Reviewer's report

Title: Artificial neural network (ANN) velocity better identifies benign prostatic hyperplasia but not prostate cancer compared with PSA velocity

Version: 2 Date: 16 April 2008

Reviewer: Jonas S Almeida

Reviewer's report:

This report describes an application of artificial neural networks to distinguish between prostate cancer (PCa) and benign prostate hyperplasia (BPH) using two PSA related measures, %free PSA and PSA velocity. A number of molecular forms of PSA are used in an attempt to find predictive covariates.

This is an important study and the authors have gathered a critical amount of clinical information. This reviewer is of the opinion that the analysis of this data is very relevant and should be reported. This reviewer is also convinced by the authors’ argument that ANN are a good choice for non-linear regression methodology. This viewpoint had been proposed by the authors in an influential piece in Nature Clinical Practice – Urology [JUNE 2005 VOL 2 NO 6 p262].

However, the description of the ANN analysis just by itself suggests serious flaws in its configuration. Three in particular are very problematic and are often misleading and are listed below. All this reports says with regard to the ANN is that it “was constructed with the SPSS-module Neural connection 2.0 (SPSS) as described earlier [22]” (Clin Chem 2002, 48:1279-1287)

The three questions that will cross the mind of readers with a statistical inclination:

1. How was the ANN configured – how many hidden nodes? If an adaptative method was used, was it by pruning, by forward addition? Or was this not optimized and 3 hidden nodes were preset as in [22]?

2. The ROC curves are derived using the same type of data that was used to train the ANN. Is there any kind of cross-validation? External validation (with data not at all used for training)? Is there any optimization halting procedure to avoid over-fitting? Early stopping criteria are a standard feature of ANN training. Was any of that in place? Whatever the specific answers, the critical issue in this objection is that of validation. As per the description in the report itself this reviewer would expect the ANN to be over-trained.

3. If a number of clinical parameters such as age are already known to be co-variates, why weren’t they considered in the analysis? That is where machine learning approaches become the most useful – by being able to include a large variety of parameters without concern for their scales or distributions. If a variable selection procedure were in place (such as boosting or more advanced methods
using evolutionary approaches) then the proposed interpretations could be backed by sensitivity analysis (the report says “results not shown” on the topic of assigning predictive sensitivity).

Most of the questions posed here are addressed in that report. However, that was a different study and it is not clear what exactly can be assumed to apply to this much smaller study. Some probably can – such as a the 10-fold cross-validation – but for others I find no clue. The suggestion is that the relevant material be pulled in and the missing answers be added. This could be listed in a compact, small, “ANN analysis section”, in a tabular format would be fine, so the computational statisticians in the audience can put their concerns to rest.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests