Improved Detection of Prostate Cancer Using Classification and Regression Tree Analysis

Mark Garzotto, Tomasz M. Beer, R. Guy Hudson, Laura Peters, Yi-Ching Hsieh, Eduardo Barrera, Thomas Klein, and Motomi Mori

ABSTRACT

Purpose
To build a decision tree for patients suspected of having prostate cancer using classification and regression tree (CART) analysis.

Patients and Methods
Data were uniformly collected on 1,433 referred men with a serum prostate-specific antigen (PSA) levels of ≤ 10 ng/mL who underwent a prostate biopsy. Factors analyzed included demographic, laboratory, and ultrasound data (ie, hypoechoic lesions and PSA density [PSAD]). Twenty percent of the data was randomly selected and reserved for study validation. CART analysis was performed in two steps, initially using PSA and digital rectal examination (DRE) alone and subsequently using the remaining variables.

Results
CART analysis selected a PSA cutoff of more than 1.55 ng/mL for further work-up, regardless of DRE findings. CART then selected the following subgroups at risk for a positive biopsy: (1) PSAD more than 0.165 ng/mL/cc; (2) PSAD > 0.165 ng/mL/cc and a hypoechoic lesion; (3) PSAD ≤ 0.165 ng/mL/cc, no hypoechoic lesions, age older than 55.5 years, and prostate volume ≤ 44.0 cc; and (4) PSAD ≤ 0.165 ng/mL/cc, no hypoechoic lesions, age older than 55.5 years, and 50.25 cc less than prostate volume ≤ 80.8 cc. In the validation data set, specificity and sensitivity were 31.3% and 96.6%, respectively. Cancers that were missed by the CART were Gleason score 6 or less in 93.4% of cases. Receiver operator characteristic curve analysis showed that CART and logistic regression models had similar accuracy (area under the curve = 0.74 v 0.72, respectively).

Conclusion
Application of CART analysis to the prostate biopsy decision results in a significant reduction in unnecessary biopsies while retaining a high degree of sensitivity when compared with the standard of performing a biopsy of all patients with an abnormal PSA or DRE.

INTRODUCTION

Prostate cancer is the most common cancer in men and the second leading cause of male cancer deaths in the United States. In the absence of effective treatment options for advanced prostate cancer, intensive efforts to detect low-stage, curable cancers may help to improve prostate cancer–specific survival.1 Recent data suggest that prostate-specific antigen (PSA) screening reduces the clinical stage at presentation and prostate cancer mortality; however, validation of these results awaits the completion of ongoing randomized screening trials.3

Optimal strategies for selecting appropriate patients for prostate biopsy have yet to be defined. Screening for prostate cancer detects the majority of prostate cancer patients; however, their effectiveness is severely hampered by a lack of specificity. Among men who are screened, 22% to 26%
of men will have an abnormal serum PSA and/or digital rectal examination (DRE) and are likely to receive a recommendation for a prostate biopsy. Furthermore, 20% to 25% of these screen-positive men will be found to have cancer detected on biopsy. Thus, the majority of patients who are biopsied because of screening abnormalities undergo biopsy unnecessarily.

To reduce the rate of unnecessary biopsies, efforts have focused on characterizing patient groups with an abnormal PSA and/or DRE who have a low likelihood of a positive biopsy. Multiple factors that have been associated with the detection of prostate cancer include age, race, family history, hypoechoic lesions on transrectal ultrasound (TRUS), PSA density (PSAD), PSA velocity, transition-zone PSAD, and percentage of free PSA. Prior work has demonstrated that the number of positive biopsies can be reduced by a modest degree with the use of percentage of free PSA, nomograms, predictive indices, and artificial neural networks. Instruments that accurately predict the presence of cancer have the potential to reduce the number of unnecessary biopsies along with their accompanying pain, morbidity, and cost.

Classification and regression tree (CART) analysis belongs to a family of nonparametric regression methods and is based on the recursive partitioning method. The CART builds a decision tree structure and classifies subjects into high- and low-risk groups. It can be used simply to explore the data, identify possible high-risk subgroups, and uncover interactions or effect modifications among prognostic factors. Unlike the commonly used logistic regression method, CART analysis does not assume a multiplicative risk model or a specific parametric probability model, does not require a specification of the risk function (eg, a linear or quadratic effect of age), and is not affected by outlying observations. Most importantly, the results of CART analysis are presented as a decision tree, which is intuitive and easier to understand than the results of many other statistical methods. In this study, CART analysis was used to create a decision tree that can assist clinicians in making a prostate biopsy decision.

**Statistical Methods**

The data were divided randomly into a model building set (80%) and validation set (20%). CART uses a binary recursive partitioning method that produces a decision tree that identifies subgroups of patients at higher likelihood of testing positive for a disease state. Unequal misclassification costs were specified so that there was a three times higher cost associated with misclassifying a cancer case as a noncancer case. CART was performed on the model building data set. To simulate the clinical decision process, a two-step approach was carried out. First, a decision tree was created based on the PSA and DRE results, and then the subgroup identified as having a high risk of prostate cancer was analyzed using the remaining variables in the data set. The sensitivity and specificity of the resulting tree analysis were evaluated using the validation set. The results of CART for the prediction of prostate cancer were compared with the results derived from a model based on logistic regression using the area under the curve (AUC) from the receiver operator characteristic curves. To visually confirm the cutoff points selected by the CART, generalized additive models were used to evaluate the functional relationship between the cancer risk and the continuous variables (ie, age, PSA, PSAD, and prostate volume).

This analysis was designed to identify predictors of biopsy outcome and not of true disease status. Current prostate diagnosis techniques do not allow for the complete elimination of false-negative biopsies; therefore, true disease status (presence or absence of prostate cancer) cannot be completely determined. Because the presence of false-negative biopsies introduces an asymmetric bias, the results should not be extrapolated to definitively predict true disease status.

**RESULTS**

### Patient Characteristics

The median age of the study group was 65.1 years (Table 1). Most of the patients were classified as white...
A family history of prostate cancer was reported in 17.6% of patients. The median PSA level in this group was 5.0 ng/mL (mean PSA, 4.8 ng/mL). The DRE was normal in 48.8% of patients, asymmetric in 5.5% of patients, suspicious in 41.7% of patients, and cancer-likely in 4.0% of patients.

**Ultrasound and Biopsy Data**

The median prostate volume was 34.2 cc (range, 4.9 to 205 cc; Table 2). The median PSAD was 0.12 ng/mL/cc. Hypoechoic lesions were demonstrated in 44.9% of patients. Cancer was detected 24.4% of patients. The majority of tumors (51.8%) was determined to be Gleason score 6.

**CART Analysis**

The initial CART procedure was carried out on the model building set (n = 1,173) using DRE and PSA data only to determine value of these primary factors in the prostate decision process. The initial CART selected a PSA cutoff level of more than 1.55 ng/mL alone for the identification of patients at risk for a positive prostate biopsy (Fig 1). Using this cutoff, 96.6% (sensitivity) of cancer patients (281 of 291 patients) were identified for further analysis, whereas 193 (21.9%) of 882 subjects without cancer were correctly identified (specificity). Using this PSA cutoff alone, the percent overall reduction in prostate biopsies was 17.3% in the model building set (203 of 1,173 procedures).

In the group with PSA more than 1.55 ng/mL (n = 970), prostate cancer was detected in 29.0% of patients (281 of 970 patients). The second CART analysis was carried out using all remaining variables, and the resulting decision tree was merged with the initial CART-derived tree. This supplementary decision algorithm identified the following four groups for a biopsy of the prostate: (1) PSAD more than 0.165 ng/mL/cc; (2) PSAD \(\leq 0.165 \text{ ng/mL/cc and TRUS hypoechoic lesion} \); (3) PSAD \(\leq 0.165 \text{ ng/mL/cc, absence of a hypoechoic TRUS lesion, age older than 55.5 years, and prostate volume \(\leq 44.0 \text{ mL; and (4) PSAD} \leq 0.165 \text{ ng/mL/cc, absence of a hypoechoic TRUS lesion, age older than 55.5 years, and 50.25 cc less than prostate volume \(\leq 80.8 \text{ cc.} \)

The incidences of cancer detection in these groups were 48.8%, 26.4%, 21.6%, and 16.5%, respectively. For subjects with a PSA more than 1.55 ng/mL, CART detected 98.9% of the remaining prostate cancers (278 of 281 cancers) and correctly identified patients without cancer in 20.5% of the remaining subjects (141 of 698 subjects).

The complete model was found to have an overall sensitivity of 95.5% (278 of 291 cancer patients) and a specificity of 37.9% (334 of 882 noncancer patients) in the model building set. The positive predictive value was 33.7%, and the test carried a negative predictive value of 96.3%. CART analysis was then carried out using the randomly selected validation set (n = 260). In this study, the sensitivity was 96.6% (57 of 59 patients), and the specificity was 31.3% (63 of 201 patients). Assuming the prostate cancer prevalence of 24% in this population, the positive and negative predictive values from the validation set were 29.2% and 96.9%, respectively.

The results generated by the CART were then compared with a logistic regression model created using the same factors. These analyses were performed using the validation data set. The CART decision tree had a receiver operator characteristic curve AUC of 0.74. When the same variables were entered into the logistic regression model, the AUC was 0.72. Thus, both CART- and logistic regression–based models are comparable in their effectiveness in discriminating biopsy-positive cases from biopsy-negative cases.

**Table 1. Patient Characteristics (N = 1,433)**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>65.1</td>
</tr>
<tr>
<td>Median</td>
<td>66.0</td>
</tr>
<tr>
<td>Range</td>
<td>41-85</td>
</tr>
<tr>
<td>Race, %</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>93.3</td>
</tr>
<tr>
<td>African American</td>
<td>4.0</td>
</tr>
<tr>
<td>Other</td>
<td>2.3</td>
</tr>
<tr>
<td>Family history, %</td>
<td>17.6</td>
</tr>
<tr>
<td>Vasectomy, %</td>
<td>33.8</td>
</tr>
<tr>
<td>PSA, ng/mL</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.8</td>
</tr>
<tr>
<td>Median</td>
<td>5.0</td>
</tr>
<tr>
<td>DRE, %</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>48.8</td>
</tr>
<tr>
<td>Asymmetric</td>
<td>5.5</td>
</tr>
<tr>
<td>Suspicious</td>
<td>41.7</td>
</tr>
<tr>
<td>Cancer-likely</td>
<td>4.0</td>
</tr>
</tbody>
</table>

**Table 2. Ultrasound and Biopsy Findings**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate volume, cc</td>
<td>40.0</td>
</tr>
<tr>
<td>Mean</td>
<td>34.2</td>
</tr>
<tr>
<td>Median</td>
<td>4.9-205</td>
</tr>
<tr>
<td>Hypoechoic area, %</td>
<td>44.9</td>
</tr>
<tr>
<td>PSAD, ng/mL/cc</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.14</td>
</tr>
<tr>
<td>Median</td>
<td>0.12</td>
</tr>
<tr>
<td>Range</td>
<td>0.01-0.8</td>
</tr>
<tr>
<td>Cancer present, %</td>
<td>24.4</td>
</tr>
<tr>
<td>Gleason score, %</td>
<td>2-4</td>
</tr>
<tr>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>8-10</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: PSA, prostate-specific antigen; DRE, digital rectal examination.
Generalized Additives Models

Generalized additive models (Fig 2A through 2D) were used to visually assess functional relationships between the continuous covariates (ie, age, PSA, PSAD, and prostate volume) and the risk of a positive biopsy. This analysis was conducted using both logarithmic transformed and untransformed data. Log(relative risk) can be converted to a relative risk by taking antilog. For example, a log(relative risk) of 0 implies the relative risk of 1 (no impact on the probability of having a prostate cancer), whereas a log(relative risk) of 1 implies the relative risk of 2.71 (ie, 2.71-fold increase in the probability of having prostate cancer).

There was a monotone increasing relationship between the risk of prostate cancer and both patient age and PSAD. Serum PSA was found to have a biphasic distribution, with an initial linear increase in the low PSA range that was followed by a plateau phase in the intermediate PSA range from 4 to 10 ng/mL. Prostate volume demonstrated a sinusoidal pattern relative to prostate cancer risk.

Biopsy Gleason Scores

The biopsy Gleason scores were compared in the group who had CART-detected cancers compared with the group who had cancers that were missed in the analysis. Collectively, between the model building (n = 13) and validation datasets (n = 2), the CART model missed 15 (4.3%) of 350 cancer cases. These missed cases were generally of lower grade than the cancers that were detected by the CART analysis (Fig 3). The missed cancers were classified as Gleason score 6.
or lower in 93.4% of patients and Gleason score 5 or below in 46.7% of patients.

DISCUSSION

The optimal strategy to screen patients for prostate cancer has yet to be defined. Screening for prostate cancer with the combination of PSA and DRE has been thought to be a sensitive means of detecting prostate cancer; however, recent data suggest these tests have diagnostic limitations. A study of men who had a normal DRE and a serum PSA level less than 3.0 ng/mL demonstrated that standard prostate screening practice is able to detect 85% of prostate cancers. The biologic significance of these tumors is not known; although these subclinical cancers may belong to a pool that are destined for a more indolent course.

