Estimating the number of people living with HIV and the undiagnosed fraction in Spain in 2013

Nuñez Olivier, Hernando Victoria, Díaz Asunción

HIV Surveillance and Behavioural Monitoring Unit. National Center of Epidemiology, Carlos III Institute of Health, 28029 Madrid.

Corresponding Author:

Olivier Nuñez

National Center of Epidemiology

Instituto de Salud Carlos III

Monforte de Lemos 5 28029 Madrid Phone: +(34) 91 822 2697

Fax: +(34) 91 387 7815

Email: onunez@isciii.es

Running Title: The undiagnosed fraction of HIV in Spain

Word Count: 3,365

1

Abstract

Objective: We aim to estimate the number of people living with HIV and the undiagnosed fraction in Spain, where coverage of the HIV surveillance system has only recently become complete.

Methods: The reconstruction of all HIV diagnoses and infections was obtained by combining HIV and AIDS surveillance data. The imputation of the diagnoses and back-calculation of the infection incidence are integrated in a Bayesian framework to take into account the uncertainty associated with unavailable data.

Results: An estimated 141,000 [95% credible interval (CI): 128,000-155,000] persons were living with HIV by the end of 2013 in Spain and 18% [95% CI: 14.3%-22.1%) were unaware of it. A similar fraction of undiagnosed infections was obtained in MSM and heterosexuals (18.8% and 20.1%, respectively) but for IDU this fraction was 3.5%.

Conclusion: This study provides the first estimates of the number of people living with HIV and the undiagnosed fraction in Spain, using routine surveillance data. The proposed method could be useful for countries where the geographical coverage of the HIV surveillance system is partial or was completed only recently.

Note: Preliminary results were presented in the abstract 565 at the 35th Conference of Spanish Society of Epidemiology [XXXV Reunión Científica de la Sociedad Española de Epidemiología], 6-8 September 2017, Barcelona, Spain; oral communication number CO1.6 at the 18th Spanish Congress on AIDS and STI [XVIII Congreso Nacional sobre el sida e ITS], 22-24 March 2017, Seville, Spain.

1 Introduction

The estimation of the number of people living with HIV (PLHIV), including those who do not know their status, is essential for planning control measures, prevention and care resources, as well as for implementing HIV screening strategies.

These data are basic for building the two first steps of the HIV continuum of care and for monitoring progress towards the 90-90-90 target proposed by the Joint United Nations Program on HIV/AIDS (UNAIDS) [1].

However, the incidence of HIV infection is difficult to estimate. European countries have been using estimation methods according to the specificities of their respective surveillance systems. Some (Austria, Denmark, Belgium, Greece and the Netherlands) [2] have used the ECDC HIV modeling tool [3] based on the

3

back-calculation method [4, 5, 6, 7]. This method allows the reconstruction of the HIV infection incidence and the undiagnosed fraction, when HIV surveillance data are available from the beginning of the epidemic. Others, usually larger countries, with more complex surveillance systems or data availability limitations, such as France and Germany, have proposed modified back-calculation models [8, 9], or other customized methods, in Italy [10] and the United Kingdom [11, 12].

In Spain, the first estimates of the HIV incidence were obtained at the end of the 1990s [13] through back-calculation from the incidence of AIDS diagnoses [4]. However, after the introduction of highly active antiretroviral therapy (HAART) in mid 1990s, this method was no longer valid to reconstruct the HIV epidemic. The extended back-calculation method [6, 7] (as implemented in the ECDC HIV Modelling tool) solves this problem by using the first positive HIV test as an alternative endpoint for back-calculation. However, this method requires complete historical data on new HIV diagnoses and CD4$^+$ counts close to the date of diagnosis.

The Spanish HIV surveillance system (SINIVIH by its Spanish acronym) was set up in 2003 with a coverage of 34% (9 out of 19 Spanish regions) of the population and did not reach complete national coverage until 2013. Based on the available

4

information, the back-calculation method cannot be directly applied. The main objective of this article is to provide an estimate of PLHIV and undiagnosed fraction in Spain in 2013, overall and by transmission category, using available surveillance data.

2   Methods

We used a two-step method similar to the extended back-calculation approach proposed by Hall *et al.* [14]. In the first step, the HIV diagnosis history was reconstructed using a bivariate Poisson regression based on both AIDS and HIV surveillance data. In the second step, the number of infections was estimated using a modification of the standard back-calculation method, which takes into account the uncertainty about the diagnosis history obtained in the previous step. The two steps were carried out in a Bayesian framework using the JAGS software [15]. Finally, to obtain the number of PLHIV, the number of deaths among infected people taken from the Spanish National Statistics Institute (for AIDS-related deaths) and from the CoRIS [16] observational cohort (for non AIDS causes of death) was subtracted.

The previous method was stratified by transmission category (ie, MSM, IDU, heterosexual, and others) and overall estimates of PLHIV and undiagnosed infections were obtained by adding the corresponding counts obtained in each category.

2.1 Reconstruction of the HIV diagnosis history

The number of people diagnosed with HIV is the visible part of the epidemic.

In Spain, the available information does not allow for a direct estimation of the annual number of new HIV diagnoses across the country. The aim of this section is to reconstruct the evolution of the HIV diagnoses in Spain during the 1984-2013 period. For this purpose, a joint model approach was used to borrow information on HIV diagnosis included in the AIDS surveillance system, which collects data since 1981 with a national coverage.

2.1.1 Data sources

The SINIVIH began in 2003, although some regional information systems started earlier. Variables concerning new HIV diagnoses included in the SINIVIH are age,

sex, HIV transmission mode, country of birth, date of diagnosis, first CD4$^+$ cell count after the diagnosis (complete for about 85% of the cases) and concurrent AIDS diagnosis. We considered data from 2003 to 2013, updated in June 2016.

To complete this information, the National Registry of AIDS (RNS) was used. The RNS collects epidemiological data such as age, sex, transmission mode, country of birth, date of diagnosis of all AIDS cases diagnosed in Spain. We used data for the 1981-2013 period, reported by June 2016 to RNS. This database is not yet linked with the SINIHIV database, but it includes the date of the first positive HIV test in about 90% of the cases. It should be noted that this percentage varies slightly among transmission categories and stabilizes after 1986. Therefore, the RNS provides valuable information on the HIV diagnosis rates, especially for the period prior to HAART. A description of main characteristics of both surveillance systems are available in an Appendix (Table A2).

The spatio-temporal pattern of the HIV diagnoses rates varies substantially by transmission category. However, due to the uniformity of HIV policies in Spain regarding HIV testing and treatment recommendations, the trend of the diagnosis rates is quite similar between the different regions within the same category of transmission. This allows us to assume that in a Poisson regression

framework, the effects of the periods and regions are additive. This assumption greatly simplifies the model and allows the historical information included in the RNS to help reconstructing the evolution of the HIV diagnosis rates.

2.1.2 Joint model approach

In order to combine the information on HIV diagnosis included in the AIDS and HIV surveillance systems, the annual numbers of HIV diagnoses reported in both systems were jointly modeled. We used a bivariate Poisson regression model in which the HIV diagnosis rates and their restriction to people who develop AIDS share a common spatio-temporal component. This model assumes that there is no interaction between the temporal (year) and spatial (region) effects, within the same transmission category. This allows estimating these effects in each surveillance system, as well as the potential correlation between the two systems.

Let H_{rt} denote the number of new HIV diagnoses observed in a transmission category at year t and in region r. We denote $\widetilde{H}_{rt}$ as the restriction of H_{rt} to people who developed AIDS over 1981-2013 period, which is the number of HIV diagnoses reported in the AIDS surveillance system during this time-frame. These

two observed numbers are likely to share common spatio-temporal variations and are jointly modeled in the following way:

$$H_{rt} = \text{Poisson}(N_{rt}\rho_{rt}),$$

$$\widetilde{H}_{rt} = \text{Poisson}(N_{rt}\tilde{\rho}_{rt}),$$

where the offset N_{rt} is the general population size, which accounts for demographic variations by region and period (for MSM, this size corresponds to the number of men in the population). The rates ρ_{rt} and $\tilde{\rho}_{rt}$ vary according to the following equations:

$$\log\tilde{\rho}_{rt} = \tilde{u}_r + \tilde{v}_t,$$

$$\log\rho_{rt} = \beta_1\tilde{u}_r + u_r + \beta_2\tilde{v}_t + v_t,$$

where the spatial terms u_r and $\tilde{u}_r$ are assumed to have independent variations across the country, while the temporal terms v_t and $\tilde{v}_t$ are assumed to have smooth variations over calendar time. Note, that this model is over-parametrized and allows the evolution of HIV diagnosis rates to be estimated from the information included in the AIDS surveillance system (RNS) when the data available in the HIV surveillance system (SINIVIH) are scarce. The terms u_r and

v_t account for spatial and time variations of the diagnosis rates observed in the SINIVIH not explained by the variations of the corresponding rates in the RNS.

In the following section, the estimates provided by the previous model were used as input data for a back-calculation method in order to reconstruct the HIV incidence history in Spain.

2.2   HIV Infection incidence estimation

In these analyses, we use an extension of the model proposed by Sweeting *et al.* [6] that accounts for the specificities of the Spanish HIV/AIDS surveillance systems to estimate the history of the HIV epidemic. This multi-state model describes the annual progression of the disease as a unidirectional flow through three pre-AIDS stages characterized by CD4+ counts (Figure 1). After infection and in the absence of antiretroviral treatment, individuals progress to AIDS through up to three different CD4+ strata. Moreover, at each stage of this natural disease progression, the individuals can be diagnosed at rates which were assumed to vary smoothly over calendar time.

Annual data on new diagnoses of late HIV and non-late HIV (endpoints 4 and 5 in Figure 1) are required for this model. However, in our framework the annual numbers of new HIV diagnoses are not directly observed but estimated. The uncertainty about the annual number of new diagnoses decreases over calendar time, but is quite large before 2000. To reflect this uncertainty in the HIV incidence inference, Hall *et al.* [14] used a multi-imputation procedure [17]. Here, we have chosen an augmented model strategy which takes advantage of our Bayesian framework and is much less time-consuming.

