Author's response to reviews

Title: Estimating the size of the MSM populations for 38 European countries by calculating the survey-surveillance discrepancies (SSD) between self-reported new HIV diagnoses from the European MSM Internet Survey (EMIS) and surveillance-reported HIV diagnoses among MSM in 2009

Authors:

Ulrich Marcus (MarcusU@rki.de)
Ford Hickson (ford.hickson@emis-project.eu)
Peter Weatherburn (peter.weatherburn@emis-project.eu)
Axel J Schmidt (axel.j.schmidt@emis-project.eu)

Version: 5 Date: 26 June 2013

Author's response to reviews: see over
Response to the reviewer

We thank the reviewer for helpful comments and suggestions. We tried our best to take all of them into account when revising the manuscript.

General feedback:
This manuscript presents a methodological approach to estimate the size of the MSM population in 38 countries by using two large international datasets on HIV infected persons. A mathematical formula was developed based on comparison of notification rates from two sources: the internet survey EMIS and the European surveillance database. This approach focuses on estimation of denominators data in a specific population. However, issues related to data quality are mentioned regarding numerators data as well.

To address data quality issues (especially completeness), could the authors refer to or explore the possibility to use a capture-recapture approach which could be complementary to the calculations presented in this article?
Response: To use capture-recapture approaches to address data quality issues in surveillance data would be quite challenging in most European countries due to data protection regulations, and because HIV surveillance data are in many countries not collected by named reporting. Also the internet survey data are completely anonymous.

The length and the style of the manuscript rather correspond to a report but an article. The output of this work could probably be beneficial to the reader if the authors could summarise and avoid repetitions of the information.
Response: We are well aware of the unusual length of the manuscript. We tried to avoid repetitions as far as possible when revising the manuscript. However, discussing data from 38 countries just requires some space.

Major compulsory revisions:
1. Figure 1: would require some clarity. All these intermediate information might not be needed to be displayed:
Response: Figure 1 has been deleted, the information has been transferred to table 1.

- It is not clear how the following information is obtained? “on the HIVpop required to achieve at least 1% or at most 3% of MSM at given SSD?”
Response: This is described in the Methods section on page 10/11: “Estimating a plausible range for the number of MSM newly diagnosed with HIV in 2009 for countries with missing or unreliable surveillance system-derived data: For countries reporting zero HIV diagnoses among MSM in 2009 (Estonia) in their surveillance system, countries not reporting HIV diagnoses among MSM at all (Austria), and countries reporting implausibly low numbers of MSM (resulting in MSM population estimates well below 1% of the male adult population: Bulgaria, Belarus, Romania, Russia, Turkey, and Ukraine), we estimated the minimum number of MSM expected to have been diagnosed with HIV in 2009 for an assumed MSM population size of at least 1% (resp. 2% for Estonia – comparable with Latvia and Lithuania - and 3% for Austria – comparable with Germany and Switzerland) given the SSD estimated on the proportion of households with internet access (see Table 1, column I).”

- Where does this formula come from ?SSD = 1.67 x household internet access /-0.6
Response: The formula comes from the empirical data from consecutive internet surveys conducted in several European countries between 2003 and 2010. This is described on page 10: “To test this hypothesis, we looked for data from repeated MSM internet surveys from the
same countries to calculate the respective SSD values. If the surveys included different time-points with increasing levels of household internet access, and if we assume that over the time period in which these surveys were conducted the size of the MSM population didn’t change substantially, an SSD can be calculated for the respective consecutive surveys [see Additional file 1].

When different SSD values were calculated at the same household internet access level, the median was determined. Obvious outlier values are likely explained by targeted offline promotion of internet surveys (SSD values increased) or mis-interpretable wording of respective questions (values too low) and were disregarded. The mathematical function which best described the observed correlation between SSD and household internet access from the countries with repeated surveys \[SSD=1.67\times[\text{national household internet access}]^{0.6}\] - see Additional file 1 and Figure 2) was then used to calculate an SSD for all 38 EMIS countries. The resulting SSD was further used to estimate the total MSM population size for all 38 countries \[N_{pop}=N_{HIVpop}\times N_{svy}\times SSD/HIV_{svy}\].

Response: see above

o Percent of the adult male population represented by \(N_{pop} (=M)\)
Response: This is the proportion of the calculated value of \(N_{pop}\) (see above) from \(N_{tot}\) (data from population statistics)

2. Sentence: “According to information provided by the respective national contact points the data reported to ECDC or published elsewhere were adjusted for reporting delays, redistribution of cases with unknown transmission risk, and incomplete national coverage of surveillance data (HIV data reported from Spain and Italy do currently not cover the whole country)”

Comment: Data are adjusted for reporting delay, and this information is presented by transmission group. However information on unknown transmission risk and incomplete national coverage are taken into account for the calculation of the notification rate, but not as an adjustment.
Response: Thanks for this comment. We have revised the sentence accordingly: “According to information provided by the respective national contact points the data reported to ECDC or published elsewhere were adjusted for reporting delays. Cases with unknown transmission risk were proportionally redistributed to transmission groups, and incomplete national coverage of surveillance data (HIV data reported from Spain and Italy did not cover the whole country in 2009) was taken into account for the calculation of the notification rates.”

3. Sentence in the discussion: “However, after adjustment for risk redistribution of cases with unknown transmission risk, in Poland and Russia the ratio dropped to a level comparable with other countries. This strongly suggests that surveillance data are unreliable in terms of transmission risk categorization (or number of reported cases) also in Bulgaria, Belarus, Romania, Turkey, and Ukraine”

Comment: This is a strong statement that would require stronger evidences.
Response: In Poland, the proportion of unknown transmission risk is around 80%. In Russia (unknown transmission risk around 50%), Belarus, Ukraine and Turkey, same sex behaviors are only queried in a small proportion of newly diagnosed cases in voluntary counseling and testing sites. However, most HIV diagnoses are made through routine mass testing in hospitals and other settings where no detailed informations on transmission risk are collected. While there are efforts to distinguish between drug- and sex-related transmission risks, no further efforts are made to distinguish between heterosexual and homosexual
transmission. MSM are therefore mostly subsumed in the heterosexual or unknown risk categories. See also newly added references 21 and 22.

In Romania, the number of EMIS participants reporting to have been diagnosed with HIV in 2009 was 19, which is 2.7-fold higher than the total number of MSM reported for the whole country in 2009 (n=7), although the EMIS sample size was only 2338. The total adult male population of Romania aged 15-64 is around 10.5 Million. Thus it is quite obvious that something must be wrong about the official surveillance data. One possible explanation is underreporting of HIV cases from the private sector.

The author must respond to these before a decision on publication can be reached. For example, additional necessary experiments or controls, statistical mistakes, errors in interpretation.

The authors comment on the following parameters that can explain a SSD above or below 1. Could this be illustrated in a figure (and rather less text)?

SSD>1
- MSM survey oversamples MSM
- Undercounting of MSM in the surveillance system
- Overestimation of MSM in the country

SSD<1
- MSM survey undersamples MSM
- Overcounting of MSM in the surveillance system
- Under-estimation of MSM in the country

Response: We created a new Figure (Figure 1) to illustrate the possible reasons for different SSD values.

Survey conducted from 2003-2010: this represents a big time lag. What is the common way of administration? Could the method from the survey be summarised?

Response: Maybe there was a misunderstanding. There were different consecutive surveys conducted in several European countries between 2003 and 2010. Common for all surveys was online recruitment of participants. All surveys were self-administered and anonymous. We used these surveys to determine the relationship between household internet access rates and SSD.

Sentence “For Italy and Spain, the most likely explanation may be an insufficiently adjusted number of newly diagnosed infections among MSM. In both countries the reporting system does not cover the whole country and misses some regions (although we explicitly tried to take this into account when adjusting the national surveillance data). In addition, transmission risk group assignment may underestimate MSM and overestimate heterosexual cases.”

Comment: could the authors clarify this statement?

Response: The sentence has been revised into: For Italy and Spain, the most likely explanation may be errors in the calculation of notification rates of newly diagnosed infections among MSM. In both countries the reporting system does not cover the whole country and misses some regions (although we explicitly tried to take this into account when calculating adjusting the national notification rates). In addition, transmission risk group assignment may underestimate MSM and overestimate heterosexual cases due to physicians not asking about and patients not reporting sexual preferences. While this is true for all countries, it may be truer for countries in which same sex behaviours are more stigmatized. According to data collected in EMIS [8] this could be the case for Italy, but less so in Spain.

Minor essential revisions:
1. Could the authors specify who are the 6 partners and 77 collaborating partners (35 countries)?
   Response: Associated and Collaborating partners are listed as EMIS network at the end of the manuscript.

2. Sentence: HIV diagnoses among MSM reported in the national systems (data not shown).
   Comment: Data are not show, but could references be provided?
   Response: A reference to the ECDC HIV surveillance report has been added.

3. Sentence: “The most likely explanation for the findings in all these countries - except Macedonia - is the underreporting of MSM cases in the national surveillance systems.”
   Comment: Could this be supported by references?
   Response: Two references for underreporting of MSM in Eastern European surveillance data have been included: