Reviewer's report

Title: Estimates of observed sensitivity and specificity can be biased when reporting the results of the second test in a screening trial conducted in series

Version: 1 Date: 6 August 2009

Reviewer: Angel Cronin

Reviewer's report:

This is a well written paper and is very relevant, as more and more diagnostic tests are being conducted in series. I have only minor comments to improve the clarity of the paper.

Minor essential revisions:
1. In the section describing the single test design (pages 3-4), it would be helpful to have a sentence or two stating that this scenario is referred to as differential verification bias, and that it has been shown that this type of bias leads to overestimates of both sensitivity and specificity.
2. On page 11, last paragraph, it is discussed that the observed sensitivity and specificity of test 2 for the series design are lower than the observed sensitivity and specificity for the single test design. I think that this discussion should be expanded. First, this is explained because “in the series design, only a portion of cases proceed to test 2, which decreases both the numerator and denominator of the observed sensitivity and specificity of test 2. The net result is a decrease in the observed sensitivity and specificity…” This follows only if the numerators decrease proportionately more than the denominators decrease. Can the authors elaborate on why, in this setting, the numerators would decrease proportionately more than the denominators? Second, it is not obvious to me why the specificity of test 2 would be lower (and therefore have less bias) for the series design compared to the single test design. In order to get the second test in the series, you need to test negative on the first test (generally speaking). It would seem to me that the true negatives would be oversampled here with respect to the truth, even more so than in the case of the single test.

Discretionary revisions:
1. For tables 1 and 4, I personally did not like the notation with “X” to indicate that the gold standard result “does not add more information about the disease status”. To me, confirmation of being disease positive by the infallible reference test does add information about the disease status.
2. On page 10, the means of test 1 and test 2 are given, but the standard deviations of the two tests are not.

5. On page 11, middle of the page, it states “Both the numerator and denominator of the observed estimates of specificity include misclassified cases.” For clarity, I suggest stating that this is because some patients with negative test results truly have the disease; this is not observed because they do not undergo
the gold standard assessment, and so they are assumed to not have the disease.

6. For the results, it was assumed that both test scores followed a normal distribution. The authors note that most screening test data are strongly left skewed. Would the relationships depicted in the figures differ with skewed data?

7. In a couple of places, the results are interpreted in the context of comparing the performance two screening tests (abstract conclusion, and on page 10). But these results hold even if the investigators are solely interested in characterizing the performance of a single test, as was the case in the Lehman paper with MRI. I would particularly suggest rewording the sentence in the abstract, because I wouldn’t want someone to think that this methodology is relevant solely when comparing two screening tests.

**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.