Reviewer's report

**Title:** Diagnostic performance of line-immunoassay based algorithms for incident HIV-1 infection

**Version:** 1  **Date:** 26 January 2012

**Reviewer:** Ming-Wei Lin

**Reviewer's report:**

In this manuscript, Jörg Schüpbach et al sought to determine the diagnostic performance of Inno-Lia algorithms for identifying incident infections in patients with known duration of infection and evaluated them in annual cohorts of HIV notifications.

Their results showed that The 10 best algorithms had a mean raw sensitivity of 59.4% and a mean specificity of 95.1%. They also tried to validate the 10 best algorithms in four annual cohorts of HIV-1 notifications totalling 2'595 patients and they found a mean IIR of 0.35 in 2005/6 (baseline) and of 0.45, 0.42 and 0.35 in 2008, 2009 and 2010, respectively. The following questions and comments need to be addressed or clarified.

**Major Compulsory Revisions:**

1. It is not clear why the authors only used patients with incident infection for determination of the diagnostic sensitivity of the Inno-Lia algorithms, while used patients with older infection for determination of the diagnostic specificity of the algorithms. It will be necessary to calculate the information of sensitivity, positive predict value (PPV), and negative predict value (NPV) in both patient cohorts with incident and older infection, respectively.

2. Were the patients with older infection used for evaluation of specificity the same as those patients in the previous study published in 2011 BMC Infectious Disease? It should be clarified and addressed in the manuscript.

3. A total of 26 algorithms (Algs) for incident HIV-1 infection were developed and employed in the study, however, the twenty-six algorithms are not completely independent. It may be problematic to apply this approach to evaluate the diagnostic performance of Inno-Lia antibody patterns simultaneously.

4. The authors suggested the use of the 10 best-performing algorithms in combination to minimize the impact of their individual limitations and the effects of sample bias and to monitor changes in IIR in a population over time. The cost-effectiveness should be considered and evaluated in the manuscript. Moreover, the diagnostic sensitivity remains far below the standards for tests used for diagnostic purposes. It would be advisable to combine other assays with high sensitivity.

5. It is not clear if Cohorts A, B, C, D were came from the same cohorts? If they were not the same cohorts, it will be inappropriate to use paired t test to make
comparisons.

6. What is Fisher's r to z test? It is quite confused why the authors evaluate correlation by Fisher's r to z test.

Minor Essential Revisions:

1. The references should be modified to have a consistent format.

Level of interest: An article whose findings are important to those with closely related research interests

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'