Reviewer's report

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

Version: 1 Date: 20 February 2013

Reviewer: Matthew Cooperberg

Reviewer's report:

EVALUATION OF PREDICTION MODELS FOR THE STAGING OF PROSTATE CANCER

General

The authors perform a rigorous assessment of various mathematical models with the goal of predicting pathological stage from a partial set of biopsy information (i.e., PSA, partial stage information, biopsy Gleason score), with notable absence of T2 sub-classification and measures of extent of biopsy involvement (see below). The conclusion—that logistic regression performed best—is reassuring, since that technique underlies most popular nomograms, the CAPRA score, etc. But ultimately I am not clear what is new here. About ten years ago, similar papers were published with similar goals, again employing a range of techniques including random forests, CART analysis, and neural networks. None outperformed regression and none really caught on. 5-fold cross-validation does not replace true external validation, which has been performed for the Partin tables, nomograms, and CAPRA score. These existing tools should be used as reference groups, rather than starting de novo, since this question has been so thoroughly investigated in the past. Ultimately, this is an admirably thorough description of a well-characterized problem, but does not add much new to the existing knowledge base.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. The outcome groups discussed in the introduction are NOT mutually-exclusive. While ECE and SVI define pT3a and T3b respectively, the two together are worse than either alone, and LNI can co-exist with either, neither, or both. Multiple, well-validated postoperative risk instruments (e.g., Kattan and Stephenson nomograms, CAPRA-S score) have shown this in the past. This is a major problem with the methods and study design.

2. The introduction is much too long – should be limited to 1-2 paragraphs to frame the specific question being asked. This is particularly relevant because this is a well-established question. The majority of the material in the intro can be shortened or moved to the discussion.

3. A “nomogram” is a graphical representation of a prediction model, not the
model itself. So searching Pubmed for “nomograms” will miss important tools such as the Partin tables and CAPRA score, both of which predict pathologic stage without use of a nomogram per se.

4. Extent of biopsy involvement (# of cores, % of cores, % of tissue etc.) is a key predictor of outcome in multiple models (e.g., Stephenson, CAPRA), so its absence here is quite problematic and may help explain the poor performance of these models.

5. There are dozens of candidate markers in development. If the authors wish to raise this issue in discussion, they need to be more thorough in their review rather than just citing their own marker set in development.

Discretionary Revisions (which the author can choose to ignore)

1. none

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests