Reviewer’s report

Title: Evaluation of Prediction Models for the Staging of Prostate Cancer

Version: 1 Date: 21 June 2013

Reviewer: Robert Veltri

Reviewer’s report:

BMC Med Info_1414589052888883_PCa: “EVALUATION OF PREDICTION MODELS FOR THE STAGING OF PROSTATE CANCER”

TITLE: Acceptable based upon the subject of the review article.

AUTHOR’S COMMENTS:

General Design: The author’s conducted a retrospective review of the Irish Prostate Cancer Research Consortium database and 603 patients. The clinicopathological variables used to predict prostate cancer (PCa) stage (organ-confined vs. non organ-confined) included biopsy Gleason score, age, PSA and clinical stage. Statistical models were built using multivariate logistic regression (MLR), random forests and k nearest neighbors. The predictive ability of the models was examined using discrimination metrics (sensitivity, specificity, positive predictive value, negative predictive value, Youden index, Brier score and area under the curve (AUC)), calibration curves and clinical relevance was explored using decision curve analysis.

Abstract: No comments.

Introduction: This introduction is outdated and should be improved to address several key issues on the subject. Current cancer statistics data has recently been updated (1). The current decision to perform a biopsy is provided in the 2010 NCCN prostate cancer guidelines (2). One of the most current clinical dilemmas involves overdiagnosis and overtreatment though has been noted often as a subject of the active surveillance option for CaP (2-6) and how to manage these decisions. Not fully addressed in the introduction is the probability of missing significant cancer (high grade) during a systematic biopsy of 12 or more cores that include the lateral apex and central zone (4-5). As to the subject of Gleason grading, besides the original work of Donald Gleason, this subject and its importance to the management of CaP should include the work of Dr. Jonathan Epstein and Peter Albertsen (8-11). Further, several other approaches to PCa staging, besides those mentioned, are not presented (12-15). Today the Partin Tables remain a standard of care for patients diagnosed with PCa among residents and practicing urologists.

2. Mohler JL. The 2010 NCCN clinical practice guidelines in oncology on prostate

Materials and Methods

Study Population
Comments and Queries:
1. The authors only used the Partin criteria of biopsy Gleason score, DRE status (Clinical stage) and total PSA value.
2. Why not introduce PSA Density (PSAD) as an additional variable?

Clinical and Pathological Assessment
Comments and Queries:
1. Define organ-confined PCa clearly!

Statistics
Comments and Queries:
1. For the multivariate logistic regression (MLR) modeling, what variable cut-off value was employed to construct the MLR models?
2. Did the authors apply forward or backward MLR modeling?
3. Did the authors supply the final equations for their MLR models (i.e. model constant, Beta coefficients, etc.) to test their MLR models?
4. The equation #1 has poor resolution.
5. The last sentence in the paragraph uses different fonts.

RESULTS
Table 1. Mathematical formulae underlying the classification models
Comments and Queries:
1. These are not mathematical formulae; rather these are the variables to input for the models and in its present form the Table is not useful.
2. Is the logistic Regression actually Multivariate Logistic Regression and is it backward or forward stepwise?

Table 2. Prostate Cancer Research Consortium patient cohort characteristics.
Comments and Queries:
1. There were 384 cases with no DRE reported and that will have a severe effect on the analysis and the outcomes since this variable is critical to any predictive solution.
2. Where are the data for the number of organ-confined and non-organ confined cases?
3. Where is the data for ECE, LN, SV, etc. involvement status?
4. If you have TRUS and serum PSA data, you should have PSAD data. Why not use it?

Table 3. Percentage of Gleason score upgrading or downgrading.
Comments and Queries:
1. The upgrading error is quite significant but not necessarily unexpected. Was this error due to sampling errors contributed by changes in biopsy sampling methodology over the years?

2. Once again the PSAD information could be of value to evaluating grading errors and also to possibly improve the models.

Table 4. Discrimination of prediction models and individual clinical variables

Comments and Queries:

1. The LR analysis of the data (Youden and Brier Indices) on pages 11- and 12 is quite useful. It clearly appears to me that Logistic Regression was the best performer and that is usually the case (i.e. Partin Tables) and other predictive algorithm publications.

2. However, with 63.3% of the DRE status data missing, how much would the loss of so much data impact the predictive models?

3. Should the authors have used only the cases with all variables available?

4. How well would have the Partin Tables worked with cases where all the variables were available? This can be done on line at the Brady Urological Research Institute website.

Figures 2a-g: Calibration curves which illustrate predicted probabilities on the x-axis and the actual outcome on the y-axis. In this case the outcome is a binary variable hence Loess smoothing (red line) was used to estimate the actual outcomes for (a) logistic regression model, (b) random forests model, (c) k nearest neighbours model, (d) biopsy Gleason score, (e) age, (f) PSA and (e) clinical stage.

Comments and Queries:

1. Calibration curves support conclusions for the three models compared and Logistic Regression appears to be the best.

2. Also, the analysis of the three input variables seem acceptable.

Figure 3: Decision curves for (a) logistic regression model, random forests model and k nearest neighbours model and (b) biopsy Gleason score, age, PSA and clinical stage. In each decision curve the solid, thin black line represents assuming no one has NOC PCa and the thin grey line represents assuming everyone has NOC PCa.

Comments and Queries: No comments and I agree with conclusions.

Summary of the Results Section:
The author’s constructed evaluated three multivariate modeling statistical tools to predict organ-confined and non-organ-confined PCa outcome. They used Logistic Regression, Random Forests and k nearest Neighbors. The author’s
methods to present and analyze their data were most acceptable and their conclusion with regards to the best approach (Logistic Regression) and the quality of the three variables used to test the various models was thorough. However, I would like to emphasize the following points of criticism:

• There was a 30% error in staging in the patient cohort to start with. This is a serious defect in design since the Partin Tables and other algorithms have a very high degree of accuracy to predict OC.

• The missing information for clinical staging (DRE) may have compromised the modeling. The weakest variable on calibration assessment.

• Perhaps the use of PSAD since the information was available may have been useful to include.

• The authors should have tested the Partin Nomogram and Tables to confirm their results.

• The final conclusion would suggest to treat everyone since the predictive probability data to determine OC and NOC were not more than 70%.

• Below are some reviews that may be useful:


DISCUSSION:

Comments and Queries:

1. The authors clearly define weaknesses in their study as it relates to post-operative staging (30% incorrect) and the high degree of upgrading noted from biopsy to radical prostatectomy Gleason scores.

2. Adding a fourth variable, age, was of little consequence to the algorithm solutions.

3. No mention of other additional potential markers (i.e. PSAD, surface of area of cancer in the biopsy etc.).

4. What about new serum and urine biomarkers to make such predictions (a-f):


2. Points not discussed were: race, new molecular (protein, genetic and epigenetic biomarkers), new PSA derivative etc.
3. The solution is not more statistical models, rather better and more meaningful biomarkers to improve the prediction outcomes.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

Responses:
1. no
2. no
3. no
4. no
5. no.
6. no

'I declare that I have no competing interests'