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

Table A1. Characteristics of National Registry of AIDS and new HIV diagnoses surveillance system in Spain

<table>
<thead>
<tr>
<th>National Registry of AIDS</th>
<th>HIV new diagnoses surveillance system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collects new AIDS cases following European case definition</td>
<td>Collects new HIV diagnosis following European case definition</td>
</tr>
<tr>
<td>Initiated in 1983 at national level.</td>
<td>Initiated in 2007 at national level although some regional systems started earlier. Collects data retrospectively since 2003 in 9 out of 19 Autonomous Regions in Spain</td>
</tr>
<tr>
<td>Collects data retrospectively since 1981</td>
<td></td>
</tr>
<tr>
<td>Coverage: 100% from the beginning</td>
<td>Coverage increased from 34% in 2003 (9 Autonomous Regions reported data) to 100% in 2013.</td>
</tr>
<tr>
<td>Compulsory notification</td>
<td>Compulsory notification</td>
</tr>
<tr>
<td>Notifiers are clinicians but data are complemented from other data sources (mortality registry, clinical records, hospital discharge registries)</td>
<td>Notifiers are clinicians but data are complemented from other data sources (laboratory registries, clinical records, hospital discharge registries)</td>
</tr>
<tr>
<td>Main variables: age, sex, transmission mode, country of birth, date of diagnosis of all AIDS cases, date of the first positive HIV test</td>
<td>Main variables: age, sex, transmission mode, country of birth, date of diagnosis, CD4 count, date of first CD4 count after HIV diagnosis</td>
</tr>
</tbody>
</table>
Table A2. Estimated cumulative deaths, fraction of undiagnosed infections and number of persons living with HIV in Spain in 2013 by transmission category.

<table>
<thead>
<tr>
<th>Transmission category</th>
<th>Figure</th>
<th>Estimates (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Undiagnosed fraction (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLHIV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>deaths</td>
</tr>
<tr>
<td>MSM</td>
<td></td>
<td>Undiagnosed fraction (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLHIV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>deaths</td>
</tr>
<tr>
<td>IDU</td>
<td></td>
<td>Undiagnosed fraction (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLHIV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>deaths</td>
</tr>
<tr>
<td>Heterosexual</td>
<td></td>
<td>Undiagnosed fraction (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLHIV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>deaths</td>
</tr>
</tbody>
</table>
Figure A1: Probabilities of diagnosis from CD4 stages over the studied period.
**Figure A2**: Estimated proportion of non-late HIV diagnoses coming from each CD4 stage over the period 2003-2013. Dots are the proportions observed in the HIV surveillance system.
Figure A3: Estimated probability of HIV as a cause of death by transmission category.
Figure A4: Estimation of the annual number of new diagnoses for different Spanish regions. Dots are the new diagnoses reported in the HIV surveillance system.
Figure A5: Proportion of each transmission category among observed deaths in the AIDS surveillance system over the period 1985-2013.
Bayesian formulation of the model:

**Step 1: Estimation of the annual number of new HIV diagnoses**

Let $H_{rt}$ denote the number of new HIV diagnoses observed in a transmission category at year $t$ and in region $r$, and $\tilde{H}_{rt}$ the restriction of $H_{rt}$ to people who developed AIDS. These numbers are jointly modeled in the following way:

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

where the offset $N_{rt}$ is the general population size, and the rates $\rho_{rt}$ and $\tilde{\rho}_{rt}$ vary according to the following equations:

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

The spatial terms $u_r$ and $\bar{u}_r$ are random effects with normal prior $N(0, \tau_R)$ and $N(0, \bar{\tau}_R)$, respectively. The smoothness on temporal terms $v_t$ and $\bar{v}_t$ is induced by assuming a second order random walk prior:

$$v_{t+2} = N(2v_{t+1} - v_t, \tau_T),$$
$$\bar{v}_{t+2} = N(2\bar{v}_{t+1} - \bar{v}_t, \bar{\tau}_T),$$

where $v_1, v_2, \bar{v}_1$ and $\bar{v}_2$ are treated as fixed unknown parameters with a normal prior $N(0, \tau_0)$.

**Step 2: Estimation of the HIV infection incidence**

As previously, the smoothness over time of the annual HIV incidence $h_t$ and the diagnosis probability $d_{kt}$ from each CD4 stage ($k=1,2,3$) was obtained by assuming second order random walk priors on this terms:

$$h_{t+2} = N(2h_{t+1} - h_t, \tau_h),$$
$$\eta_{t+2} = N(2\eta_{t+1} - \eta_t, \tau_\eta),$$

where

$$d_{kt} = \text{logit}(a_k + b_k \eta_t),$$

and $a_k, b_k$ are unknown parameters with a normal prior $N(0, \tau_0)$. 


Let \( p_{tkl} \) denote the one-step transition probability from state \( k \) to state \( l \) at time \( t \) (year) associated with the back-calculation model formulated in Figure 1 of the manuscript. Using the notations used in this figure, we have that the transition probability matrix

\[
P_t = (p_{tkl}) = \begin{pmatrix}
p_{t11} & q_{12}(1 - d_{1t}) & 0 & 0 & d_{1t} \\
0 & p_{22} & q_{23}(1 - d_{2t}) & 0 & d_{2t} \\
0 & 0 & p_{33} & q_{34}(1 - d_{3t}) & d_{3t}
\end{pmatrix},
\]

where \( p_{tkk} = 1 - \sum_{l>k}^5 p_{tkl} \) for \( k = 1,2,3 \).

Given this matrix, the expected number (prevalence) \( E_{tk} \) of persons in state \( k \) at time \( t \) is obtained using the following recurrence equations [1]:

\[
E_1 = I_1,
E_t = P_t' E_{t-1} + I_t, \quad t \geq 2
\]

Where \( P_t' \) is the transpose of \( P_t \) and \( I_t \) is a 5-dimensional column vector with the annual infection incidence \( h_t \) as the first component, and all other components equal to 0.

In the same way, the expected number of new arrivals (incidence) \( e_{tk} \) of persons in state \( k \) at time \( t \) is obtained using the recurrence relations:

\[
e_1 = I_1,
e_t = \bar{P}_t' E_{t-1} + I_t, \quad t \geq 2
\]

where \( \bar{P}_t \) is the same matrix as \( P_t \) but with zeros in its diagonal.

From this formulation, the expected proportion of non-late HIV diagnoses coming from each CD4 stage \( (k = 1,2,3) \) over time is

\[
\theta_{tk} = \frac{E_{tk} d_{tk}}{e_{t5}}
\]

Let denote \( r_{t1}, r_{t2} \) and \( r_{t3} \), the observed numbers of new non-late HIV diagnoses at time \( t \) with an available CD4 count at diagnosis greater than or equal to 500, between 200 and 500, and less than 200, respectively. Then, it is assumed that

\[
r_t = \text{Multinomial}(\theta_t, R_t),
\]

where \( R_t = r_{t1} + r_{t2} + r_{t3} \).

The expected number of new diagnoses for year \( t \) is
\[ n_t = \sum_r N_r \rho_{rt}. \]

Let \( \alpha_t \) 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 two steps described in this Bayesian formulation are connected using an augmented model strategy by the creation of faked zero observations:

\[ 0 = (\log e_t^4 - \log \hat{e}_t^4) + e_t^4, \]
\[ 0 = (\log e_t^5 - \log \hat{e}_t^5) + e_t^5, \]

where \( e_t^4 \) and \( e_t^5 \) are normal random deviations with arbitrary high precision:

\[ e_t^4 = N(0, \tau_e), \]
\[ e_t^5 = N(0, \tau_e). \]

Model implementation:
The following code is an implementation using JAGS [2] of the Bayesian model described in the paper to estimate the infection incidence. Flat priors were used for all parameters, except for temporal components whose prior distribution was slightly informative in order to obtain a smooth estimate of the HIV incidence curve. The estimation was performed using 40 MCMC chains with a burn-in period of 2,000. Algorithm convergence was monitored using the Gelman-Rubin diagnostic [3].
variables

Mu[N] # Expected annual number of diagnoses
r[N], # region index
t[N], # time index
u[N], # region effects
v[N], # time effects (autoregressive prior)
mu[N], # expected rates of HIV diagnosis in the HIV surveillance system
mu_s[N], # expected rates of HIV diagnosis in the AIDS surveillance system
P[N-1,1,3], # Matrix of transition probabilities
P0[N-1,1,3], # Working matrix
e[N], # Expected annual prevalence in the different states
e[N], # Expected annual incidence in the different states
d[N-1,2], # Annual diagnosis probabilities
prop[N-1,2], # Expected proportion of non-late HIV diagnoses coming from each CD4
h[N], # Expected annual incidence of the HIV infection
w[N], # autoregressive prior for the time variations of the HIV infection
g[N], # autoregressive prior for the time variations of the diagnosis probabilities

model{

presso

go on

to

for(i in 1:N) { 
  y[i] ~ dpois(mu[i]) #HIV diagnosis counts in the SINI-HIV
  v_s[i] ~ dpois(mu_s[i]) #HIV diagnosis counts in the RNS
}

for(i in 1:N) { 
  # Expected response
  log(mu[i]) <- log(pop[i]) + u[i] + v[i] + beta.t*u_s[i] + beta.r*v_s[i]
  log(mu_s[i]) <- log(pop[i]) + u_s[i] + v_s[i] #
}

for(k in 1:N) { 
  # prior for region effects
  u[k] ~ dnorm(0,tau.s)
  u_s[k] ~ dnorm(0,tau.s.r)
}

for(i in 1:N) { 
  # prior for period effects (2nd order random walk)
  v[i] ~ dnorm(v[i-1]*2-v[i-2], tau.t)
  v_s[i] ~ dnorm(v_s[i-1]*2-v_s[i-2],tau.s.t)
}

# Priors for the parameters
v[1] ~ dnorm(0,prec)
v[2] ~ dnorm(0,prec)
v_s[1] ~ dnorm(0,prec)
v_s[2] ~ dnorm(0,prec)
beta.r ~ dnorm(0,prec)
tau.r ~ dgamma(1,prec)
tau.s ~ dgamma(1,prec)
beta.t ~ dnorm(0,prec)
tau.t ~ dgamma(1,prec)
tau_s.t ~ dgamma(1,prec)

Mu <- X.t %*% mu #annual number of HIV diagnoses across the country

# Faked zero observations to link the two steps
for (i in 1:Nt){
  zeros[i] ~ dnormal(log(e[i,3]) + eps) 
  if (i > 1) zeros[i] ~ dnormal(log(e[i,3]) + eps) 
  zeros_s[i] ~ dnormal(log(e[i,3]) + eps) 
}

for(i in 1:(Nt-1)){
  CD4[i,1] ~ dmulti(prop[i], nCD4[i,1])
}
# Prior on diagnosis probabilities variations

---

# Prior on HIV incidence curve (2nd order random walk)

---

# Expected proportion of non late HIV

---

# Expected prevalence and incidence

---

# Transition Probability Matrix

---

# Probabilities of natural progression

---

# Prior on HIV incidence curve (2nd order random walk)

---

# Prior on diagnosis probabilities variations (2nd order random walk)


```r
# initial values
w[1] ~ dnorm(0, prec)
w[2] ~ dnorm(0, prec)
g[1] ~ dnorm(0, prec)
g[2] ~ dnorm(0, prec)

for(k in 1:3){
  b[k] ~ dnorm(0, prec)
  a[k] ~ dnorm(0, prec)
}

# smooth precision
tau.w ~ dgamma(10, prec)
tau.g ~ dgamma(10, prec)

```
References


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.
AIDS: Author’s paper submission checklist

| Title of paper: | ► Estimating the number of people living with HIV and the undiagnosed fraction in Spain |
| Names of authors: | ► Nuñez Olivier, Hernando Victoria, Díaz Asunción |

**AUTHORS SHOULD PLEASE ENSURE THAT ALL APPROPRIATE INFORMATION (EG. CONFLICT OF INTEREST STATEMENTS) ARE ALSO INCLUDED IN THE TEXT OF THE ARTICLE.**

1. **DUPLICATE PUBLICATION** is not acceptable and includes papers, or letters to the Editor reporting the same data previously published in any journal. Abstracts of papers presented at meetings and published in the proceedings of such meetings do not constitute duplicate publication, but should be disclosed by including a note at the beginning of the paper, i.e., "Data presented previously at (state meeting) and published as abstract in (give reference)". Have you published these data previously?

