**Supplementary Digital Content**

**Supplementary Methods**

***Description of the method***

Each year, a proportion of the undiagnosed population develops clinical AIDS [1] and is diagnosed with HIV at the same time (referred to as simultaneous HIV/AIDS diagnosis). Another part of the undiagnosed population can be diagnosed before developing clinical AIDS (because of risky behaviors leading to HIV test or through mandatory/suggested testing, e.g., blood donation, pregnancy, check-up, etc.). A fraction of them would have developed clinical AIDS that year if they had not been diagnosed and had not benefited from antiretroviral therapy (i.e. fraction prevented). Thus, the number of the observed simultaneous HIV/AIDS diagnoses summed to the fraction prevented can be used to estimate the size of the undiagnosed population if the progression rate to AIDS for this population is available, as a sort of simplified back-calculation.

Given that the progression to AIDS depends mainly on the CD4-count and that the distribution of this parameter is unknown in the undiagnosed HIV-population, we assumed that the CD4 count distribution of the undiagnosed population was similar to that of asymptomatic subjects reported to the HIV surveillance system. A newly diagnosed subject was defined as asymptomatic in absence of HIV-related symptoms according to the 1993 CDC classification system.

We assume that annual progression to clinical AIDS in HIV undiagnosed people could be approximated by that estimated in untreated patients before introduction of antiretroviral therapies, as derived by CASCADE Collaboration [2].

On the basis of the previous considerations, the annual number of cases with undiagnosed HIV, denoted by ***n***, can be estimated by the following formula:

where ***pi*** is the proportion of patients in CD4 stratum***i*** among newly diagnosed asymptomatic HIV, **ri** is the annual AIDS rate in CD4 stratum ***i***, ***k*** is the number of CD4 cell count classes considered, ***S***is the estimate of simultaneous HIV/AIDS cases, obtained as the sum of simultaneous HIV/AIDS diagnoses reported to surveillance (*O*) and the estimated prevented fraction (*P*).

We excluded new HIV diagnoses in those <15 years old, usually infected by vertical transmission and for whom the previous assumptions could not be valid.

***How to estimate the prevented fraction (P)***

We should consider at the numerator not only those with simultaneous HIV/AIDS diagnosis but also the estimated number of simultaneous HIV/AIDS diagnosis averted because of diagnosis (meaning, treatment). In particular, for each person reported to surveillance (without a simultaneous HIV/AIDS diagnosis) was calculated a probability to develop AIDS within the year. This individual probability was computed considering his/her CD4-count at diagnosis and the days (at risk) from the date of HIV diagnosis to the end of the year.

For the *h-th* newly diagnosed person without simultaneous HIV/AIDS, having CD4-count in the *i-th* stratum, the probability of developing clinical AIDS within the end of the year was calculated by the following formula [3]:

For example a person diagnosed without clinical AIDS the 15th of march 2012 and having a CD4 count of 120 cells/mm3 had a probability to develop AIDS within the year equal to

Where 291 is the days between the 15th of March and the end of the year, and 0.21992 is the annual AIDS rate in the CD4 stratum 100-149 cells/mm3 as reported in Table 1. The prevented fraction is then the following:

Where, derive from (2) and *m* is the total of new diagnoses reported to surveillance without a simultaneous HIV/AIDS diagnosis (i.e. asymptomatic stage or symptomatic non-AIDS stage).

***Under reporting/ascertainment of HIV/AIDS***

In Italy to date, there are no precise estimates about under-reporting/ascertainment of both AIDS and HIV diagnoses, hence we provide three reliable scenarios: no under-reporting/ascertainment, 10% and 20% under-reporting/ascertainment. We based our decision on the results reported for Lazio region [4], where a HIV surveillance system was set up in 1985, the longest period of observation available in Italy.

We assumed that under-reporting/ascertainment of both AIDS and HIV diagnoses was not different by clinical stage, CD4 cell count at diagnosis and the other characteristics considered here.

***Multiple imputation and estimate of uncertainty***

**Surveillance data presents missing information about the HIV exposure category, CD4 cell count and** presence/absence of HIV-related symptoms or clinical AIDS **at HIV diagnosis.**

**We assumed that they were missing at random (i.e. the probability of being missing depends only on the covariates we can observe).**

**Using logistic regression models (data not shown) we evaluated that the significant predictors (p<0.05) of missing exposure categories among females were age (two categories: ≥35 or <35), country of birth (Italy or other countries), clinical stage (un-reported or reported) and CD4 cell count at diagnosis (three categories: un-reported, <350 or ≥350 cells/mm3); only nationality was predictive among males. Significant predictors (p<0.05) of missing CD4 cell counts were: HIV exposure category (un-reported, reported), clinical stage (un-reported, reported) and country of birth (Italy or other countries). Finally, significant predictors of a missing clinical stage were: age (four categories: <25, 25-34, 35-44 or ≥45), HIV exposure category (un-reported, reported), country of birth (Italy or other countries), CD4 cell count (four categories: <100, 100-199, 200-499 or ≥500).**

**Missing data were estimated by a multiple imputation procedure [5] creating 1,000 imputed datasets from the original data; in each of those, we first imputed the missing HIV exposure category (based on age, country of birth, clinical stage and CD4 count); then, conditionally on the previously imputed HIV exposure category, missing CD4 cell count (based on age, HIV exposure category, clinical stage and country of birth), and thereafter, the missing clinical stage (based on age, HIV exposure category and CD4 cell count) was imputed.**

Two sources of uncertainty have to be considered: one deriving from the multiple imputation of missing values, and the second deriving from the annual progression to AIDS rate by CD4 cell count. Thus, an (), obtained by random generation from the Normal distribution, was associated to each imputed dataset *j (j=1,..,1,000)* with a mean equal to the AIDS rate *ri* and a standard deviation equal to its standard error *.*

Applying formulas (1) and (2) to each imputed dataset, we obtained *P,* , , and *n*.

