Reviewer's report

Title: Tracking type specific prevalence of human papillomavirus in cervical pre-cancer: a novel sampling strategy

Version: 1 Date: 25 August 2011

Reviewer: Mark Myatt

Reviewer's report:

Abstract : Paragraph 2 : Method : The use of the term "precision" is wrong since the reported method classifies rather than estimates prevalence. I think that the authors mean "accuracy" or, perhaps, "performance" or "utility".

Abstract : Paragraph 3 : Results : The use of "false negative" and "false positive" is confusing since a three-class method is being tested. This reflects a confusion throughout the report. It is not clear whether a two or three class method is required and which is being tested.

Page 5 : "... in public health" : Some references to public health applications of sequential sampling methods is required. These methods (commonly called "LQAS") are very common in service epidemiology.

Page 5 : "... method proposed ..." : The truncated sequential sampling method has been adopted by the WHO and CDC and is in current use in many settings.

Page 5 : "... prevalence estimate ..." : Do the authors mean "classification"?

Page 5 : "... false positive or a false negative" : See above.

Page 5 and elsewhere : The text seems to switch between reporting a two class method and a three class method.

Page 5 : "... estimate of prevalence" : See above.

Page 5 : "... high precision" : See above.

Page 6 : Formula 1 : This formula seems to be incorrectly specified. The numerator should be:

-\ln((1 - \alpha) / \beta))

Page 6 : Formula 3 : The numerator might be simplified to:

\ln(q1 / q2)

Page 6 : Paragraph 1 : The text is referring to a two class approach.

Page 6 : Paragraph 1 : It is unclear what is intended by the "... plus tolerance boundaries".
Page 6: Paragraph 1: q2 should be defined as 1 - p2 (not just "p2").

Page 6: Formula 4: This is the formula for a two-class binomial sequential sampling classifier. It seems that the authors are confused between the two-class binary sequential sampling method and the three class TSS method.

Page 6: Paragraph 3: It is unclear what is intended by the "... plus tolerance boundaries".

Page 7: Paragraph 3: Substituting:

\[ p_1 = 0.05 \] <- from Page 7: Paragraph 2
\[ p_2 = 0.15 \] <- from Page 7: Paragraph 2
\[ \alpha = 0.01 \] <- from Page 7: Paragraph 2
\[ \beta = 0.025 \] <- from Page 7: Paragraph 2

yields \( N_{\text{max}} = 138 \) not the 48 quoted in the text. This (large) discrepancy should be accounted for.

Page 8: Paragraph 1: "... uniform distribution ...": Myatt and Bennett did not use a uniform distribution of prevalences but a distribution reflecting reported prevalences of HIVDR (see Table 3 in Myatt and Bennett). Uniform distribution within band was used but the frequency of each band was non-uniform. The presentation of the use of a triangular distribution as "... a methodological enhancement intended to simulate the most frequent sampled value for the prevalence ..." is not, therefore, correct.

Page 8: Paragraph 2: The fitting of separate probability of classification curves is a methodologically flawed approach. The three curves are not independent of each other. For example, if \( P(\text{Low}) = 0.2 \) and \( P(\text{High}) = 0.1 \) the \( P(\text{Moderate}) \) is constrained to be 0.7. A better approach would have been to use a larger number of simulation replicates in at each of the tested prevalences and plot the calculate probability of classification.

Page 8: Paragraph 4 / Figure 2: It is unclear whether prevalences in excess of 30% are likely to be found in real-world applications of this method. If such prevalences are unlikely then testing at prevalences above 30% makes little sense.

Page 8: Paragraph 4: The \( R^2 \) values are without clear use or meaning here.

Page 9: Paragraph 2 / Figure 2: Figure 2 looks wrong:

1. In the sense that it represents poor performance of the classifier at returning a moderate prevalence classification. At \( M_{\text{max}} = 47 \) the HIVDR classifier works better than this \( N_{\text{max}} = 48 \) classifier.

2. It seems that the probability of classification varies considerably between prevalences that are just 1% apart. This plot looks more like 100 prevalences
sampled 50 times each (i.e. 5000 replicates) was used rather than 100 prevalences sampled 5000 times each (i.e. 500 000 replicates was used). The authors should check their computer code.

Table 1: The average sample numbers appear to be incorrect. The average sample number should be much higher for moderate prevalence classifications.

**Level of interest:** An article of insufficient interest to warrant publication in a scientific/medical journal

**Quality of written English:** Not suitable for publication unless extensively edited

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

**Declaration of competing interests:**

I declare that I have no competing interests