Author's response to reviews

Title: Can HIV incidence testing be used for evaluating HIV intervention programs?

Authors:

Agnés Fiamma (afiamma@ucla.edu)
Pascale Lissouba (pascale.lissouba@inserm.fr)
Amy E Oliver (aoliver4@jhmi.edu)
Beverley Singh (beverleys@nicd.ac.za)
Oliver Laeyendecker (plaeyen@jhmi.edu)
Thomas C Quinn (Tquinn2@jhmi.edu)
Dirk Taljaard (dirk@progressus.co.za)
Bertran Auvert (bertran.auvert@uvsq.fr)

Version: 4 Date: 21 April 2010

Author's response to reviews: see over
Can HIV incidence testing be used for evaluating HIV intervention programs?

The authors would like to thank the reviewers for their insightful and valuable remarks, which have been very useful in improving the manuscript.

Editors:

<table>
<thead>
<tr>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns. We look forward to receiving your revised manuscript by 21 April 2010. If you imagine that it will take longer to prepare please give us some estimate of when we can expect it. Ethics - Experimental research that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee. Research carried out on humans must be in compliance with the Helsinki Declaration (<a href="http://www.wma.net/e/policy/b3.htm">http://www.wma.net/e/policy/b3.htm</a>), and any experimental research on animals must follow internationally recognized guidelines. A statement to this effect must appear in the Methods section of the manuscript, including the name of the body which gave approval, with a reference number where appropriate. Answer: We now indicate the name of the body which granted the approval, the University of the Witwatersrand Human Research Ethics Committee, and the reference number in the methods section. We also indicate that a written informed consent form was signed by each participant.</td>
</tr>
</tbody>
</table>
## Reviewer #1: Major Compulsory Revisions

| R1A.1 | In describing the avidity assay it is not clear how the assay is performed. Specifically, it should be stated that each sample is loaded into 2 wells of the ELISA plate and the Avidity Index was calculated as the ratio of the OD values. OD of treated specimen/OD of untreated specimen x 100 = AI  
| Answer: We now describe precisely how the avidity assay was performed. |
| R1A.2 | An avidity index should never be reported as a value greater than 100% as this is biologically not relevant. A number of greater than 100% can be obtained but this occurs usually at the extremes of the serologic assay. The suggestion would be to report any value greater than 100% as 100%.  
| Answer: All values greater than 100% are now reported as 100%. |
| R1A.3 | It is not clear from the methods how samples that would be classified as negative based on the Bio-Rad Kit negative cutoff are handled for the AI calculation. For example, assume that the cutoff of the assay for negative vs positive is .280 and you obtain values of - for treated of .150 and untreated .275. The calculated AI would be 54.5 and would likely indicate a long term infection when in reality it would be a recent infection because the sample was negative in the assay and an AI could not be calculated. How this type of situation was handled in regards to the data presented should be addressed. It is important because it could influence the analysis done looking at individual assay performance results with time. Page 7. If samples were included that were in fact negative it would appear the slope over time would decrease when in fact it does not. This could also affect the number of false recent seroconverters calculated for the Avidity assay.  
| Answer: There were 6 individual time points with untreated and treated avidity OD values lower than the avidity OD threshold. These individuals were considered as recently infected for the avidity test. The results were recalculated to take into account these 6 individual time points. |
| R1A.4 | To assess the ability to use incidence testing from cross-sectional data to evaluate an intervention does require that the authors demonstrate the accuracy of the assay. Sensitivity and specificity were calculated. In addition, FPR estimates were computed from longitudinal data. Though the derivation provided in the annex is useful, it was not necessary to provide estimates for crude and adjusted incidence rate ratios. The paper would be better served if it focused the attention on the adjusted rate ratios computed from the two assays and the dual combination using the cutoff values that result in an estimated window period which compares as precisely as possible to the follow-up duration used in the original manuscript of 21 months. The presentation of estimates at multiple cutoff values, window periods, and with or without adjustment detracts from presentation of the methods and results related to the well-posed objective of comparing the two methods to evaluate the MC intervention as a way of determining whether cross-sectional data and incidence testing can be used for this purpose.  
| Answer: There is currently a debate on the benefit of correcting results given by HIV incidence assays and how to calculate such corrections. We think that it can be useful for those willing to use these HIV incidence assays to provide a comparison between the corrected and uncorrected results. There is also a debate on which cut-off values to chose. This is why we tested several values. Finally, we compared the HIV incidence rate ratio (IRR) calculated using survival analysis and the time period between V3 and V9 (about 9 months) with the IRR calculated using the two HIV incidence assays and a window period of 9 months. |
| R1A.5 | Based upon the PLOS Medicine manuscript published by Auvert et al in 2005, there were 3059 HIV-, 69 recent HIV+, and 146 prevalent HIV+ participants, representing the total 3274 that were randomized to the MC intervention trial. A subset of these totals was used in this study from |
subjects who completed the 21 month follow-up as identified in the annex. Comparisons of prevention effect should be made based upon this subset in which the efficacy was 62% instead of 60%. The original paper reports that “When considering only those participants who completed their M21 visit, the RR was 0.38 (0.22-0.67)”.

Answer:

As indicated in the paper we have added additional last follow-up visit data to the original data set and consequently obtained a new data set. This was described in another publication (Auvert et al., J Acquir Immune Defic Syndr 2010;53(1):111-6). The effect was 60% (34% to 76%). Before adding these new data, the effect was almost the same: 60% (32% to 76%).

As described in the annex, we also calculated the IRR using a) survival analysis and b) only the HIV testing results obtained at V3 and V4 and c) the new data set. This analysis gave an effect of 61% (21% to 83%). This is the value of the effect that has to be compared with the IRR given by the two HIV incidence assays.

R1A.6: The mean follow-up duration of subjects used in estimation of the intervention effect from the conventional longitudinal analysis was 21 months. Though it would be ideal to choose cutoff values that best distinguish recent from long-standing infection, for comparative purposes and to avoid use of the window period, cutoff values should be chosen to equalize follow-up and the window periods. This can be provided as the rationale for choice of cutoff values that result in an estimated window period of 21 months. Alternately, methods could be employed to annualize incidence or standardize to a common time period; or strong assumptions about the constancy of incidence and incidence ratios could be made.

Answer:

We wanted to explore the possibility of using these HIV incidence assays in a wide range of cut-off values. This is important to guide future researchers willing to use these assays to estimate the effect of an intervention using cross sectional data.

Nevertheless, as stated in our response to R1A.4, we compared the HIV incidence rate ratio (IRR) calculated using survival analysis and the time period between V3 and V9 (about 9 months) with the IRR calculated using the two HIV incidence assays and a window period of 9 months.

R1A.7: To verify how accurately these assays predict incidence, they must be validated against an independent measure of incidence. Two or more assays for recent infection can be compared to each other, though limited by the fact that none may be considered gold standards. The ultimate validation of these assays requires testing of cross-sectional specimens obtained from cohorts for which there is conventional measurement of observed incidence. Agreement between two incidence assays using the Kappa statistic, in which neither assay can be considered a gold standard, is not informative. Agreement between each assay and true incidence is preferable. Compare the test results for each incidence assay and for the combined assay to the known status of those specimens. Report the Kappa statistic for agreement in these comparisons.

Answer:

We reported the kappa statistic of the comparison of BED and AI results. We have now added the comparison of the results given by each of these two assays with the known HIV status of the individual.

R1A.8: Provide one table of the results for BED, AI, and BED/AI. Show the cutoff value that results in a window period of 21 months, sensitivity and specificity, Kappa, each of the parameters use in the corrected IRR (NTR, N+, #2, and N-), and the corrected IRR with bootstrap confidence limits.

Answer:

Unfortunately it is not possible to use a cut-off value of 21 months for AI and BED/AI. For BED, the cut-off value corresponding to an 18 month window
period is usually judged as the maximum value it is possible to use. As indicated above in our answer to R1A.7, we have now added the comparison of the results given by the 2 assays with those given by BED and AI for a window period of 9 months, which corresponds with the duration of time between V3 and V4. We have also provided all the details in Table 3 (sensitivity, specificity, number of individuals tested positive...)

R1A.9: The fact that variance of the bioassay measurements increased over time suggests that a data transformation may be required if parametric methods are used to explore the relationship of these measures with time since seroconversion. How is this relevant to the methods used in this manuscript?

Answer:
We didn't calculate any results that need to take this fact into account.

R1A.10: Ideally, the window period estimates and the false recent parameter should be calculated from independent data that are representative of the study population. If the same data were used to calculate the correction as was used for the estimation of incidence ratios, the estimates are biased towards a better outcome. This limitation should be discussed.

Answer:
As shown in this paper, the window period and the false recent parameter were not used to calculate the uncorrected HIV incidence rate ratio (IRR). We did use the window period to calculate the corrected IRR (i.e. to estimate the coefficient, A) but this was only a corrective factor and it is not strongly dependent on the value of the window period. This point is now mentioned in the discussion as a limitation of the study. We did use the false recent parameter to estimate the corrected IRR. The participants used to estimate this false recent parameter were also used to estimate the corrected IRR. This point is now also mentioned in the discussion as a limitation of the study.
Reviewer #1: Minor Essential Revisions

R1B.1: Page 2 In the abstract the complete name of the Bio-Rad assay should be used as there are multiple EIA’s produced by Bio-Rad.
Answer: The abstract has been modified accordingly.

