PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Estimating the HIV undiagnosed population in Catalonia, Spain: Descriptive and comparative data analysis to identify differences in MSM stratified by migrant and Spanish-born population.
AUTHORS	Reyes-Urueña, Juliana; Campbell, Colin; Vives, Nuria; Esteve, Anna; AMBROSIONI, Juan; Tural, Cristina; Ferrer, Elena; Navarro, Gemma; Force, Luis; García, Isabel; MASABEU, Àngels; Vilaró, Josep; Garcia de Olalla, Patricia; Caylà, Joan; Miro, Josep; Casabona, Jordi

VERSION 1 – REVIEW

REVIEWER	Dr Jesus M Garcia Calleja WHO Switzerland
REVIEW RETURNED	18-Jul-2017

GENERAL COMMENTS	The paper uses a ECDC tool develop to estimate the number of HIV infections and the under diagnosed proportion. The paper should refer to "estimates" and not HIV cases in the results and use CL or uncertainty ranges in their figures. For instance in results on page 7, the first section talks about HIV diagnosis but in the second paragraph all the data is estimates from the ECDC tool. The estimated data is fro 2013, and yet we are in 2017, why not to produce estimates proportion for 2016 ? or at least 2015 ?, it is feasible ??. Geographical distribution : is possible to add some geographical distribution to the new HIV cases and estimated underdiagnosed ? There are two factors that the authors do not consider or talk about it: 1) the fact that ART and other care services are free in Catalonia, may have an effect of attraction for migrant populations, specially from LMIC. 2) Barcelona is well know for the European gay week and all the cultural events and gathering or large numbers of MSM in the city. Can be possible to add some information on this and the possible factor for new infections in cat ?

REVIEWER	Alessia Mammone Department of Epidemiology, National Institute for Infectious Diseases 'Lazzaro Spallanzani', Rome, Italy

REVIEW RETURNED	21-Jul-2017

GENERAL COMMENTS	Reviewer's report In their paper Reyes-Uruena et al. describe a method to evaluate the number of HIV undiagnosed along with a tentative estimate of the continuum of care in Catalunia, using a combination of data sources. The paper is well written in general, but would profit from some revisions and clarifications. A. Major Compulsory Revisions In general, the part regarding the cascade of care is not really focused; it seems that the authors are uncertain about the opportunity to add this point. The result is that this part about the cascade of care estimation is not reported in the objective section of the abstract (for example as a side objective); it is briefly reported in the method section of the abstract, but it is not well explained in the methods section of the paper. The authors should decide if they want to address this estimation; if the answer is yes, (and in my opinion, it is worthy) the manuscript should be entirely revised, by adding some clarifications; if the authors decide to not consider the cascade, every reference to it should be removed from the manuscript. Methods section Page 5 Row 30: a concurrent AIDS/HIV diagnosis is defined as within 6 weeks? Page 5 Row 39: the sentence is not clear to me: what does it means "...by MSM that are under active follow up on the cohort"; please explain; Page 6 row 13: "we considered 5 distinct...", but then you list just 4 periods; please add some comments about the last period you list "2002-2013"; Page 6 row 37 and on: this part should be explained better; please add a brief description of the cohort, number of persons enrolled, starting year of the cohort; Also the representativeness of the cohort with respect to surveillance data should be reported (i.e. proportion of M and F, risk factor, immigration status). Discussion section Page 10 row 50 and on: the sentence "While limited.....as we could for HIV testing", it is not very clear to me; please rephrase the sentence; Page 11 row 26: also self test should be discussed as a possible strategy to reduce the fraction of undiagnosed; Page 18: it would be useful to add in the table the number of new diagnoses reported in the same years to the surveillance system (stratified as well as the estimated incidence in spanish MSM and migrants MSM); Page 21: in the caption of figure 3 is reported several times VIH instead of HIV; Page 28: add a panel with the general Cascade of care and then report the cascade by immigration status and by risk factor; Page 27, Table 1 of Supplement material please verify the sum

	about sex; it would be useful to present in the same table the same characteristics for the PISCIS cohort and the same statistics; General data regarding the Surveillance System, i.e. number and characteristics of new diagnoses reported to surveillance during the study period, stage of disease (cd4 level, that is strictly related to ART initiation), that could help in results' interpretation, should be reported.

REVIEWER	Ruiguang Song Centers for Disease Control and Prevention USA
REVIEW RETURNED	28-Jul-2017

GENERAL COMMENTS	Major comments: The incidence estimates shown in Figures are not consistent. In Figure 1, the total incidence was increasing after 2010, while the incidences in its two subgroups (Spanish born-population and migrants) were decreasing after 2010. This could happen when the total is derived from a separate model, not from the sum of subgroup results. Also, the results for the total could depend on the variable (sex, age, or risk) used to form subgroups. How to make results consistent across all population subgroups needs to be discussed. In the abstract and Figure 4(A), it says that there were 8,458 (8,101, 9,079) Spanish-born MSM. However, in Figure 3(A), the number seems only around 6,000. Also in the abstract, it says the proportion of migrant MSM than local MSM was undiagnosed (32% vs. 16%), but based on the proportions shown in Figure 4, the one for local MSM is 22%, not 16%. Minor comments: There are two tables in this manuscript, but both are labeled as Table 1. The second one on page 27 could be labeled as Table A since it is considered as supplemental material. There are places using the word "VIH". Should it be HIV? Page 5 line 20, what is the meaning of late diagnosis, CD4 < 200 or 350? What is the advanced HIV, clinical symptoms only or including CD4 < 200? Page 6 line 13, should the number of distinct historical periods be four (nor five) because only four periods listed in the paragraph? Page 11 line 56, what's durable viral suppression? Is it the duration of vial suppression? Page 18, the title of the table indicates that the estimated average time from infection to HIV diagnosis is presented in this table, but it doesn't.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1. Dr Jesus M Garcia Calleja

The paper uses a ECDC tool develop to estimate the number of HIV infections and the undiagnosed proportion.

1. The paper should refer to "estimates" and not HIV cases in the results and use CL or uncertainty ranges in their figures.

Response: According to the reviewer' suggestion all the numbers into the results have been refereed as estimated numbers.
In the cases of confidence intervals, the tool determines 95% confidence intervals using a simple simulation method. These confidence intervals include two sources of uncertainty for the estimated number of people living with undiagnosed HIV: the stochastic uncertainty concerning the CD4 count-specific rate of symptoms and the stochastic uncertainty associated with the possibility that the observed number of HIV diagnoses with HIV-related symptoms may not correspond to the expected number based on the CD4 count-specific rate of symptoms. In the graphs these CIs have been included as a shadowed straps around the estimated numbers to show lower and upper bounds.

2. For instance in results on page 7, the first section talks about HIV diagnosis but in the second paragraph all the data is estimates from the ECDC tool.

Response: In the first paragraph from the results section, authors wanted to give an insight of the data that were introduced into the model, as these data were from the Catalan Surveillance system (CSS). These data referred to HIV diagnosis reported to the CSS.
The next paragraphs made reference to the estimated numbers calculated by the ECDC tool (estimated numbers).

3. The estimated data is fro 2013, and yet we are in 2017, why not to produce estimates proportion for 2016 ? or at least 2015 ?, it is feasible ??.

Response: The reviewer is right in point out the old dated data (2013), but the CSS has suffered some improvements in the number of information sources feeding the system, therefore it is still under assessment and analysis, as it is using multiple sources collected into one or more databases and there are some data need to be verified in deep. So, the more accurate data is from 2013. Currently is unfeasible to use more update data to estimate the numbers of total number of people living with HIV and undiagnosed in Catalonia.

4. Geographical distribution: is possible to add some geographical distribution to the new HIV cases and estimated underdiagnosed ?

Response: Unfortunately the geographical distribution of cases is not well collected into the CSS as it has a high rate of under reporting, thus, any assumption using this variable will be very imprecise and even more in the undiagnosed cases.

5. There are two factors that the authors do not consider or talk about: the fact that ART and other care services are free in Catalonia, may have an effect of attraction for migrant populations, specially from LMIC.

Response: This is a very important point to be considered. However, in previous studies, we saw that through the care continuum of care in Catalonia, a lower proportion of migrants are enrolled in HIV care, on ART and have viral suppression, as compared to the Spanish-born population. Our findings would suggest that in Catalonia, some of the differences between migrants from different regions (and

by extension between migrants and the Spanish-born population) might be explained by differential patterns of migration rather than only differential access, so still despite the free access to ART in Catalonia, migrants might experience barriers to access HIV treatment (Ref: Reyes-Urueña, J., Campbell, C., Hernando, C., Vives, N., Folch, C., Ferrer, L., ... & Casabona, J. (2017). Differences between migrants and Spanish-born population through the HIV care cascade, Catalonia: an analysis using multiple data sources. Epidemiology & Infection, 145(8), 1670-1681).

In the other hand, there are not studies published yet related to the call effect of free access to treatment on Latin Americans HIV positive patients, and from our CSS data what we have seen is that the epidemiological profile of migrants with a HIV positive diagnosis in 2013 was different to that of the Spanish-born population, there was a higher proportion of women (24.6% and 16.7%, respectively) and they were younger at diagnosis (median age of 33 vs. 37, respectively). This is in line with characteristics of Latin American (LAC) migrant population, as the demographic profile of LAC migrants in Catalonia reveals a young population with high rates of labour force participation, with relatively high levels of education and significant feminization. (Ref: Hierro, M. (2016). Latin American migration to Spain: Main reasons and future perspectives. International Migration, 54(1), 64-83). This might explained that the main reason to migrate to Catalonia is an economic reason rather that to get access to free ART.

Finally, raw analyses performed with the CSS data have showed that migrants, and especially LAC MSM appear at particular risk of HIV acquisition post-migration, as around 55% of migrants with a new HIV diagnoses notified to the Catalan surveillance notification system might have acquired HIV after their arrival to Spain, and among Latin Americans MSM is around 60%. This is also in line with recent studies performed in Europe, where among migrants from Latin America and the Caribbean, the rate of post-migration HIV acquisition was 71% in Spain. (Ref: Alvarez-del Arco, D., Ibidun, F., Thomadakis, C., Pantazis, N., Touloumi, G., Gennotte, A. F., ... & Boesecke, C. (2017). High levels of post-migration HIV acquisition within nine European countries. AIDS.)

Therefore, the authors believe that there is a small impact into the migration flow to Catalonia, specially among LMIC, due to the fact that ART and other care services are free of access, still an important point to be considered for future studies. Nevertheless, authors don't think that this point should be included as a point of discussion of the manuscript.

6. Barcelona is well know for the European gay week and all the cultural events and gathering or large numbers of MSM in the city. Can be possible to add some information on this and the possible factor for new infections in cat ?

Response: As the reviewer mentioned Barcelona is a well-known city for gay gatherings and events not just during the pride week but also during the whole year, however it is difficult to determine to what extent these events can impact the HIV local epidemic. This might be explained by different facts:

- It has been described that around 25% of MSM reported having sex while travelling abroad in the previous 12 months, is with a person not resident in the respondent's country of residence (Ref: European Men-Who-Have-Sex-With-Men Internet Survey (EMIS)). From the Catalan HIV surveillance system this is difficult to measure and follow, as the system does not collect data from tourist.

- Men over 25 years of age, those with a higher education, from large cities, or with HIV infection were more likely to report sex abroad Ref: European Men-Who-Have-Sex-With-Men Internet Survey (EMIS)). The high risk factors of men who assist gay events, have been associated to a high risk to acquire HIV, but not just in mass gathering gay events, but also in a general basis, as among these

MSM prevalence of HIV infection is higher, as well as, high sex-risky behaviours, when compared with other men.