Using PSA and DRE, up to 25% of men who present for prostate screening will have abnormalities that could lead to a biopsy of the prostate. Efforts have been made to reduce the number of unnecessary biopsy procedures through the use of PSA derivatives, PSA subspecies analysis, and multivariable models based on either logistic regression or artificial neural networks. Many of these models have not gained wide acceptance because of a lack of sensitivity, specificity, or both. Recently, we reported on the development of a prebiopsy nomogram, which incorporates the results of patient age and DRE and TRUS findings to determine the relative likelihood of a positive biopsy for patients with a serum PSA of ≤ 10 ng/mL. This nomogram has a sensitivity of 92% and a specificity of 24% (AUC = 0.73). Furthermore, it was able to reduce the total number of biopsy procedures by 20%. In the current study, we used CART analysis to develop a prostate biopsy decision algorithm in patients referred for an abnormal PSA, DRE, or both. Using these covariates, CART analysis selected PSA more than 1.55 ng/mL alone as an indication for further work-up. With the application of this cutoff, 96.6% of all cancers were detected, while reducing the necessity of both ultrasound procedures and biopsies by 21.9%. Thus, in contrast to our previous nomogram model, the CART was able to find a low-risk group that could not only avoid the biopsy procedure, but also forgo the ultrasound procedure. Among patients with a PSA more than 1.55 ng/mL, patients were selected for biopsy based on PSAD, presence of a hypoechoic lesion, age, and prostate volume. Overall, the CART model was able to detect 96.6% of all cancers and reduce the number of unnecessary procedures by 31.3%, with an overall reduction in biopsy procedures of 25.0%.

Fig 2. General additive models demonstrate the relationship between continuous variables and the risk of prostate cancer. The resulting figures show the predicted log(relative risk) in the y-axis and the continuous covariate in the x-axis.
Thus, the CART model significantly reduced the number of unnecessary biopsies, while retaining a high degree of sensitivity, when compared with standard screening practice. Although the sample size of the validation set was adequate for the evaluation of overall performance, its size precludes the rigorous evaluation of model performance at each individual terminal node. However, we compared the accuracy for exploratory purposes and found the overall probabilities in each terminal node to be comparable to those obtained in the model building set. An exception was noted for the high-risk group 3 (PSA/H11022 \(1.55 \text{ ng/mL}\), PSAD/H11349 \(0.165 \text{ ng/mL/cc}\), TRUS nonhypoechoic, age \(55 \text{ years}\), prostate volume \(44 \text{ mL}\)), for which the model building set showed a biopsy-positive rate of 21.6% (n = 125) compared with 6.1% (n = 33) in the validation set. It seems that this difference accounted for the decrease in specificity; however, importantly, model sensitivity was preserved. This may be an effect caused by the small sample size at this node or by true inconsistencies between the validation and training datasets.

In this study, a serum PSA level of greater than 1.55 ng/mL alone was identified as an indication for further testing in a population referred for prostate evaluation. The DRE lost predictive value when the PSA test was coanalyzed as a continuous variable. It is possible that, at this threshold, tumors may be too small for prostate cancer detection by DRE as previously suggested by Vis et al.\(^{19}\) Schroder et al\(^{20}\) found that the positive predictive value of the DRE was negatively associated with increased PSA levels. The positive predictive value for the DRE was 45% when the PSA level was between 4.0 and 10.0 ng/mL; however, it was reduced to 10% when the PSA level was between 1.0 and 1.9 ng/mL and was reduced to 4% when the PSA level was between 0.0 and 0.9 ng/mL. These investigators calculated that, for patients with DRE abnormalities and PSA values between 1.0 and 1.9 ng/mL, 38 biopsy procedures would be necessary to discover one cancer. For patients with PSA values from 0.0 to 0.9 ng/mL, 46 biopsies were required to detect a single cancer, which is a rate they determined to be “unacceptably high.”\(^{4}\) Others have similarly reported the DRE to have a low positive predictive yield in patients with a low PSA.\(^{21,22}\)

Recently, it has been proposed that the normal PSA cutoff level be decreased to 2.5 or 3.0 ng/mL.\(^{23,24}\) A higher rate of organ-confined disease and the potential for curability have been cited as justification to changing the current screening recommendations.\(^{20,25,26}\) Additional evidence lending support to a lowering of the PSA cutoff level comes from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.\(^{27}\) In this trial, men with a PSA level of less than 2.0 ng/mL had a 97.4% chance of maintaining a PSA \(\leq 4.0 \text{ ng/mL}\) for 3 years, whereas the risk for men with an initial PSA level between 2.0 and 3.0 ng/mL was 76.4%. The rate of PSA conversions in the initially low PSA group was similar to the incidence of missed cancers in patients with low PSA values in our study. In the current study, the relationship of PSA to prostate cancer biopsy detection was examined in a referred population. The risk of a positive biopsy was found to be 4.9% (10 of 203 subjects) for men with a PSA \(\leq 1.55 \text{ ng/mL}\), whereas the risk of cancer was 29.0% above this value. These data support using a lower PSA cutoff level for the detection of prostate cancer; however, further study is needed to determine the net benefit of detecting cancers at this lower PSA level. The presence of a hypoechoic lesion on TRUS is well known to be suggestive of prostate cancer.\(^{28,29}\) Prostate cancer tissue contains a reduced number of acoustic interfaces primarily because of its cellular density and, thus, is less echogenic (ie, hypoechoic) relative to normal prostate tissue. Similar to our previously published model, the presence of a hypoechoic lesion by TRUS was found to aid in the prediction of cancer using the CART model. In patients with a PSA level more than 1.55 ng/mL and a PSAD more than 0.165 ng/mL/cc, the incidence of prostate cancer in patients with a hypoechoic lesion was 26.4% compared with 12.8% in patients without lesions (relative risk = 2.1).

In the intermediate PSA range, the most important disease to discriminate from prostate cancer is benign prostatic hyperplasia.\(^{30}\) Many studies have shown a reduction in prostate cancer risk with increasing prostate size. In the current study, an overall reduction in prostate cancer risk...
was observed with increase in gland size. However, a trend towards increased risk was observed on the low and high end of the volume scale (Fig 2D). Because there are relatively small numbers of patients in these ranges, it is unclear what their significance is. Additional studies with greater numbers of patients in these categories would be required to make assumptions about prostate cancer risk in these areas. The CART model chose select patients with smaller prostate glands (volume ≤ 44.0 mL) to be at increased risk of prostate cancer. Interestingly, it also chose a group of patients for biopsy who would normally be considered low-risk patients (volume, 50.3 to 80.8 mL). The identification of this subgroup combined with other clinical features illustrates some of the advantageous features of CART analysis, which, unlike other methods, does not assume a multiplicative risk model or a specific parametric probability model and, thus, allows for the unveiling of unique interactions between covariates.

There were limitations to the current study that warrant mention. Because percentage of free PSA was not available at the onset of this study, we were unable to assess its utility in the current model. The CART model did include PSAD, which has been shown to have equivalent or improved predictive capacity compared with percentage of free PSA. Future studies that incorporate percentage of free PSA or its subtypes into multivariable models will likely result in further improvements to such models. Similar to other reports, this study included a low proportion of black men; thus, it is not known how well the CART model would perform in this population or other racial or ethnic groups. Our model was designed to predict the presence of cancer in men on initial prostate biopsy only and not whether cancer would be detected on subsequent biopsies. Therefore, similar to those patients who have biopsy-negative results, patients who forego a biopsy based on the results of the CART should continue to be monitored. Recent studies have suggested that extended biopsy schemes taking more cores improves the rate of cancer detection. Then again, a randomized controlled trial comparing a six-cores with 12-cores biopsy procedure showed no difference in cancer detection rates. Although it is possible that extended biopsy schemes may find cancer more frequently, it may also contribute disproportionately to the overdiagnosis of prostate cancer. Nonetheless, similar to all studies of prostate cancer diagnosis, this analysis is limited by the bias introduced by false-negative biopsies. This limitation of the biopsy procedure precludes the definitive determination of the true disease status. Therefore, the results of this analysis should be used to aid in predicting the outcome of a biopsy procedure, which is the primary goal of the analysis, and should not be used to predict true disease status.

In summary, CART analysis chose a PSA cutoff level of ≤ 1.55 ng/mL for the identification of patients at minimal risk for a positive biopsy regardless of DRE findings. The model correctly identified 96.6% of patients with prostate cancer and was able to reduce the number of unnecessary biopsies by 31.3%.

Authors’ Disclosures of Potential Conflicts of Interest
The authors indicated no potential conflicts of interest.

REFERENCES

11. Carlson GD, Calvanese CB, Partin AW: An algorithm combining age, total prostate-specific antigen (PSA), and percent free PSA to predict prostate cancer: Results on 4298 cases. Urology 52:455-461, 1998
18. Carter HB, Pearson JD: Prostate-specific antigen velocity and repeated measures of