Integrated predictive model for prostatic cancer using clinical, laboratory and ultrasound data

**Modelo preditivo integrado para a presença de câncer de próstata utilizando dados clínicos, laboratoriais e ultrassonográficos**

GUSTAVO DAVID LUUDWIG, ACBC-SC¹; HENRIQUE PERES ROCHA¹; LUCIO JOSÉ BOTELHO²; MAIARA BRUSCO FREITAS².

**ABSTRACT**

**Objective:** to develop a predictive model to estimate the probability of prostate cancer prior to biopsy. **Methods:** from September 2009 to January 2014, 445 men underwent prostate biopsy in a radiology service. We excluded from the study patients with diseases that could compromise the data analysis, who had undergone prostate resection or used 5-alpha-reductase inhibitors. Thus, we selected 412 patients. Variables included in the model were age, prostate specific antigen (PSA), digital rectal examination, prostate volume and abnormal sonographic findings. We constructed Receiver Operating Characteristic (ROC) curves and calculated the areas under the curve, as well as the model’s Positive Predictive Value (PPV). **Results:** of the 412 men, 155 (37.62%) had prostate cancer (PC). The mean age was 63.8 years and the median PSA was 7.22ng/ml. In addition, 21.6% and 20.6% of patients had abnormalities on digital rectal examination and image suggestive of cancer by ultrasound, respectively. The median prostate volume and PSA density were 45.15cm³ and 0.15ng/ml/cm³, respectively. Univariate and multivariate analyses showed that only five studied risk factors are predictors of PC in the study (p<0.05). The PSA density was excluded from the model (p=0.314). The area under the ROC curve for PC prediction was 0.86. The PPV was 48.08% for 95% sensitivity and 52.37% for 90% sensitivity. **Conclusion:** the results indicate that clinical, laboratory and ultrasound data, besides easily obtained, can better stratify the risk of patients undergoing prostate biopsy.

**Keywords:** Prostatic Neoplasms. Biopsy. Prostate-Specific Antigen.

**INTRODUCTION**

Prostate cancer (PC) is a major cause of morbidity and mortality worldwide¹². In the United States, PC is the most commonly diagnosed visceral cancer; in 2015, it is estimated that there were over 221,000 new cases and about 27,500 deaths³, a mortality of 12.4%. In Brazil, it is the second most common cancer in the male population, after nonmelanoma skin cancer, and the second leading cause of cancer death in men⁴. According to the National Cancer Institute (INCA) 61,200 new cases are estimated in 2016⁵.

The prostate specific antigen (PSA) was first used for detection in the 90s. This method revolutionized the disease panorama, causing a considerable increase in the number of men diagnosed with PC, by indicating prostate biopsy. This allowed an early diagnosis of the disease and theoretically increased curability⁶⁷. However, PC is detected in only 30% and 45% of men undergoing initial biopsy, with even lower rates for subgroups with PSA of 4-10 ng/ml, for example⁸⁹, showing a low specificity. For some of these men, the tumor could be very small and the biopsy sensitivity was not enough, but most of the time the patient did not even had PC. This is due to the inability to adequately predict PC positivity likelihood using only PSA and digital rectal examination (DRE). Thus, it is necessary to accurately assess the pretest probability of a positive biopsy, since this procedure is not without risk.

Many risk factors have been correlated with the detection of PC, but their combined contribution can be difficult to quantify. Different predictive models were created in order to work around this problem. Garzotto et al.¹⁰ used data of age, PSA density, DRE and ultrasound data to build their model, but the population was in its majority white and all Americans. Zhao et
al.\textsuperscript{11} developed a model with the Chinese population, restricting PSA values in the 4-10 ng/ml range. These predictive models may have reduced accuracy when used in other target populations, such as the Brazilian. It is known that afrodescendants have a high risk of PC and this population amounts to only 4.2\% of the population present in the work by Garzotto, for example\textsuperscript{10,11}.

The aim of this study was to develop a predictive model for detection of prostate carcinoma by incorporating clinical, laboratory and ultrasonographic data. This would therefore reduce the need for prostate biopsies in patients at low risk, and consequently, the morbidity associated with this procedure.

**METHODS**

We analyzed the records of 445 patients treated between September 2009 and January 2014 in a reference radiology service in Florianopolis – Santa Catarina State, Brazil. We included patients older than 40 years, with seven variables into account (age, DRE, PSA, prostate volume, PSA density, transrectal prostate ultrasound and ultrasound-guided prostate biopsy with at least 12 fragments). We excluded patients with associated diseases that could compromise the data analysis, those previously submitted to prostatic resection and those in use of 5-alpha-reductase inhibitors. We then selected 412 patients for the analysis.

All patients underwent DRE performed by a member of the urology team, classified as normal or abnormal, the latter including prostate hardening, presence of nodulation or irregularities. After DRE, we performed the ultrasound-guided transrectal biopsy. The device used was the Samsung UGEO H60 model USS-H60NF40/US. We measured the prostate in three dimensions and estimated the prostate volume using the modified formula for elongated ellipsoid (0.52 x [length(cm) x depth(cm) x height(cm)]). We checked suspicious areas for the presence of PC. We considered as highly suspicious the hypoechoic nodules and diffusely heterogeneous prostates. We calculated the PSA density by dividing the serum PSA the calculated prostate volume. All patients underwent transrectal prostate biopsy using an 18 gauge, 20cm biopsy needle. We obtained a minimum of 12 fragments from each patient, with harvesting of additional fragments should there be highly suspicious areas. The same pathology laboratory was in charge of examining the biopsy specimens for the presence of adenocarcinoma.

We organized and registered data in a Microsoft Office Excel 2007\textsuperscript{\textregistered} database, with double entry. We performed statistical analysis using the Statistical Package for Social Sciences (SPSS), version 16.0 for Windows.

We describe and present the quantitative variables age, prostate volume, PSA density and PSA as mean, standard deviation, median, minimum and maximum, and the qualitative variables, in frequency ranges according to the appearance in the groups. For comparison between groups, we used the Student’s t test when parametric, the Mann-Whitney test when nonparametric, and the chi-square test when the variables were categorical.

We carried out a logistic regression analysis, having as the outcome variable the presence or absence of PC. In the crude analysis, the variables studied were age, DRE, PSA, prostate volume, PSA density and ultrasound abnormalities suggestive of cancer. In the final model, we included the variables with p<0.20 (age, DRE, PSA, prostate volume, sonographic abnormalities suggestive of cancer). We considered as variables associated with the outcome the ones with p<0.05. We evaluated the goodness of fit by means of sensitivity and specificity metrics and by the construction of the Receiver Operating Characteristic (ROC) curve. We constructed the ROC curve using the MedCalc Statistical Software, version 14.8.1 (Software bvba, Ostend, Belgium). Areas under the curve greater than 0.9 have high accuracy, while 0.7-0.9 indicates moderate precision, 0.5-0.7, low precision, and 0.5, test due to chance\textsuperscript{12}.

**RESULTS**

Table 1 shows the characteristics of the study population. The patients’ age ranged from 40 to 85
years (mean 63.85±8.51). The median PSA level was 7.22ng/ml. DRE was classified as altered in 21.6% of patients (Table 1).

When dividing the age groups, level of PSA and DRE according to biopsy result (positive and negative), was found statistical significance for all: p=0.005 for age; p<0.001 for PSA levels; and p<0.001 for DRE. We divided the study population into PSA lower than 4.0ng/ml, between 4.0 and 10 ng/ml, and greater than 10.0ng/ml, and classified them as for the presence or absence of PC (Table 2).

**Sonographic Findings**

We observed lesions suggestive of prostate cancer in 20.6% of patients. The median prostate volume was 45.15cm³. The median PSA density was 0.15ng/ml/cm³ (Table 1). When dividing these variables into positive and negative biopsy groups, we found statistical significance for all, with p<0.001.

**Biopsy Results**

We obtained a minimum of 12 specimens from all patients during the procedure. Prostate adenocarcinoma was identified in 37.62% (155 of 412 patients – Table 1).

**Development of Predictive Model**

For the univariate logistic regression, the significant predictors for a positive biopsy were: age, with *odds ratio* (OR) of 1.04 (p=0.005); prostate volume, OR 0.96 (p<0.001); altered DRE, OR 1.51 (p<0.001); ultrasound suggestive of cancer, OR 6.2 (p<0.001); PSA levels between 4-10 ng/ml, OR 2.25 (p=0.007); and PSA value ≥10.0ng/ml, with OR 4.80 (p=0.007). We did not observe statistical significance for the variable PSA density: OR 1.53 (p=0.314 – Table 3).

The multivariate logistic regression appointed as significant predictors for the presence of prostate carcinoma: age (p=0.017); prostate volume (p<0.001); altered DRE (p<0.001); ultrasound suggestive of cancer (p<0.001); and PSA (p=0.012) (Table 3).