2.2.1 Augmented model strategy

For each year t of the period 1984-2013, let e_t^4 and e_t^5 denote the expected number of late and non-late HIV diagnoses, respectively. These expectations correspond to the observed stages of the epidemic in the multi-state model pictured in Figure 1 and are complicated functions of the unknown infection incidence h_t and diagnosis rates d_{jt} (see the Appendix for further details). They can be estimated using the output of the Poisson regression model given in the previous section. The expected number of new diagnoses for year t is

$$n_t = \sum_r N_{rt} \rho_{rt}.$$

11

Let α_t denote the corresponding proportion of late HIV diagnoses observed in the surveillance system. Expectations e_t^4 and e_t^5 can be approximated by the estimated number of late and non-late HIV diagnoses, respectively:

$$\hat{e}_t^4 = \alpha_t n_t,$$

$$\hat{e}_t^5 = (1 - \alpha_t)n_t.$$

The uncertainty about these estimates is taken into account in the multi-state model inference by the creation of faked zero observations [18]:

$$0 = (\log e_t^4 - \log \hat{e}_t^4) + \varepsilon_t^4$$

$$0 = (\log e_t^5 - \log \hat{e}_t^5) + \varepsilon_t^5,$$

where ε_t^4 and ε_t^5 are random deviations with arbitrary high precision. Thus, this augmented model strategy links the two steps of our estimation method in a unique Bayesian framework and allows us to propagate the uncertainty about the diagnosis rates in the infection incidence estimation.

2.2.2   Diagnosis process and natural disease progression

To disentangle the contributions of the infection dynamic and the diagnosis process in the observed HIV diagnoses, additional data of CD4$^+$ counts close to

diagnosis (before start of treatment) were used. These data are based on a representative subsample of non-late HIV diagnoses that fall into each of the CD4+ count strata and are assumed to follow a multinomial distribution [6]. These additional data provide information that allow us to identify the diagnosis probabilities d_{jt} (Figure 1), which are assumed to vary according to the following equation

$$d_{jt} = \text{logit}(a_j + b_j\,\eta_t),$$

and the annual HIV incidence h_t, where η_t and h_t are assumed to vary smoothly over calendar time (see the Appendix for further details about the model).

Moreover, the probabilities q_{ij} of natural disease progression between CD4+ strata (Figure 1) are assumed to be constant over calendar time and are chosen as in Aalen *et al.* [5]. These probabilities correspond to mean sojourn times of 5.5, 4 and 1 years in stages CD4[≥500], CD4[200-500) and CD4[<200], respectively, which gives a mean incubation time for an untreated person of 10.5 years.

2.2.3   People living with HIV and the undiagnosed fraction

The number of PLHIV was calculated as the cumulative number of infections minus the cumulative number of deaths during the study period (1977-2013). Data for HIV/AIDS related deaths were obtained from the records of the National Statistics Institute (INE) which covers the entire national territory and accounted for about 40% of all causes of death among patients with HIV, at the end of the study period (2010-2013). However, the analysis of mortality data from the cohort of the Spanish HIV Research Network (CoRIS[20]) reveals that this percentage varies considerably between transmission categories (about 75% for MSM and 30% for IDU and Heterosexuals) in the same period. Therefore, the calculation of the number of deaths for all causes among HIV infected patients was stratified by period and transmission category (see the Appendix for further details).

The number of undiagnosed persons with HIV is defined as the cumulative number of infections minus the cumulative number of diagnoses during the study period. Then, the undiagnosed fraction is obtained as the ratio of the number of undiagnosed persons to the number of persons living with HIV.

3   Results

Using the method described in the previous section, we estimated that 141,000 (95% CI, 128,000-155,000) persons were living with HIV by the end of 2013 in Spain and 18% (95% CI, 14.3%-22.1%) were unaware of it. The estimated prevalence of persons living with HIV and undiagnosed infections were 0.36% and 0.06%, respectively. By category of transmission, we estimated 55,100 MSM (95% CI: 52,000-58,500), 28,400 heterosexuals (95% CI: 26,100-31,250) and 23,750 IDU (95% CI: 19,250-28,650) were living with HIV in 2013. The corresponding proportions of undiagnosed infections were 18.8% (95% CI: 15.0% -22.8%) for MSM, 20.1% (95% CI: 15.2% -26.3%) for heterosexuals and 3.5% (95% CI: 2.3% - 6.1%) for IDU.

The model estimate of the overall HIV incidence shows a sharp increase after 1977, with a peak in 1985-1986 of approximately 25,000 infections per year (Figure 2). The incidence declined after 1986 to a minimum of 2,500 infections per year in the early 1990s. Then, in the late 1990s, there was a small resurgence of the epidemic that, since 2000, progressively stabilized. The incidence among MSM (Figure 3) reflected this general trend, but the resurgence of the epidemic

is more pronounced in this category of transmission, accounting for about 40%

of new infections at the end of the study period. The trend of HIV incidence in

IDUs is also similar to the overall trend as of the late 1990s. Though this

accounted for more than half of the incidence in this period, incidence has

steadily declined since 2000 to reach very low levels at the end of the study

period. Among heterosexuals, the incidence increased more gradually after

1977, peaking in the late 1980s, declining in the early 1990s and then remaining

relatively stable around 1,500 infections per year.

The previous estimates were based on the reconstruction of the evolution of the

number of HIV diagnoses. Figure 4 shows in log-scale the model estimates of the

annual number of diagnoses for the whole population and among people who develop

AIDS. From this figure, we can see that both curves behave similarly before 2000.

Beginning in the year 2000, however, the number of HIV diagnoses in the whole

population remains fairly stable, while the HIV cases reported in the RNS continues to

decrease.

A similar representation by transmission categories offers more insights on the

evolution of the annual number of diagnoses. Figure 5 reveals that the apparent

stability of the overall HIV diagnoses hides heterogeneous trends across

transmission categories. For MSM, for example, the number of HIV diagnoses

tends to increase in the last period, whereas for IDU it has been constantly decreasing since the 1990s.

The model also provided an estimate of the probabilities of diagnosis at the different stages of $CD4^+$ (see Appendix, Figure A1). There was a notable uncertainty in the diagnosis probabilities before 2000 due to the sparsity of $CD4^+$ data in that period. After the HAART implementation, in all transmission categories (except IDU) there was a clear growing trend in diagnosis probability for patients with $CD4^+$< 500 cells/µl, while for patients with high $CD4^+$ counts (> 500 cells/µl) it remains almost constant at around 10%. Moreover, compared with the other transmission categories, MSM had a higher proportion of individuals diagnosed in the first two stages of $CD4^+$ (> 200 cells/µl), 90% for MSM versus 75% for the other categories (see Appendix, Figure A2). The median time to diagnosis decreased slowly (especially in heterosexuals and others) during the study period and reached 5 years around 2007 (see Appendix, Figure A3).

4 Discussion

This is the first study that provides credible intervals for the number of PLHIV and the undiagnosed fraction in Spain, using routine surveillance data. The main advantage of our approach is that it does not depend on prevalence data and size estimates of key populations, but it is based on data collected routinely, which allows for a periodic update of the estimates.

Our HIV incidence estimates are consistent with our knowledge of the epidemic — that HIV spread rapidly during the 1980s among IDUs and to a lesser extent among MSM [13]. From 1980 to 1990, AIDS rates in Spain were the highest in Europe, with a peak of 19 cases per 100,000 in 1994. Today, the number of new HIV diagnoses has decreased markedly; our rate of new HIV diagnoses is close to other European countries, although it is higher than the European Union average [21]. The estimated number of PLHIV (around 141,000) is comparable with the results of UNAIDS that account for 130,000-160,000 people living with HIV in Spain in 2013 [22]. The estimate of HIV prevalence in Spain (0.36%) is higher than in other European countries with similar concentrated epidemics, such as France (0.29%) or Italy (0.25%) [2].

We estimated that there was a range of 14-22% of people with undiagnosed HIV in 2013. This estimate is in the range of recent estimates for other European countries: 16% in France [24], 19% in the United Kingdom [2] and higher than in Italy (11-13%, [10]), as well as in other countries with concentrated epidemics: 16.5% in 2013 in the United States [25]. The average estimates of some EU/EEA countries using ECDC HIV Modelling Tool in 2015 were around 15% [26]. However, the different methods used in these countries, according to the availability of data, make these comparisons difficult. The proportions of undiagnosed infections among MSM and heterosexuals are similar to those in the United States [25] and the United Kingdom [12]. However, we obtain a lower proportion among IDUs [25]. This difference can be explained by the fact that our HIV epidemic has historically been concentrated on IDUs and that many resources have been allocated to control HIV infection in this subpopulation.

The purpose of the regression imputation method proposed in this article was to reconstruct the evolution of the number of HIV diagnoses in order to estimate the incidence of infection using a standard back-calculation approach. The imputation of the diagnosis and back-calculation of the infection incidence are integrated in a Bayesian framework to take into account the uncertainty associated with unavailable data. One of the strengths of our approach is that it

provides national estimates even if some regions have a young HIV surveillance system, drawing on available information from other regions with a longer collection period. Though one may argue that a single model for both diagnosis and infection rates [27, 8], would be a more satisfactory approach, it turned out to be very unstable and time-consuming in our framework.

This study, however, has some limitations. The method is based on reported cases, therefore the delay in reporting AIDS and HIV diagnoses could affect the estimates. To address this issue, we used data updated to 2016 to estimate the infection in 2013. Another problem could be the lack of information in CD4+ count. In the Spanish surveillance registers, the completeness of this variable is higher than the EU/EEA average [21]. CD4 data are assumed to be missing at random and no imputation methods have been applied.

Unlike other modeling [7,28], our model assumes that all individuals start with CD4 count above 500 cells/µl. The fraction of newly infected individuals with lower levels of CD4 varies considerably with the viral load (not available in our surveillance system) and remains difficult to specify. In addition, the time between the sharp decline in the number of CD4s after infection and their recovery is usually very short and we assume, as in Song et al. [25], that this

violation of our assumption has little effect as few infected individuals are diagnosed in this short period.

Information about HIV infected people who moved abroad is not available in our system, as in most European countries. This problem is very difficult to address and could potentially lead to overestimating the number of PLHIV [2]. Finally, to avoid overestimating the number of infected people who are still alive, we complemented the national mortality statistics with data provided by the CoRIS cohort, in order to capture mortality from causes other than HIV/AIDS, mainly in recent years [16].