R1B.2: In results page 6- it is suggested that the results reported for the incidence assay be recent and not recent and not negative and positive. Negative and positive are confusing terms in that Negative indicates long-term and positive indicates recent infection.
Answer: We changed "positive" to "recent" and "negative" to "not recent".

R1B.3: In the discussion on page 10/11 it is assumed that increasing prevalence effects misclassification the same way in BED and the Avidity assay described. There are not sufficient data reported for this Avidity assay to make this generalization.
Answer: We have changed the sentence.

R1B.4: Page 11- It should be noted that the avidity assay used in this study is not that same as that described in reference 13 so this could account for the only moderate agreement between BED and the Avidity test described in the manuscript.
Answer: We now provide a reference corresponding to the assay used in this study.

R1B.5: The abstract does not stand on its own. Provide the numbers of specimens collected at last follow-up and how they were used. Focus on the best estimates for the protective effect, instead of presenting crude and adjusted rates. State the best algorithm, estimates for the parameters, and provide the best estimate possible of efficacy. "Calculation of the effect was independent of the value of a window period.” This sentence should be included in the methods section.
Answer: We have updated the abstract. The sentence "The uncorrected IRR is independent of the window period." has been added to the methods section.

R1B.6: Describe which subjects have the additional 596 follow-up observations.
Answer: We refer to the publication where they are described.

R1B.7: The term window period or recency period should be defined up front. Prefer the wording “estimated” instead of “crude” window period.
Answer: We have added a definition of the window period. We have replaced the word "crude" by "uncorrected".

R1B.8: Refer to the annex in the text for explanation of numbers of specimens and how they were used, and for methods related to the window period.
Answer: This change has been made.

R1B.9: Page 5, the proportion of false long-term seroconverters should have been calculated from known recent seroconverters, whereas false recent seroconverters are calculated from known long-term seroconverters. Since the false long-term rate is not adjusted for in the IRR algorithm, this method and result is of less interest.
Answer: We estimated the proportion of false long-term seroconverters as the proportion of those with a value given by BED, AI or BED-AI higher than the cut-off (i.e. tested not recent) among those who became HIV-positive during a time interval between the last HIV-negative test and the first HIV-
positive test shorter than the uncorrected window period. Thus, we estimated this proportion of false long-term seroconverters from known recent seroconverters.

We estimated the proportion of false recent seroconverters among those with long-term infection as the proportion of those with a value given by BED, AI or BED-AI lower or equal to the cut-off (i.e. tested recent) among those who became HIV-positive during a time interval between the last HIV-negative test and the first HIV-positive test greater than twice the uncorrected window period. Thus, we estimated the proportion of false long-term seroconverters from known recent long-term seroconverters.

We have indicated the reference describing how to correct the number tested recent (Welte et al. AIDS Res Hum Retroviruses 2009;25(1):125-6). We have followed this referenced method with a minor modification to take into account the prevalence by age in our sample (see Annex).

<table>
<thead>
<tr>
<th>R1B.10: Are both methods subject to bias related to generalizability? Is there a cohort effect?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer: We have considered a population of young men. The men in this population are characterised by low HIV prevalence and high HIV incidence. This is discussed as a limitation of our study. We think that, despite this limitation, our results and experience will be useful to other researchers.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R1B.11: If incidence testing algorithms, avidity vs. BED vs. AI/BED, are to be compared with respect to their ability to accurately estimate the prevention effect, this should be stated in the introduction as an objective.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer: We cannot formally compare the results given by the incidence assays because of the lack of power as seen by the wide confidence interval obtained for the estimation of the effect. The objective was to explore the possibility of using these incidence assays to estimate the effect of an intervention using cross sectional data. We agree that more studies are needed to compare these assays. We now mention this point in the discussion.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R1B.12: Comment on the comparability of the window period to the follow-up period as it relates to the comparability of the estimates of efficacy between the two methods.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer: As indicated in R1A.7, we compared the HIV incidence rate ratio (IRR) calculated using survival analysis and the time period between V3 and V9 (about 9 months) with the IRR calculated using the two HIV incidence assays and a window period of 9 months.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R1B.13: It should be made clear in the methods (as is mentioned in the discussion) that HIV-infected individuals were also included in the cohort.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer: This has been made clear.</td>
</tr>
</tbody>
</table>
**Reviewer #1: Discretionary Revisions**

<table>
<thead>
<tr>
<th>R1C.1: Vernacular such as “censored” on page 4 of the methods such be written for understanding by a broader audience.</th>
</tr>
</thead>
</table>
| **Answer:**  
We now used the word "locked". |