With the data obtained, we built a ROC curve with all model variables to evaluate the accuracy compared with PSA and DRE alone (Figure 1). We also constructed a ROC curve for the comparison of the model with PSA and DRE combined (Figure 2). The area under the curve was 0.86 for the model, in contrast to isolated PSA 0.65, 0.69 for isolated DRE and 0.71 for combined PSA and DRE.
Setting the sensitivity to 95% for the proposed model and isolated PSA, we found a specificity of 38.15% and 16.34%, respectively. From these values, we calculated the Positive Predictive Value (PPV) using the prevalence of PC in the study patients, and found 48.08% for the model and 40.64% for isolated PSA. This would imply a reduction of 15.46% in the number of biopsies. By setting the sensitivity to 90%, specificity increases to 51.36% for the model and to 20.33% for PSA. We found a PPV of 52.37% for the model and 40.52% for isolated PSA, resulting in a reduction of 22.62% in biopsies.

**DISCUSSION**

The screening for prostate cancer based on PSA and DRE still has important limitations, since the PSA is highly sensitive, but it is not cancer-specific and...
most men with elevated PSA do not have PC.\textsuperscript{13} The difficulty of screening for this disease is to establish protocols that have high positive predictive values, to stratify high-risk individuals for PC.

A branch of the Prostate Cancer Prevention Trial (PCPT) investigated the prevalence of PC in 2950 men who used placebo and had PSA levels below 4.0ng/ml and DRE considered normal, i.e. patients considered of low risk for PC.\textsuperscript{14} The results showed that the disease could be diagnosed in all PSA levels, including high-risk tumors. This indicates that the PSA should not be considered as the only factor in choosing patients for prostate biopsy.\textsuperscript{15} The findings of this study corroborate this, since approximately 20\% of patients with PSA less than 4.0ng/ml had PC diagnosis (Table 2).

Because of these limitations, statistical models began to be developed to more accurately predict the risk of PC in the biopsy. Eastham \textit{et al.}\textsuperscript{16} published, in 1999, the first study demonstrating a model that included age, ethnicity and PSA. Only PSA was an independent predictor of positive biopsy in their analysis, with an area under the curve of 0.75. However, the study was conducted during the period in which the default was the harvesting of six prostate fragments. This may limit the analysis results, as this pattern has less sensitivity to the currently used twelve fragments.\textsuperscript{17} In the previously mentioned Prostate Cancer Prevention Trial, Thompson \textit{et al.}\textsuperscript{18} used the placebo arm results to assess the risk of PC considering age, ethnicity and family history. Although this study has been innovative and had wide acceptance, there are certain limitations. In PCPT, 89\% of the 5519 patients had a level of PSA in the range considered "normal", i.e., <4.0ng/ml, and only 150 patients had PSA levels greater than 6ng/ml, unlike what we find in many clinical settings. Furthermore, the PCPT was limited to men over 55 years, excluding its use a large number of patients.

Karakiewicz \textit{et al.}\textsuperscript{1} developed two predictive models with three independent cohort data, where men were referred for prostate biopsy based on PSA values, percentage of free PSA and alterations in DRE. They collected the data from the first and second cohorts in Montreal, Canada, where 4193 men underwent biopsy guided by ultrasound and had six fragments removed, after digital rectal examination and measurement of PSA values. Of these, 514 also underwent measurement of free PSA. The third cohort consisted of 1762 patients from the University Hospital Hamburg - Eppendorf, Germany. These men had criteria for sextant biopsy and had collected PSA, percentage of free PSA and DRE. The predictive model based on age, DRE, PSA and percentage of free PSA showed better accuracy than the model that used only age, DRE and PSA, with areas under the ROC curve of 0.77 and 0.69, respectively.\textsuperscript{1} One limitation of this study was the failure to assess the impact of ethnicity, all patients being Caucasian. Another limitation was the use of only six biopsy fragments.\textsuperscript{19}

This study evaluated, within the same population, the best combination of variables to be used for PC prediction and then created models that meet these characteristics. We saw that the most commonly used criteria for screening of patients with prostate cancer, PSA and DRE, have low accuracy, with values of area under the ROC curve of 0.71 when used together. The model developed and demonstrated in this work presented the best accuracy among the tested combinations, with values of the area under the

<table>
<thead>
<tr>
<th>PSA</th>
<th>Total</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 4.00</td>
<td>52</td>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td>4.00-10.00</td>
<td>255</td>
<td>89</td>
<td>166</td>
</tr>
<tr>
<td>&gt; 10.00</td>
<td>105</td>
<td>56</td>
<td>49</td>
</tr>
<tr>
<td>Total</td>
<td>412</td>
<td>155</td>
<td>257</td>
</tr>
</tbody>
</table>

\textit{PSA: prostate specific antigen; n: number.}
Ludwig
Integrated predictive model for prostatic cancer using clinical, laboratory and ultrasound data


ROC curve of 0.86 for predicting the risk of PC. The results obtained are consistent with those obtained in other studies\textsuperscript{1,10,16}. Most of the published studies have been limited to PSA values less than 10.0ng/ml, with the justification that any patient with values above that would be subjected to a prostate biopsy\textsuperscript{10,11,16}. In this study, we chose not to limit the PSA, as there was a rate of nearly 50% negative biopsies in this population subgroup, which would open room for a better patient's selection for biopsy including these PSA values. It would be a new paradigm that needs further study and deepening, but would have the main benefit of avoiding repeated biopsies in such patients.