Our results are useful for guiding HIV policy in Spain. The estimated number of PLHIV is the key denominator for building the continuum of HIV care. In addition, this figure is important for reporting the burden of this disease and the allocation of resources. Reducing the undiagnosed fraction of HIV should be a priority in our country. In Spain, HIV tests are free at health centers and community services. Several autonomous regions have implemented HIV testing programs in pharmacies that have demonstrated good ability to reach and diagnose previously unexplored populations [29]. The methodology developed here will help evaluate such new HIV testing strategies and preventative measures in the near future.

In summary, we estimated that there were 141,000 people living with HIV in Spain in 2013 and between 14-22% were not aware of their HIV status. The proposed method could be useful for countries with historical data on AIDS cases, whose HIV surveillance systems have a geographical coverage partial or complete only recently.

Acknowledgments

References

1.   UNAIDS. 90–90–90 - An ambitious treatment target to help end the AIDS epidemic.   2014.   URL: http://www.unaids.org/en/resources/documents/2017/90-90-90 (visited on 2018).

2. Gourlay A, Noori T, Pharris A, Axelsson M, Costagliola D, Cowan S, et al. The Human Immunodeficiency Virus Continuum of Care in European Union Countries in 2013: Data and Challenges. *Clin. Infect. Dis.* 2017; 64:1644–1656.

3. ECDC. HIV modelling tool. Stockholm: European Centre for Disease Prevention and Control, 2015. URL: http://ecdc.europa.eu/en/publicationsdata/hiv-modelling-tool (visited on 2018).

4. Brookmeyer R and Gail MH. A Method for Obtaining Short-Term Projections and Lower Bounds on the Size of the AIDS Epidemic. *Journal of the American Statistical Association* 1988; 83:301–308.

5. Aalen OO, Farewell VT, De Angelis D, Day NE, and Gill ON. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Stat Med* 1997; 16:2191–2210.

6. Sweeting MJ, De Angelis D, and Aalen OO. Bayesian back-calculation using a multi-state model with application to HIV. *Statistics in medicine* 2005; 24:3991–4007.

7.   van Sighem A, Nakagawa F, De Angelis D, Quinten C, Bezemer D, Coul EO de, et al. Estimating HIV Incidence, Time to Diagnosis, and the Undiagnosed HIV Epidemic Using Routine Surveillance Data. *Epidemiology* 2015; 26:653–660.

8.   Sommen C, Alioum A, and Commenges D. A multistate approach for estimating the incidence of human immunodeficiency virus by using HIV and AIDS French surveillance data. *Statistics in medicine* 2009; 28:1554–1568.

9.   Heiden M An der, Marcus U, Gunsenheimer-Bartmeyer B, and Bremer V. Undiagnosed HIV infection in Germany: who do we need to target? [TUPEC137] *21st International AIDS Conference* 2016, Durban.

10.   Mammone A, Pezzotti P, Regine V, Camoni L, Puro V, Ippolito G, et al. How many people are living with undiagnosed HIV infection? An estimate for Italy, based on surveillance data. *AIDS* 2016; 30:1131–1136.

11.   Goubar A, Ades AE, De Angelis D, McGarrigle CA, Mercer CH, Tookey PA, et al. Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance

data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2008; 171:541–580.

12. Nakagawa F, Sighem A van, Thiebaut R, Smith C, Ratmann O, Cambiano V, et al. A Method to Estimate the Size and Characteristics of HIVpositive Populations Using an Individual-based Stochastic Simulation Model. *Epidemiology* 2016; 27:247–256.

13. Castilla J and Fuente L de la. [Trends in the number of human immunodeficiency virus infected persons and AIDS cases in Spain: 1980-1998]. *Med Clin (Barc)* 2000; 115:85–89.

14. Hall HI, Song R, Rhodes P, Prejean J, An Q, Lee LM, et al. Estimation of HIV incidence in the United States. *JAMA* 2008; 300:520–529.

15. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. [JAGS] *Hornik K, Leisch F, Zeileis A (eds) Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003)* 2003, Vienna.

16. Hernando V, Alejos B, Monge S, Berenguer J, Anta L, Vinuesa D, et al. All cause mortality in the cohorts of the Spanish AIDS Research Network (RIS)

compared with the general population: 1997-2010. *BMC Infect. Dis.* 2013; 13:382.

17. Rubin DB. *Multiple imputation for nonresponse in surveys*, vol. 81. John Wiley & Sons; 2004.

18. Ruiz-Cárdenas R, Krainski ET, and Rue Ha. Direct fitting of dynamic models using integrated nested Laplace approximations—INLA. *Computational Statistics & Data Analysis* 2012; 56:1808–1828.

19. Lindgren F and Rue Ha. On the second-order random walk model for irregular locations. *Scandinavian journal of statistics* 2008; 35:691–700.

20. Sobrino-Vegas P, Gutiérrez F, Berenguer J, Labarga P, García F, Alejos-Ferreras B, et al. [The Cohort of the Spanish HIV Research Network (CoRIS) and its associated biobank; organizational issues, main findings and losses to follow-up]. *Enferm. Infecc. Microbiol. Clin.* 2011; 29:645– 653.

21. Europe EC for Disease Prevention and Control/WHO Regional Office for. HIV/AIDS surveillance in Europe 2017 - 2016 data. 2017. URL: http://ecdc.europa.eu/en/publications-data/hivaids-surveillanceeurope-2017-2016-data (visited on 2018).

22. UNAIDS. Methods for deriving UNAIDS estimates. Tech. rep. Geneva, 2016. URL: http://www.unaids.org/en/resources/documents/ 2016/methods-for-deriving-UNAIDS-estimates (visited on 2018).

23. Díaz A, Sobrino-Vegas P, and Del Amo J. La cascada del Tratamiento en España: Primeras estimaciones. [P-081] *VI Congreso Nacional de GeSIDA* 2014, Málaga.

24. Supervie V, Ndawinz JD, Lodi S, and Costagliola D. The undiagnosed HIV epidemic in France and its implications for HIV screening strategies. *AIDS (London, England)* 2014; 28:1797.

25. Song R, Hall HI, Green TA, Szwarcwald CL, and Pantazis N. Using CD4 Data to Estimate HIV Incidence, Prevalence, and Percent of Undiagnosed Infections in the United States. *J. Acquir. Immune Defic. Syndr.* 2017; 74:3–9.

26. Pharris A, Quinten C, Noori T, Amato-Gauci AJ, Sighem A van, and ECDC HIV/AIDS Surveillance and Dublin Declaration Monitoring Networks. Estimating HIV incidence and number of undiagnosed individuals living with HIV in the European Union/European Economic Area, 2015. *Euro Surveill.* 2016; 21.

27. Bellocco R and Marschner IC. Joint analysis of HIV and AIDS surveillance data in back-calculation. *Stat Med* 2000; 19:297–311.

28. Cori A.; Pickles M.; van Sighem A.; Gras L.; Bezemer D.; Reiss P.; Fraser C. CD4$^+$ cell dynamics in untreated HIV-1 infection: Overall rates, and effects of age, viral load, sex and calendar time. Aids **2015**, 29:2435–2446.