<table>
<thead>
<tr>
<th>R1C.2: Can the avidity index reach values higher than the maximum 100% (page 4 methods)?</th>
</tr>
</thead>
</table>
| **Answer:**  
An avidity index is calculated by dividing two numbers. It may happen that these numbers lead to an index >1. Such an index does not have a clear interpretation. These index >1 are now reported as 1 as said earlier in RA1.2. |

<table>
<thead>
<tr>
<th>R1C.3: It is not necessary to provide the crude IRR formula in the main text.</th>
</tr>
</thead>
</table>
| **Answer:**  
We think that it is important to show that this formula is very simple and that it does not depend on the window period term. |
Reviewer #2: Major Compulsory Revisions

R2A.1: Use of incidence assays to estimate the relative risk in two arms of a randomized trial is a unique use of these assays. The RCT design addresses both recognized and unrecognized sources of bias by randomizing and comparing between the study arms. Using incidence assays in this scenario, misclassification due to the test characteristics (i.e. the incidence assays) should be randomly distributed between the two arms. As a result, a risk ratio defined by using the assays in both arms of the trial provides a reasonable approximation of the effect of the intervention. Similarly, as the authors point out, it is not necessary to know the window period precisely when the assays are used to estimate the RCT effect.

While use of the assays in an RCT provides a reasonable estimate of the risk ratio that would be obtained by longitudinal survival analysis, the same would not necessarily be true in an intervention program where there is no randomization. For example, if the tests were compared before versus after an intervention (e.g. circumcision), changes in factors such as the incidence to prevalence ratio, the ratio of different HIV subtypes, and the use of ART might result in different operating characteristics for the incidence assays at different time points. Similarly, if men who chose circumcision were compared to men who did not choose circumcision at a single time-point, differences could exist in the incidence to prevalence ratios, ages, etc. of the populations.

If the authors feel that this type of testing could be used for a cross-sectional study of a single-arm intervention, more explanation of how they feel that the RCT results can be generalized to non-RCT settings is essential, as it is currently not clear. Alternatively, if the intent is to say that this would be an intervention that would be useful in RCT settings, this needs to be made clearer in both the title and the manuscript. Statements like, "Such results imply that HIV interventions may be assessed using HIV incidence assays on samples obtained from a cross-sectional survey by calculating incidence rate-ratios (Discussion – 1st paragraph)," may otherwise mislead readers because the text does not clarify that the conclusion only applies in RCT settings.

Answer:
There are currently several ongoing studies to assess the effect of male circumcision on HIV incidence in community situations and not in the context of RCT. To assess the effect of MC on HIV incidence in such situations requires the ability to estimate HIV incidence among various groups, including circumcised and uncircumcised men. This can be done by following a cohort of people, with the resultant problem of bias, due to loss of follow-up and selection of participants. There is also the possibility of estimating HIV incidence from cross-sectional random samples with the problem of the accuracy of this estimation.

This study is the first to explore this last possibility with the unique opportunity to be able to compare the results with those obtained by following people over time. Following this comment from the reviewer we have modified the discussion to make this clear.

R2A.2: It is curious that overall, correction for misclassification has produced results that are, for the most part, farther from the “true” estimate provided by longitudinal analysis. This should be addressed in the discussion.

Answer:
We have recalculated this correction using a formula taking into account the
young age of the participants. As a result the corrected values are slightly higher than the uncorrected effect but the increase is now reasonable.

R2A.3: On page 5, second paragraph, the paper states, “We calculated the proportion of false long-term seroconverters and the proportion of false recent seroconverters among those with long term infection.” The former category is difficult to interpret; who would be classified as a false long-term seroconverter among those whose true status was long-term HIV-infected?

Answer:
We now made this clear in the annex where we state: We estimated the proportion of false long-term seroconverters as the proportion of those with a value given by BED, AI or BED-AI higher than the cut-off (i.e. tested not recent) among those who became HIV-positive during a time interval between the last HIV-negative test and the first HIV-positive test shorter than the uncorrected window period.
**Reviewer #2: Minor Essential Revisions**

| R2B.1: | The abstract talks about the correction for misclassifications, but gives no explanation of what is involved. This makes it difficult to interpret the findings that are presented in the abstract. |
| Answer: | We have corrected the abstract on this point. |
| R2B.2: | Table 3 “NE: Not Calculable” should be “NC: not calculable” |
| Answer: | We have made this change. |

**Reviewer #2: Discretionary Revisions**

| R2C.1: | The authors point out that, in contrast to another study that they cite, this analysis found only moderate agreement between the two assays. It would be helpful to also comment on the possible reasons for this. The methods such be written for understanding by a broader audience. |
| Answer: | We have recalculated the results to take into account the 6 individual time points with untreated and treated avidity OD values lower than the avidity OD threshold. As a result the agreement is now "good". |