Some limitations are present in the model presented in this study. First, it we did not take into account the possible outcome of a repeated biopsy for those with negative findings on an initial biopsy, taking into consideration that false negatives may occur\textsuperscript{21}. Second, we collected secondary character data retrospectively, and thus, their records were not designed and completed to meet the research objectives. Finally, the proposed model has not been validated externally. This can cause it to present different results in other populations. This raises the need for other research centers to confirm and validate the results of any predictive model in use\textsuperscript{22-24}.

The results indicate that the clinical, laboratory and ultrasound information, besides easily obtained in clinical practice, can better stratify the risk of patients undergoing prostate biopsy.

Table 3. Gross and adjusted analysis of factors associated with prostate cancer.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gross Regression coefficient</th>
<th>Odds ratio (95% CI)</th>
<th>p value</th>
<th>Adjusted Regression coefficient</th>
<th>Odds ratio (95% CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.04</td>
<td>1.04 (1.005:1.071)</td>
<td>0.005</td>
<td>0.39 (1.006:1.072)</td>
<td>1.04 (1.006:1.072)</td>
<td>0.017</td>
</tr>
<tr>
<td>Prostate volume</td>
<td>-0.04</td>
<td>0.96 (0.946:0.973)</td>
<td>&lt; 0.001</td>
<td>-0.04</td>
<td>0.96 (0.945:0.973)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>DRE</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Altered</td>
<td>1.51</td>
<td>4.53 (2.308:8.800)</td>
<td>&lt; 0.001</td>
<td>1.62 (2.609:9.776)</td>
<td>5.05 (2.609:9.776)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Ultrasound</td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Altered</td>
<td>1.83</td>
<td>6.2 (3.015:12.807)</td>
<td>&lt; 0.001</td>
<td>1.99 (3.562:15.012)</td>
<td>7.32 (3.562:15.012)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>PSA Density</td>
<td>0.43</td>
<td>1.54 (0.668:3.516)</td>
<td>0.314</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PSA</td>
<td></td>
<td>0.007</td>
<td></td>
<td>0.012</td>
<td>3.15 (1.201:8.267)</td>
<td></td>
</tr>
<tr>
<td>&lt; 4.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4.00-10.00</td>
<td>0.81</td>
<td>2.25 (1.079:4.701)</td>
<td>1.27</td>
<td>3.54 (1.535:8.177)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 10.00</td>
<td>1.60</td>
<td>4.80 (2.181:10.566)</td>
<td>1.15</td>
<td>3.15 (1.201:8.267)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PSA: prostate specific antigen; 95% CI: 95% confidence interval.
RESUMO

Objetivo: desenvolver um modelo preditivo para estimar a probabilidade de câncer prostático previamente à biópsia. Métodos: de setembro de 2009 até janeiro de 2014, 445 homens foram submetidos à biópsia prostática em um serviço de radiologia. Pacientes com doenças que pudessem comprometer a análise de dados, submetidos à ressecção prostática ou usando inibidores de 5-alfa-redutase foram excluídos do estudo. Dessa forma, 412 pacientes foram selecionados. Variáveis incluídas no modelo foram idade, antígeno prostático específico (PSA), toque retal, volume prostático e achados ultrassonográficos anormais. Curvas de Características Operacionais (ROC) foram construídas e áreas sob a curva foram calculadas, assim como os Valores Preditivos Positivos (VPP) do modelo. Resultados: dos 412 homens, 155 (37,62%) tinham câncer de próstata (CAP). A média da idade foi 63,8 anos, a mediana do PSA foi 7,22ng/ml. Além disso, 21,6% e 20,6% dos pacientes apresentaram anormalidades no toque retal e imagem sugestiva de câncer pela ultrassonografia, respectivamente. A mediana do volume prostático e da densidade do PSA foram 45,15cm³ e 0,15ng/ml/cm³, respectivamente. Análises univariada e multivariada demonstraram que apenas cinco fatores de risco estudados são preditores de CAP no estudo (p<0,05). A densidade de PSA foi excluída do modelo (p=0,314). A área sob a curva ROC para predição de CAP foi 0,86. O VPP foi 48,08% para sensibilidade de 95% e 52,37% para sensibilidade de 90%. Conclusão: Os resultados indicam que informações clínicas, laboratoriais e ultrassonográficas, além de serem facilmente obtidas, podem estratificar melhor o risco de pacientes que serão submetidos à biópsia prostática.


REFERENCES