Reviewer's report

Title: Evaluation of HIV Testing Algorithms in Ethiopia: The role of the tie-breaker algorithm and weakly reacting test lines in contributing to a high rate of false positive HIV diagnoses

Version: 2
Date: 20 November 2014

Reviewer: Jörg Schüpbach

Reviewer's report:

The revised manuscript is still difficult to read, and remains unclear in many aspects.

One major difficulty arises from using the term "tiebreaker". While WHO uses this term for the 3rd decisive test after discrepant results of two initial tests always performed in parallel, the "tiebreaker strategy" implemented in Ethiopia apparently is a sequential testing approach in which only samples reactive in the first test undergo further testing. Why this should be fundamentally different from what you call a sequential approach (with a confirmatory test at the end) remains unclear. What you actually compare are variations of sequential testing algorithms – at least that is what you describe under Methods. Your confirmatory test has no other role than that of a tiebreaker. Or does Table 3 actually show results of initial parallel testing? I really do not know what you mean! And this goes throughout the paper.

Table 2. You should also add the OIC to this table (and perhaps also WB!).

Table 3. I don't understand what "Discordant" means when you use a confirmatory test at the end of the testing sequence. Was the end result correct or not? What was discordant? Was it not possible to clarify with the OIC? I also think that Table 3 is unnecessarily complicated. What interests me is how the different algorithms perform in resolving the 203 positive and the 225 negative results, as defined by the gold standard. One doesn't know when looking at the table in its current state. Why not make things clear and understandable?

Table 4. I do not understand why, with a specificity of 100%, the PPV can be less than 100%. A specificity of 100% means that you do not have any false-positive samples. Thus, as per definition, PPV = TP / (TP+FP), PPV will equal TP/TP, i.e. 1 = 100%.

Table 5. As long as the issue of Table 4 is not resolved I do not trust Table 5.

Table 6. Overlap with Table 5: if the multitest algorithms with strong positive signals have good PPV and NPV, the individual tests must also have good PPV and NPV.

Table 7.
The univariate analysis is of little value. You cannot draw any conclusions from it. You should do a multivariate analysis (logistic regression) to sort things out or otherwise leave the table away.