- The most common way of meeting sexual partners abroad was through the internet (32%), gay saunas (14%) or gay nightclubs and discos (13%) Ref: European Men-Who-Have-Sex-With-Men Internet Survey (EMIS)). These ways to contact a gay partner are not just used during gay gathering events, but also in a regular basis in Barcelona during periods without mass gathering gays events.
- Finally, in Catalonia it has been estimated that the median time from infection to diagnosis is around four years (raw data), so it is quite difficult to determine where the new diagnosed cases have acquired the HIV infection in the past, this includes gay mass events.

Response: Having into account all these previous points, locally is difficult to measure the impact of gay events in the HIV Catalan epidemic, however, authors still consider and agree with the reviewer that these events might have an important impact not just in the increase of HIV infections, but also in the number of new STIs and in the increase number of HIV undiagnosed MSM congregate in these events. However, as related data have not been published yet and it has not been possible to find studies on this matter, we, authors believe that includes this into the discussion without any reference is not advisable.

Reviewer: 2. Alessia Mammone

In their paper Reyes-Uruena et al. describe a method to evaluate the number of HIV undiagnosed along with a tentative estimate of the continuum of care in Catalonia, using a combination of data sources. The paper is well written in general, but would profit from some revisions and clarifications.

A. Major Compulsory Revisions

In general, the part regarding the cascade of care is not really focused; it seems that the authors are uncertain about the opportunity to add this point.

1. The result is that this part about the cascade of care estimation is not reported in the objective section of the abstract (for example as a side objective); it is briefly reported in the method section of the abstract, but it is not well explained in the methods section of the paper. The authors should decide if they want to address this estimation; if the answer is yes, (and in my opinion, it is worthy) the manuscript should be entirely revised, by adding some clarifications; if the authors decide to not consider the cascade, every reference to it should be removed from the manuscript.

Response: As the reviewer mentioned the comparison between migrant MSM and Spanish-born MSM is of high relevance in our context, taking into account the currently epidemiological characteristics of the HIV in Catalonia, therefore it has been included this analysis into the objectives and methods of the manuscript.

Methods section

2. Page 5 Row 30: a concurrent AIDS/HIV diagnosis is defined as within 6 weeks?

Response: Yes, it has been defined a concurrent HIV/AIDS diagnosis, as an AIDS diagnosis within 6 weeks of HIV diagnosis.

3. Page 5 Row 39: the sentence is not clear to me: what does it means “...by MSM that are under active follow up on the cohort”; please explain;

Response: We agree with the reviewer that the sentence it was not clear. Therefore it has rephrased as it follows: “The first two steps of the HIV care cascade (PLWHIV and diagnosed and proportion of undiagnosed population) were derived from the estimates calculated with the ECDC tool, whereas the last three were derived from the PISCIS cohort. Each of these steps were analysed by region of origin and by MSM.”

4. Page 6 row 13: “we considered 5 distinct...”, but then you list just 4 periods; please add some comments about the last period you list “2002-2013”;

Response: Thanks to the reviewer to highlight this typo mistake, as in fact in the model we just included four periods instead of five. Also it has been clarified why we included the last period into the model. The phase that it has been included is as follows: (4) 2002-2013, the start of the HIV notification in Catalonia.

5. Page 6 row 37 and on: this part should be explained better; please add a brief description of the cohort, number of persons enrolled, starting year of the cohort;

Response: This has been explained in a previous paragraph; however, the starting year of the cohort has been added. A more detailed description can be found in the reference.

6. Also the representativeness of the cohort with respect to surveillance data should be reported (i.e. proportion of M and F, risk factor, immigration status).

Response: Unfortunately, despite the high cohort coverage in Catalonia (around 80% of the patients on ART in Catalonia are on followed up in the PISICS cohort), all the cases under follow up are anonymized, so it is not possible identify and linkage cases with the Catalan surveillance system. Therefore, it is impossible to measure the representativeness of the cohort with respect to surveillance data.

Discussion section
7. Page 10 row 50 and on: the sentence “While limited.....as we could for HIV testing”, it is not very clear to me; please rephrase the sentence;

Response: Following reviewer’s suggestion, the sentences has been rephrase as it follows:
Resources should be allocated primarily to promote testing in high-activity MSM, as they are the key group sustaining the epidemic, but also for encouraging all MSM including migrants MSM to test regularly for HIV.

8. Page 11 row 26: also self test should be discussed as a possible strategy to reduce the fraction of undiagnosed;

Response: Authors agree with the reviewer that self testing is a prevention strategy with great potential to reduce the number of people undiagnosed as well as the proportion of late diagnosis, however in Spain the sale of self-tests remains unauthorized.

Even a study has showed that less than 5% of Spanish MSM are aware of the availability on the internet of unauthorized and unreliable self-testing kits for HIV, and less than 1% had ever used these kits shortly after having been exposed to mass media information related to them (Ref: Koutentakis K, Rosales-Statkus ME, Hoyos J, et al. Knowledge and use of unauthorized HIV self-test kits among

men who have sex with men in Spain, following approval of an over-the-counter self-test in the U.S: a cross-sectional study. BMC Public Health. 2016;16:532. doi:10.1186/s12889-016-3204-9).
So, authors think that knowledge of HIV status and access to self-testing remains low in Spain, and given the fact that self-testing is yet to be authorized, there may also be an increase in the unregulated trade of illegal kits, with the adding problem that these online sales do not offer any kind of telephone or web support for interpretation of results, counselling, referral or linkage to services for care and support. Therefore, given the complexity of this intervention in Spain authors consider that this is out of the scope of the current manuscript to be included as isolated point into the discussion.

9. Page 18: it would be useful to add in the table the number of new diagnoses reported in the same years to the surveillance system (stratified as well as the estimated incidence in spanish MSM and migrants MSM);

Response: Following reviewer's suggestion, in table 1 it has been included the number of new HIV, HIV/AIDS and AIDS cases reported by year on MSM, according with their region of origin.

10. Page 21: in the caption of figure 3 is reported several times VIH instead of HIV;

Response: Thanks to reviewer's comment by spotting these typo mistakes. All the VIH have been changed into HIV through the figure legends.

11. Page 28: add a panel with the general Cascade of care and then report the cascade by immigration status and by risk factor;

Response: Authors consider that the reviewer's request is very appropriate, that is why we have addressed these points in different papers. For the general cascade from Catalonia the ref is the following: Campbell, C. N., Ambrosioni, J., Miro, J. M., Esteve, A., Casabona, J., Navarro, G., ... & Tural, C. (2015). The continuum of HIV care in Catalonia. AIDS care, 27(12), 1449-1454, and for the care cascade by region of origin the reference is the following (Ref: Reyes-Urueña, J., Campbell, C., Hernando, C., Vives, N., Folch, C., Ferrer, L., ... & Casabona, J. (2017). Differences between migrants and Spanish-born population through the HIV care cascade, Catalonia: an analysis using multiple data sources. Epidemiology & Infection, 145(8), 1670-1681). Therefore, add all these figures will add a lot of information that will be out of the scope of the current manuscript.

12. Page 27, Table 1 of Supplement material please verify the sum about sex; it would be useful to present in the same table the same characteristics for the PISCIS cohort and the same statistics;

Response: Following reviewer's suggestion, the sum of men and women in table 1 (supplement material), has been reviewed and changed accordingly, Also, it has been included a table with the PISCIS cohort information from 1998 to 2013, to give an insight of its epidemiological characteristics.

13. General data regarding the Surveillance System, i.e. number and characteristics of new diagnoses reported to surveillance during the study period, stage of disease (cd4 level, that is strictly related to ART initiation), that could help in results' interpretation, should be reported.

Response: Table 1 on the supplement material makes reference to the new HIV cases reported to the Catalan surveillance system, and includes number and epidemiological and clinical characteristics of these cases. Moreover, the first paragraph in the results section includes a brief description of the data from the surveillance system that were included into the model.

Reviewer: 3. Ruiguang Song

Major comments:

1. The incidence estimates shown in Figures are not consistent. In Figure 1, the total incidence was increasing after 2010, while the incidences in its two subgroups (Spanish born-population and migrants) were decreasing after 2010. Also, the results for the total could depend on the variable (sex, age, or risk) used to form subgroups. How to make results consistent across all population subgroups needs to be discussed.