29. Fernández-Balbuena S, Belza MJ, Zulaica D, Martinez JL, Marcos H, Rifá B, et al. Widening the Access to HIV Testing: The Contribution of Three In-Pharmacy Testing Programmes in Spain. *PLoS ONE* 2015; 10:e0134631.
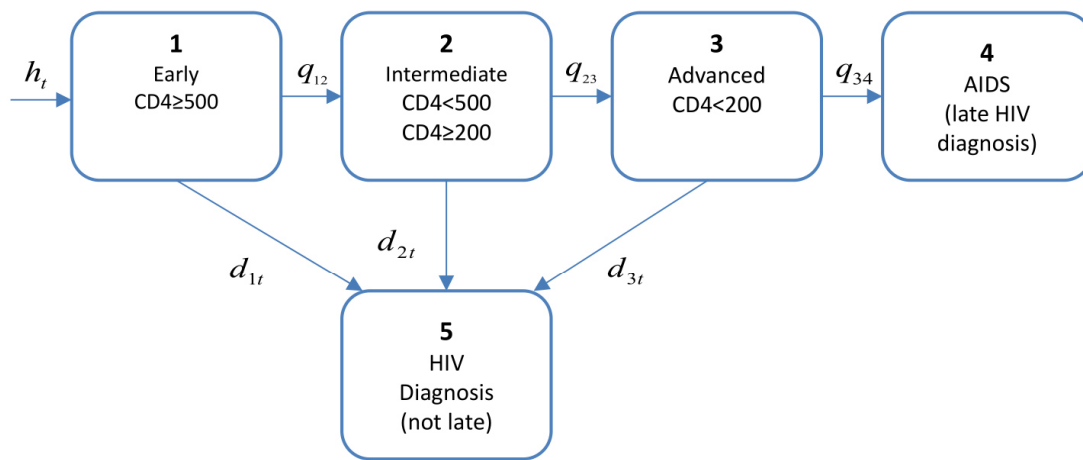
Fig. 1: Multi-state model for the disease progression from HIV infection to diagnosis from Sweeting *et al.* [6].
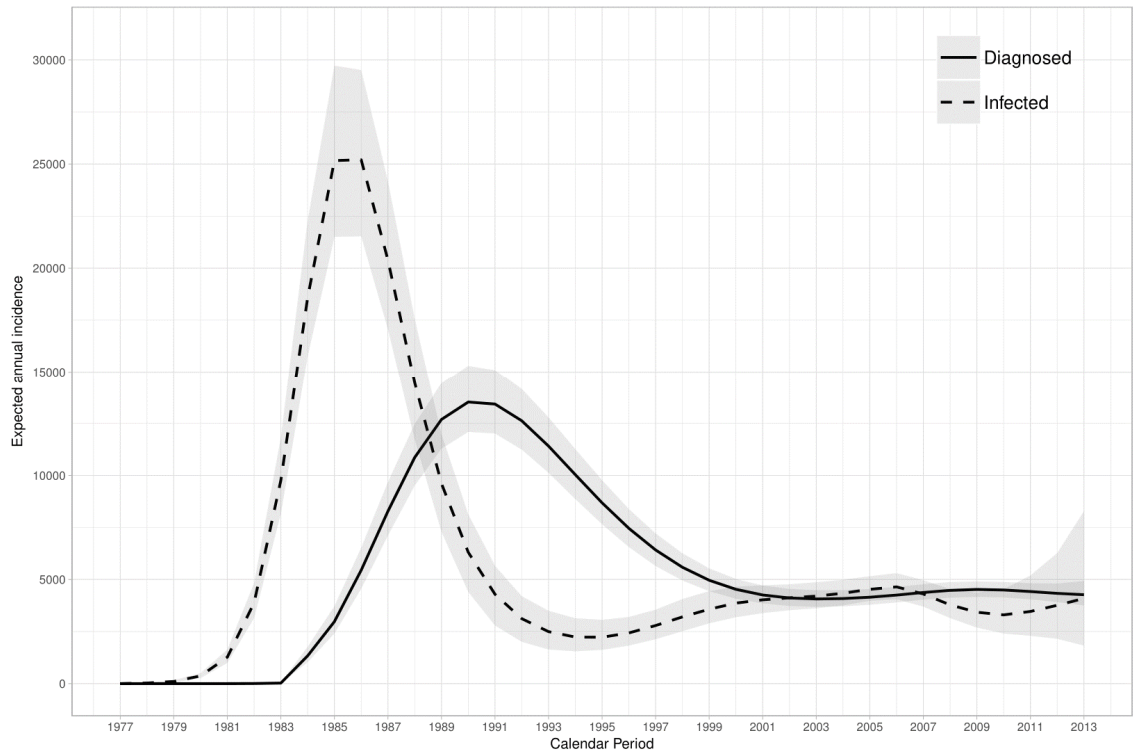
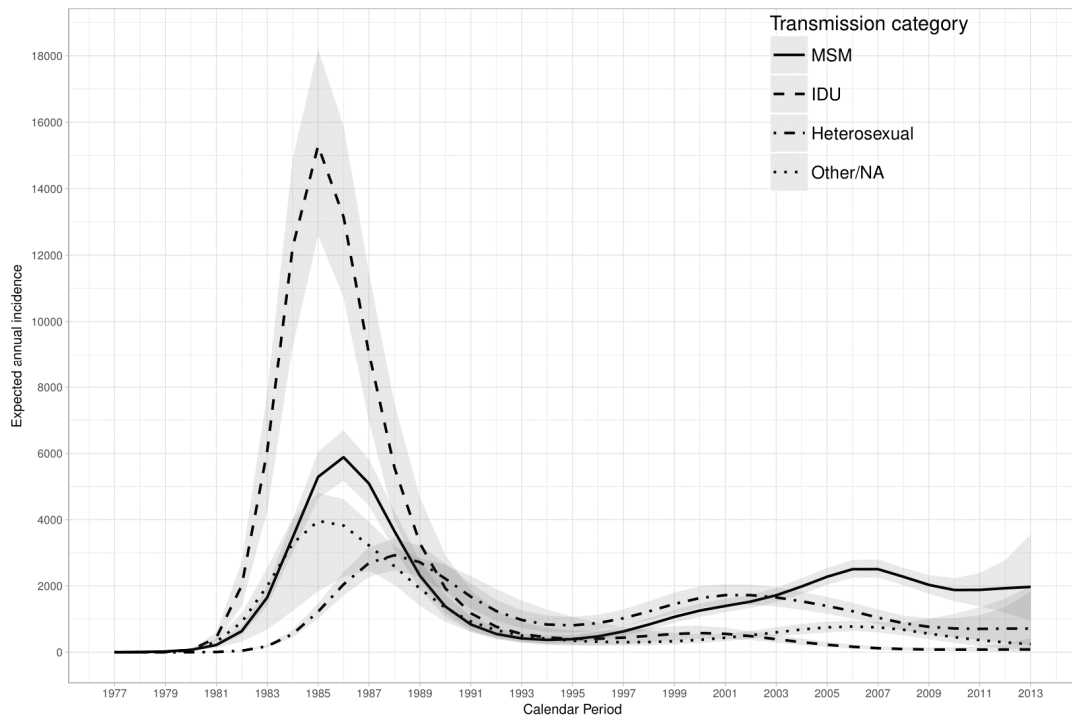Fig. 2: Overall annual new cases of HIV infections and HIV diagnoses (Spain, 1977-

2013).

