Reviewer's report

**Title:** Estimating the regional distribution of men who have sex with men (MSM) based on Internet surveys and calculation of MSM population-specific incidence of sexually transmitted infections

**Version:** 1  **Date:** 25 September 2008

**Reviewer:** Françoise Dubois-Arber

**Reviewer’s report:**

Review of the article « Estimating the regional distribution of men who have sex with men (MSM) based in Internet surveys and calculation of MSM population-specific incidence of sexually transmitted infections”, by U. Marcus et al.

In this article the authors explore the use of proportional regional distribution of participants of large internet-based surveys among MSM from Germany to estimate the regional distribution of MSM in the German general population, to calculate population- specific incidence rates of HIV and syphilis.

This type of exercise is a complex one, which includes many assumptions, and there are only few examples of it in the literature. This type of work is therefore necessary and important.

The article is very long, with many subchapters that would deserve a better organisation. One may suggest to split this article in two different articles. It seems to me that the exploration of the validity of using internet-based surveys to calculate sizes of MSM regional population – with the comparison of estimated HIV prevalences by two methods - is the main subject of this article. The use of the results of this exploration may be the subject of another article.

Overall, in spite of the great interest of this exercise, there are many assumptions, estimations, adjustments which are not clearly detailed and the consequences of which are not appreciated. In this situation, it is difficult to completely appreciate the validity of the data presented. Some tables are also not easy to read with imprecise headings (see below).

**Major revisions needed**

1. The methods should include, at the beginning, a definition of the steps that have been undertaken to attain the objectives. A figure with a summary of these steps, together with the assumptions made during the process may be useful.

2. The methods should be exposed more in depth, by defining, at each step, the rationale for the comparisons undertaken (for example the rationale for and the consequences of the comparison of the three samples), the precise assumptions made, the detailed account of adjustments made, the detailed calculation of estimates, including confidence intervals where it is relevant (for example, in the
estimation of the total MSM population in Germany, and consequently, on the regional prevalence estimates).

3. The diverse internet sites included in the studies should be mentioned with their characteristics, as well as the contribution of the diverse sites to the surveys.

4. The results should be presented more clearly, step by step. A table with the main characteristics of the three survey samples should be provided.

5. As the main outcome of the comparison seems to be the relevance of using online KaBasTi as a reasonably representative sample of German MSM this choice should be validated.

6. The headings in tables 2 and 3 are not precise, as well as methods of calculations. These two tables should be revised.

7. The similarities declared for the figures 2 and 3 should be analysed more in depth.

8. Where it is appropriate (in relation to survey data), confidence intervals should be mentioned.

9. Sensitivity analysis on the estimates should be performed. In summary, we should get an evaluation of the cumulated effect of the different assumptions, adjustments, etc., on the precision of the numbers of MSM in the regions, and on the prevalence and incidence estimates. Current differences between regions may change.

10. The discussion should take into account these changes, and this may lead to substantial modifications.

Minor revisions needed

1. The objectives are not clearly described and should be more explicit (in the abstract, they appear more clearly).

2. In the introduction, some statements such as the last sentence in the last paragraph, should be sustained with references to literature.

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests