approved the study protocol. Patient information is provided in Supplemental Table 1.

**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 was used.
This was generally 0-3 months prior to HCC detection as described in (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 Supplemental Table 2.

**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 National Cancer Institute (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 7 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 the 2009 publication describing this set in great detail(25). Patient information is provided in Supplemental Table 3. 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 co-infection with HCV). Thus both cases and controls were treated according to AASLD guidelines for their HBV and thus, were DNA negative. Patients without complete records of the analyzed variables (i.e. age, gender, AFP, ALT and ALP) were excluded. Serum levels of AFP, ALT, and ALK were determined using commercially available kits at the Thomas Jefferson Hospital or other Clinical Diagnostic Laboratories. Patient information is provided in Supplemental Table 4.

University of Texas Southwestern (UTSW) Cohort: The third validation cohort used data from UTSW and 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 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 UT Southwestern Medical Center. Patient information is provided in Supplemental Table 5.
Statistical Methods

Datasets: As stated, a dataset utilizing samples from the University of Michigan (Supplemental Table 1) and HALT-C (Supplemental Table 2) 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, showed in supplementary table 6. More information on feature selection and analysis is provided in the supplemental 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’s 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 7.

To avoid over-fitting, we applied leave-one-out cross-validation, bootstrap-validation and 3-fold cross-validation to validate the 10 candidate models (Supplemental Tables 8-10). Based upon the performance of the cross validation and the properties of the calibration, the model with logAFP, Age, Gender, ALK, ALT was selected for further development. We refer to this as the Doylestown algorithm (Supplementary Table 11). More information on the models and methods...
used for algorithm development are provided in the supplemental 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 12).

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

**Results**

**Model Development and Performance in Training set:**
In our previous efforts to develop non invasive 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 in shown in figure 1. Supplementary table 6 shows the ten clinical factors analyzed and those five were found to be associated with HCC, which included age, gender, alkaline phosphatase (ALK), alanine aminotransferase (ALT) and log transformed AFP values. The logistic regression equation is presented in figure 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 (Figure 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 figure 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 (Figure 1) consisted of a large multi-center case control study collected by the Early Detection Research Network (EDRN) of the National Cancer Institute (NCI). This case controlled study consists of 870 patients, 438 patients with liver cirrhosis and 432 patients with HCC (Supplementary Table 3). 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 (Figure 3A). This increase was statistically significant ($P<0.0001$). In addition, if only patients with early cancers ($n=225$) were
examined, as figure 3B shows, a similar increase in performance was seen (0.7856 to 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 controlled 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 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 anti-viral therapy significantly reduces the incidence of liver cancer in these patients (~50% reduction), the risk remains very high, almost 20-30 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 Thomas Jefferson University and consisted of those with HBV associated liver disease and treated for their infection following AASLD guidelines and were DNA negative at time of study. This set comprised of 699 patients, 113 that had HBV associated early HCC and 586 of which had chronic HBV infections (Supplementary Table 4). 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 figure 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 University of Texas Southwestern (UTSW) and consisted of 1229 patients—804 with a background of liver cirrhosis and 425 with HCC (Supplementary Table 5). 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 figure 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 5), 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 10-100 ng/mL, where the AUROC was increased from 0.579 for AFP alone to 0.700 for the Doylestown algorithm. As figure 4B shows, when patients were broken down into specific groups based on AFP level, the Doylestown algorithm increased the AUROC in almost all groups, from those with AFP between 10-100 ng/mL all the way to AFP levels between 10-
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 sub-group 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 10-100 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 re-evaluated the performance of the Doylestown algorithm in the TJU patients only in those patients with AFP in the zone of 10-100 ng/mL (all of these patients had early HCC). Consistent with the results shown in Figure 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, 40 cases; Figure 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 10-100 ng/mL (n=106 controls, 202 cases; Figure 4E) when patients with all stages of HCC were examined. When only patients with early stage cancer (n=109) who had AFP between 10-100 ng/mL were examined the AUROC also increased from 0.641 to 0.773 (see Figure 4F).

In the discovery sets 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 10-100 ng/mL was 0.6636 and this was increased to 0.9110 with the Doylestown algorithm. And in the in the HALT-C data set, the AUROC of AFP in patients with AFP between 10-100 ng/mL was 0.6583 and this was increased by the Doylestown algorithm increased this to 0.7077.
Discussion:

In this paper, 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 4-12% 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 <20 ng/mL was misclassified by the Doylestown algorithm (ie. 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 re-classified 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 et al. has 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 non-viral liver disease and thus expanded upon the work presented by El-Serag, et al. which only examined HCV patients. Our analysis also included both internal and external validation from multiple sources, which was different than 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 test such as ALK, ALT and AFP can vary when performed in one laboratory to another. This inter-assay 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 variable of up to 15% can occur without miss classification, 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 paper also has 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 algorithm would be used in the management of patients at risk of developing HCC.
Reference:


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

<table>
<thead>
<tr>
<th>Sample source</th>
<th>Total number of samples</th>
<th>HCC samples</th>
<th>Control samples</th>
<th>AFP Sensitivity at 95% Specificity</th>
<th>Doylestown algorithm Sensitivity at 95% Specificity</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>UM&lt;sup&gt;1&lt;/sup&gt;</td>
<td>209</td>
<td>116</td>
<td>93</td>
<td>55%</td>
<td>75%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>UM early stage HCC only&lt;sup&gt;2&lt;/sup&gt;</td>
<td>153</td>
<td>60</td>
<td>93</td>
<td>48%</td>
<td>88%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HALT-C&lt;sup&gt;3&lt;/sup&gt;</td>
<td>151</td>
<td>49</td>
<td>102</td>
<td>43%</td>
<td>55%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HALT-C-Early stage HCC only&lt;sup&gt;4&lt;/sup&gt;</td>
<td>141</td>
<td>39</td>
<td>102</td>
<td>36%</td>
<td>41%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>EDRN&lt;sup&gt;5&lt;/sup&gt;</td>
<td>870</td>
<td>432</td>
<td>438</td>
<td>42%</td>
<td>53%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>EDRN – Early stage only&lt;sup&gt;6&lt;/sup&gt;</td>
<td>656</td>
<td>224</td>
<td>432</td>
<td>31%</td>
<td>43%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TJU&lt;sup&gt;7&lt;/sup&gt;</td>
<td>699</td>
<td>113</td>
<td>586</td>
<td>38%</td>
<td>58%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>UTSW&lt;sup&gt;8&lt;/sup&gt;</td>
<td>1229</td>
<td>425</td>
<td>804</td>
<td>61%</td>
<td>63%</td>
<td>0.9378</td>
</tr>
<tr>
<td>UTSW – Early stage Only&lt;sup&gt;9&lt;/sup&gt;</td>
<td>943</td>
<td>139</td>
<td>804</td>
<td>40%</td>
<td>35%</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>UTSW only 10-100 ng/mL of AFP&lt;sup&gt;10&lt;/sup&gt;</td>
<td>229</td>
<td>98</td>
<td>121</td>
<td>10%</td>
<td>25%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>EDRN only 10-100 ng/mL of AFP&lt;sup&gt;10&lt;/sup&gt;</td>
<td>308</td>
<td>202</td>
<td>106</td>
<td>24%</td>
<td>47%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TJU only 10-100 ng/mL of AFP&lt;sup&gt;10&lt;/sup&gt;</td>
<td>128</td>
<td>40</td>
<td>88</td>
<td>12%</td>
<td>27%</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

1) The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the UM training set; 2) sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the UM training set in only those with early stage HCC; 3) sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the HALT-C training set; 4) sensitivity of either AFP or the Doylestown algorithm at 95% specificity in only those with early stage HCC; 5) The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the EDRN validation set. 6) The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in the EDRN validation set in only those with early stage HCC; 7) The sensitivity of either AFP or the Doylestown algorithm at 90% specificity in the Thomas Jefferson University (TJU) set. 8) The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in whole University of Texas Set. 9) The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in those patients with early stage HCC; 10) The sensitivity of either AFP or the Doylestown algorithm at 95% specificity in those patients with AFP between 10-100 ng/mL in each of the respective validation cohorts.
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 non viral liver disease.

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 is the line of 95% specificity.

Figure 3. Validation of the Doylestown algorithm in the NCI EDRN sample set and in an HBV infected sample set from Thomas Jefferson University. 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 Thomas Jefferson University validation group. Dotted line is the line of 95% specificity.

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, Y-axis is the AUC for either AFP or the Doylestown algorithm in the specified patients. For 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≤1000; group 11 are patients with 10<AFP<=10,000; group 12 are patients with AFP≥100,000; group 13 are all patients. In all case AFP values are ng/mL. C) AUROC of the Doylestown algorithm and AFP in the UTSW set only in patients with AFP between 10-100 ng/mL. D) AUROC of the Doylestown algorithm and AFP in the TJU set only in patients with AFP between 10-100 ng/mL. E) AUROC of the Doylestown algorithm and AFP in the EDRN set only in patients with AFP between 10-100 ng/mL. F) AUROC of the Doylestown algorithm and AFP in the EDRN set only in patients with early stage HCC and AFP between 10-100 ng/mL. Dotted line is the line of 95% specificity.
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 National Cancer Institute (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
\[ p = \frac{1}{1 + \exp \left( -[-10.307 + (0.097 \times \text{age}) + (1.645 \times \text{Gender}) + (2.315 \times \log \text{AFP}) + (0.011 \times \text{ALK}) + (-0.008 \times \text{ALT})] \right) } \]
Figure E: Receiver operating characteristic curve for the prediction of outcome A with AUC = 0.6439 and for the prediction of outcome B with AUC = 0.7591.

Figure F: Receiver operating characteristic curve for the prediction of outcome C with AUC = 0.6411 and for the prediction of outcome D with AUC = 0.7730.
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.


Updated version  Access the most recent version of this article at: doi:10.1158/1940-6207.CAPR-15-0186

Supplementary Material  Access the most recent supplemental material at: http://cancerpreventionresearch.aacrjournals.org/content/suppl/2015/12/22/1940-6207.CAPR-15-0186.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.