Author's response to reviews

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

Authors:

Jorg Schupbach (schupbach.jorg@virology.uzh.ch)
Leslie R Bisset (leslie.bisset@swissonline.ch)
Martin D Gebhardt (martin.gebhardt@bag.admin.ch)
Stefan Regenass (Stephan.Regenass@usz.ch)
Philippe Burgisser (philippe.burgisser@bluewin.ch)
Meri Gorgievski (M. Gorgievski@ifik.unibe.ch)
Thomas Klimkait (Thomas.Klimkait@unibas.ch)
Corinne Andreutti (C. Andreutti@lasource.ch)
Gladys Martinetti (gladys.martinetti@ti.ch)
Christoph Niederhauser (christoph.niederhauser@bsd-be.ch)
Sabine Yerly (Sabine.Yerly@hcuge.ch)
Stefan Pfister (stefan.pfister@ifik.unibe.ch)
Detlev Schultze (detlev.schultze@zlmsg.ch)
Marcel Brandenberger (Marcel.Brandenberger@synlab.com)
Franziska Schoni-Affolter (Franziska.Schoni-Affolter@chuv.ch)
Alexandra U Scherrer (alexandra.scherrer@usz.ch)
Huldrych F Guenthard (huldrych.guenthard@usz.ch)
Swiss HIV Cohort Study SHCS (noname@noreply.ch)

Version: 3 Date: 28 March 2012

Author's response to reviews: see over
We wish to thank the two reviewers for their critical re-assessment of our revised paper.

**Reviewer:** Maurizio Zazzi

Dr. Zazzi agrees with the improvements we have made and finds that the paper — including its statistics — is now acceptable for publication.

**Reviewer:** Ming-Wei Lin

Dr. Lin maintains that the use of paired t-test for assessment of the incident infection rates of consecutive annual cohorts of HIV notifications is inadequate. According to her statistical understanding "the use of paired t test is inappropriate unless the subjects in cohorts A, B, C, and D were repeated."

**Response:**

Indeed, the most common use of a paired t-test is the comparison of two measurements from the same individual or experimental unit. This is, however, not the only situation in which this test may be used. The test is also appropriate when some other natural pairing exists. The StatView Reference Handbook (SAS Institute Inc., 1998, second edition) gives the following example (p. 30): "A survey of husbands and wives is designed to test for differences of opinion on particular issues. Each couple's responses are viewed as a pair and tested for differences with a paired t-test." In this case, the persons are definitely not repeated. The StatView Reference continues: "The critical issue is whether a pair's responses are more likely to be similar than responses from random experimental units. When the pair's responses are likely to be consistently more similar, a paired t-test is more powerful than an unpaired t-test."

This criterion is certainly met in our comparisons. It is evident that the incidence results in two different years are more similar for a the same algorithm than for randomly selected different algorithms.

That Dr. Lin is not correct with her statement is also demonstrated by the statistics textbook Practical Statistics for Medical Research of Douglas C. Altman, a true expert (Chapman & Hall/CRC edition 1999). In the chapter *Comparing groups — categorical data* he cites another example where paired comparisons are required despite the fact that the two groups compared are different (p. 236, top). He refers to a "group of 32 marijuana users that are compared with 32 matched controls with respect to their sleeping difficulties. Seven of the marijuana users reported sleep difficulties sometimes or always compared with 13 of the controls." He continues "Because the groups were individually matched we should not treat the observations as independent and thus need different methods from those described in..."
the previous section" (the one dealing with independent groups; note of J.S.).

The analogy to our comparisons is obvious. The incident infection rates in different years are certainly paired for each of the different algorithms.

In conclusion, we maintain our standpoint and refuse to make any changes.

Minor comments:
1. What is logistic likelihood ratio (LLR) chi-square values? This term seems not commonly used in the statistical methods.

Response:

Chi-square(d) ($X^2$) distribution and Chi-square(d) tests are the basis for the comparison of categorical data and the evaluation of frequency tables and are explained in every textbook of statistics (e.g. also in the Altman book). On p. 410 of the Altman book, Dr. Lin would also find information on the basic parameters describing test performance, namely the terms diagnostic sensitivity and diagnostic specificity.

No changes made.