► A very brief summary has been included in “Gourlay A et al. The Human Immunodeficiency Virus Continuum of Care in European Union Countries in 2013: Data and Challenges. Clin Infect Dis. 2017 Jun 15;64(12):1644-1656. doi: 10.1093/cid/cix212”

2. **CONFLICT OF INTEREST** include financial support from the biomedical industry or other commercial sources in the form of research grants, bench fees, consultancy or lecture fees, travelling expenses, payment of registration fees, consultancy appointments, posts held in the biomedical industry or equipment manufacturers, stock holdings in the company, free supply of drugs and the like. These should be stated in relation to each author. Has any of the authors any conflict of interest? Please state details.

► No

3. **CONSENT** Please note that patient's, or normal control’s, written consent is needed not only for full papers, but also for case reports. The written consent needs to include not only agreement to undergo treatment, or participate in an experiment or a randomised control trial, but also agreement for anonymised data to be published in a scientific journal. Was patient’s consent obtained and in what form?

► According to Spanish legislation (Ley 33/2011, de 4 de octubre, General de Salud Pública), surveillance data not require patient’s consent. Patients included in the CoRIS cohort, have to sign a consent for being included in the cohort.

4. **ETHICS** All studies need to be approved by the local Ethical Committees. Was your study? Please provide the approval from your local Ethical Committees for any animal experimentation or human subject studies.

►

5. **AUTHOR’S CONTRIBUTIONS AND APPROVAL OF TEXT** Please state briefly how each of the authors contributed to the study, to data analysis and to the writing of your paper. Subject to your agreement, we will print this information, if the paper is accepted for publication. In addition, please confirm that all the authors have read and approved the text as submitted to AIDS. Justify individual’s contributions when the author list exceeds 10.
AD was the main study researcher. She supervised all phases of the work, reviewed surveillance data and quality, and critically revised the article. ON did the statistical analysis, interpretation of results, and drafted the manuscript. VH provided mortality data and made important contributions to successive versions of the manuscript. All authors have seen and approved the final manuscript.

6. STATISTICAL ANALYSIS Kindly please let me know who performed the statistical analysis of your data.

► Olivier Nuñez

7. ARTICLE LENGTH Please acknowledge that you have kept within the word and insert limits for your submission by ticking the relevant boxes below, and that you have indicated the word count on the title page of your article

<table>
<thead>
<tr>
<th>Article length</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Original papers, 3500 words of text excluding references with no more than five inserts (figures/tables)</td>
<td>□ Concise communications, 1800 words of text excluding references with maximum of two inserts (figures/tables)</td>
</tr>
<tr>
<td></td>
<td>□ Research Letters, 1000 words excluding summary with no more than one insert (figure/table)</td>
</tr>
<tr>
<td></td>
<td>□ Correspondence, 750 words excluding references with no more than one insert (figure/table)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title page</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Title no more than 120 characters</td>
<td></td>
</tr>
<tr>
<td>✓ Running head, no more than 40 characters</td>
<td></td>
</tr>
<tr>
<td>✓ Justification of the number of contributors greater than 10 included in this document</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Abstract page</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Abstract no more than 250 words</td>
<td></td>
</tr>
<tr>
<td>✓ 5–7 keywords listed</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word count</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Word count of text (excluding references) included on title page</td>
<td></td>
</tr>
</tbody>
</table>

8. CLINICAL TRIALS AND BEHAVIOURAL EVALUATIONS: Authors reporting results of randomized controlled trials should include with their submission a complete checklist from the CONSORT statement, see JAMA 1996; 227:637-639 or http://www.consort-statement.org. For behavioural and public health evaluations involving non-randomized designs, authors should include with their submission a complete checklist from the TREND statement, see Am J Public Health 2004; 94:361-366 or www.trend-statement.org.

REGISTRATION OF CLINICAL TRIALS: As a condition for publication of a clinical trial in AIDS, registration of the trial in a public registry is required. Registration of a trial must be at or before the enrollment of participants. The editors of AIDS also do not advocate one particular registry but require that the registry utilized meet the criteria set out in the statement of policy of the ICMJE (www.icjme.org).

Please state that your article includes a clinical trial and that the conditions of submission above have been met.

► No

Other information for the Editor:
<table>
<thead>
<tr>
<th><strong>Name of person completing this form:</strong></th>
<th><strong>Olivier Nuñez</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Date:</strong></td>
<td><strong>Friday 2 March 2018</strong></td>
</tr>
</tbody>
</table>