Response: The reviewer is right in point out that the incidence among Spanish born and migrants shows in the graphs a decreasing trend, however the confidence intervals for these two populations are wider than the CIs for the whole Catalan population. These can be related to the method that calculates 95% confidence intervals on estimated parameters and model outcomes by doing a bootstrap analysis. As the model assumes that the data are distributed according to a certain probability distribution, in this case either a Poisson or a negative binomial distribution, with a mean defined by the best-fitting model, the tool generates a new dataset by sampling from this distribution for every year for each of the relevant data items and for each population. The model is then refitted to this new dataset starting from the parameter values found in the main fit. This procedure of sampling and refitting is repeated many times, so these can increase the grade of uncertainty among the analysis by sub-groups.

2. In the abstract and Figure 4(A), it says that there were 8,458 (8,101, 9,079) Spanish-born MSM. However, in Figure 3(A), the number seems only around 6,000. Also in the abstract, it says the proportion of migrant MSM than local MSM was undiagnosed (32% vs. 16%), but based on the proportions shown in Figure 4, the one for local MSM is 22%, not 16%.

Response: Thanks to the reviewer for highlighting these mistakes. All the changes have been included and modifications have been done accordingly.

Minor comments:

3. There are two tables in this manuscript, but both are labeled as Table 1. The second one on page 27 could be labeled as Table A since it is considered as supplemental material.

Response: Following reviewer's suggestion, tables on the supplement material have been renamed as Table A and Table B.

4. There are places using the word "VIH". Should it be HIV?

Response: Thanks to reviewer's comment by spotting these typo mistakes. All the VIH have been changed into HIV through the figure legends.

5. Page 5 line 20, what is the meaning of late diagnosis, CD4 < 200 or 350? What is the advanced HIV, clinical symptoms only or including CD4 < 200?

Response: Following reviewer's comments it has been included the definition used for late diagnosis and advanced HIV infection, as it follows: Late diagnosis and advanced HIV infection were defined as a CD4 cell count below 350 cells/ml or AIDS and CD4 cell count below 200 cells/ml or AIDS at the time of HIV diagnosis, respectively.

6. Page 6 line 13, should the number of distinct historical periods be four (nor five) because only four periods listed in the paragraph?

Response: Thanks to the reviewer to highlight this typo mistake as in fact in the model we just included four periods instead of five.

7. Page 11 line 56, what's durable viral suppression? Is it the duration of vial suppression?

Response: We have assumed durable viral suppression as VL values <200 copies/mL over the 2-year period.

8. Page 18, the title of the table indicates that the estimated average time from infection to HIV diagnosis is presented in this table, but it doesn't.

Response: Thanks to the reviewer to spot this mistake, title and even the content of the table 1 have been modified accordingly.

VERSION 2 – REVIEW

REVIEWER	Jessu| M Garcia Calleja WHO Switzerland
REVIEW RETURNED	26-Sep-2017

GENERAL COMMENTS	this is the second review, changes and suggestion have been accepted.

REVIEWER	Dr Alessia Mammone National Institute for Infectious Diseases IRCCS "L. Spallanzani" Via Portuense, 292 00149 Roma, Italy
REVIEW RETURNED	04-Jan-2018

GENERAL COMMENTS	Reviewer's report In their paper Reyes-Uruena et al. describe a method to evaluate the number of HIV undiagnosed along with a tentative estimate of the continuum of care in Catalunia, using a combination of data sources. The paper improved a lot with respect to the previous version. A. Minor Revisions Page 4 last paragraph: also add something about the importance of the cascade of care; Page 30, table B: please correct Chi-Quadrado

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2. Alessia Mammone

Minor Revisions

1. Page 4 last paragraph: also add something about the importance of the cascade of care.

Response: A phrase, which refers to the importance of the cascade of care, was added, following reviewer suggestion.

2. Page 30, table B: please correct Chi-Quadrado

Response: We would like to thank to the review to point out this mistake, which was corrected in the new version of the manuscript.