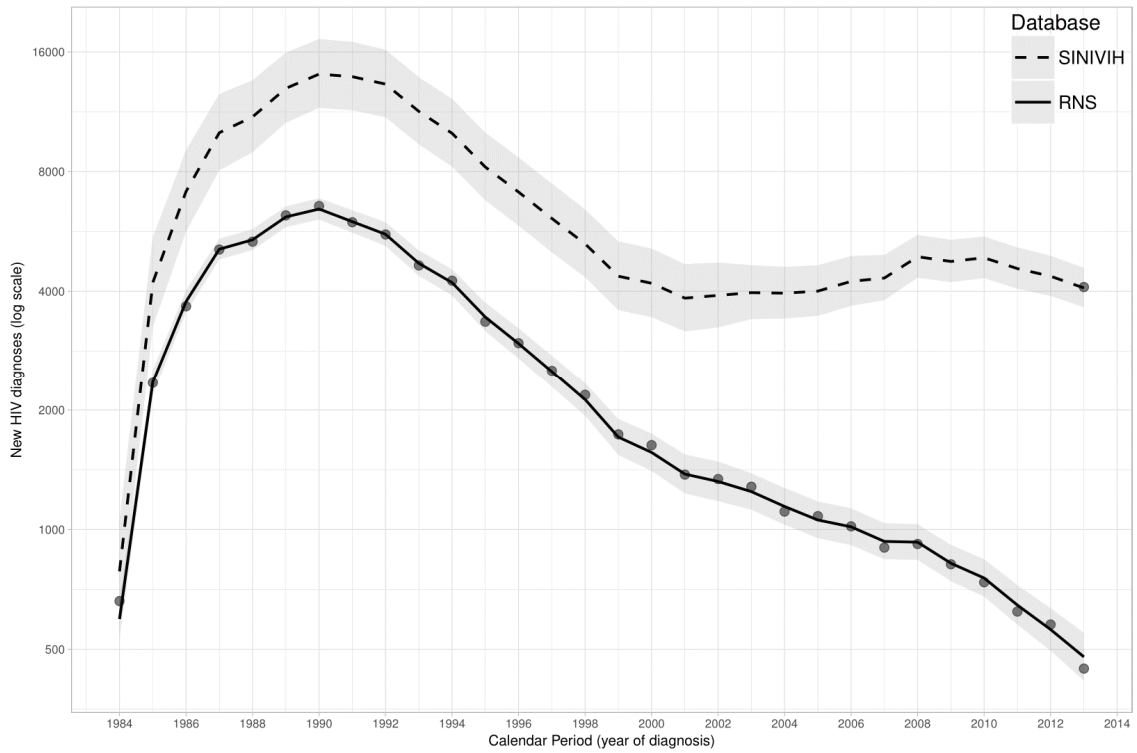Fig. 3: Annual new cases of HIV infections by transmission category (Spain, 1977-

2013).

Fig. 4: Annual new HIV diagnoses estimates (in log scales) over the 1984-2013 period, in the HIV surveillance system (SINIVIH) and in the AIDS surveillance system (RNS).
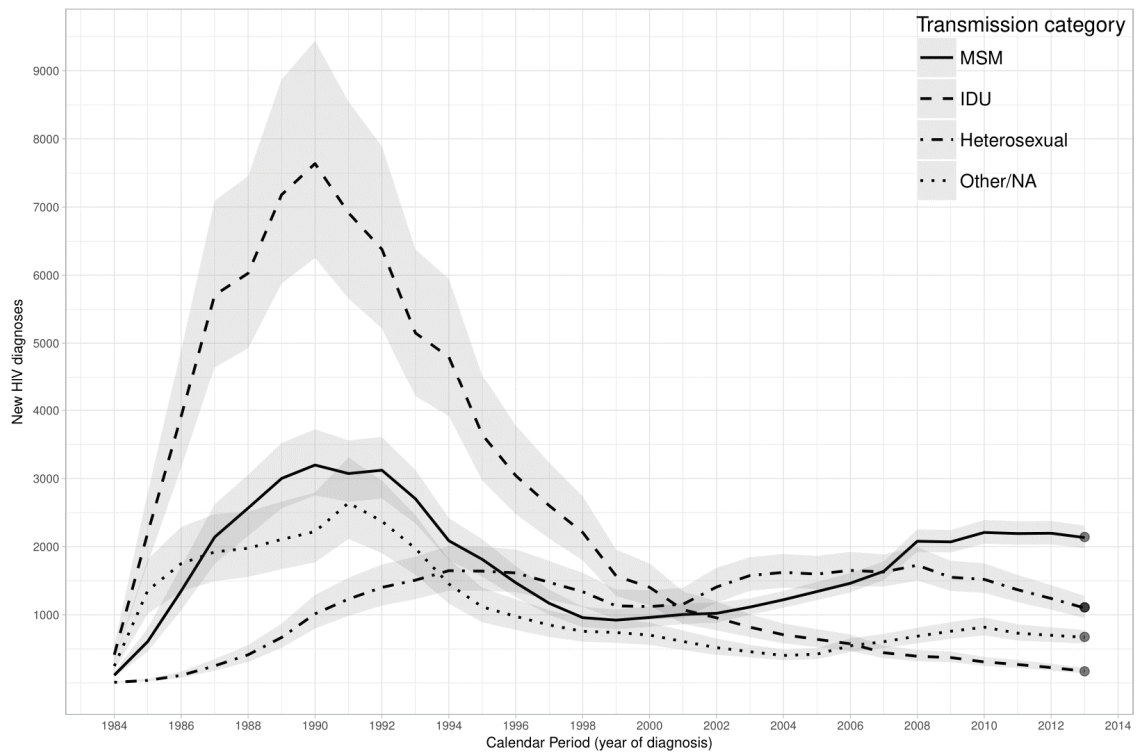
Fig. 5: Annual new HIV diagnoses estimates by transmission category (Spain, 1984-2013). MSM indicates men who have sex with men; IDU, injection drug users; Other/NA, haemophiliac/transfusion recipient, mother-to-child or unknown categories.