The reported estimates were the median values from the 1,000 datasets; 95% confidence intervals were the 2.5 and the 97.5 percentile of the ordered distributions obtained from the 1,000 datasets.

Estimates were also stratified by gender, HIV exposure category, age group and nationality.

**References**

1. Ancelle-Park RA. Expanded European AIDS case definition. *Lancet* 1993;341-441.
2. Guiguet M, Porter K, Phillips A, Costagliola D, Babiker A. Clinical progression rates by CD4 cell category before and after the initiation of combination antiretroviral therapy (cART). Open AIDS J. 2008;2:3-9.
3. Phillips A, Pezzotti P; CASCADE Collaboration. Short-term risk of AIDS according to current CD4 cell count and viral load in antiretroviral drug-naïve individuals and those treated in the monotherapy era. *AIDS*. 2004 Jan 2;18(1):51-8.
4. Porta D, Perucci CA, Forastiere F, De Luca A; Lazio HIV Surveillance Collaborative Group. Temporal trend of HIV infection: an update of the HIV surveillance system in Lazio, Italy, 1985-2000. *Eur J Public Health*. 2004 Jun;14(2):156-60.
5. Rubin DB. *Multiple Imputation for nonresponse in surveys*. Wiley: New York; 1987.

**Supplementary tables**

**Supplementary Table 1.** Main characteristics of persons with a new HIV diagnosis; Italy, 2012.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **N** | **%** | **% excluding missing** |
| **Gender** | *Female* | 861 | 21.1 | 21.1 |
|  | *Male* | 3,221 | 78.9 | 78.9 |
|  |  |  |  |  |
| **Age** | *15-24* | 325 | 8.0 | 8.0 |
|  | *25-34* | 1,298 | 31.8 | 31.8 |
|  | *35-44* | 1,224 | 30.0 | 30.0 |
|  | *45-54* | 874 | 21.4 | 21.4 |
|  | ≥*55* | 361 | 8.8 | 8.8 |
|  |  |  |  |  |
| **Exposure category** | *MSM* | 1,562 | 38.3 | 44.4 |
|  | *Male heterosexual* | 1,051 | 25.8 | 29.9 |
|  | *Female heterosexual* | 693 | 17.0 | 19.7 |
|  | *IDU* | 212 | 5.2 | 6.0 |
|  | *Missing information* | 564 | 13.8 | - |
|  |  |  |  |  |
| **Country of birth** | *Italy* | 2,991 | 73.3 | 73.3 |
|  | *Other countries* | 1,091 | 26.7 | 26.7 |
|  |  |  |  |  |
| **Symptoms** | *Asymptomatic* | 1,800 | 44.1 | 61.6 |
|  | *Non-AIDS symptoms* | 478 | 11.7 | 16.3 |
|  | *Clinical AIDS* | 646 | 15.8 | 22.1 |
|  | *Missing information* | 1,158 | 28.4 | - |
|  |  |  |  |  |
| **CD4 cell count (cells/mm3)** | *<100* | 769 | 18.8 | 24.0 |
|  | *100-199* | 415 | 10.2 | 13.0 |
|  | *200-349* | 583 | 14.3 | 18.2 |
|  | *350-499* | 575 | 14.1 | 18.0 |
|  | *≥500* | 857 | 21.0 | 26.8 |
|  | *Missing information* | 883 | 21.6 | - |
|  |  |  |  |  |
| **Total** |  | 4,082 |  |  |

*Note: 18 HIV diagnoses in people <15 years of age were excluded from this table;*

*MSM: Men who have sex with men;*

*IDU: Injecting drug users.*

**Supplementary Table 2.** Percentages of undiagnosed HIV+ prevalence estimates in Italy according to HIV exposure category, gender, country of birth and age; Italy 2012.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **HIV+ diagnosed (A)** | **HIV+ undiagnosed (B)** | **Overall prevalence HIV+ (C=A+B))** | **% HIV+ undiagnosed [(B/C)x100]** |
| **In care\*** | **Not retained\*\*** | **Total** | **No under-** **reporting/****diagnosis** | **10% under- reporting/****diagnosis** | **20% under- reporting/****diagnosis** | **No under- reporting/****diagnosis** | **10% unde- reporting/diagnosis** | **20% under- reporting/****diagnosis** | **No under- reporting/****diagnosis** | **10% under- reporting/diagnosis** | **20% under- reporting/****diagnosis** |
| **Total** | 94,146 | 18,076 | 112,222 | 13,729 | 15,102 | 16,475 | 125,951 | 127,324 | 128,697 | 10.9 | 11.9 | 12.8 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **HIV exposure Category** |  |  |  |  |  |  |  |  |  |  |  |
| *IDU* | 28,714 | 5,513 | 34,227 | 895 | 984 | 1,074 | 35,122 | 35,211 | 35,301 | 2.5 | 2.8 | 3.0 |
| *Heterosexual*  | 37,511 | 7,202 | 44,713 | 6,291 | 6,919 | 7,549 | 51,004 | 51,632 | 52,262 | 12.3 | 13.4 | 14.4 |
| *MSM* | 27,921 | 5,361 | 33,282 | 6,382 | 7,020 | 7,659 | 39,664 | 40,302 | 40,941 | 16.1 | 17.4 | 18.7 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Gender**  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Male* | 66,014 | 12,675 | 78,689 | 11,352 | 12,488 | 13,623 | 90,041 | 91,177 | 92,312 | 12.6 | 13.7 | 14.8 |
| *Female* | 28,132 | 5,401 | 33,533 | 2,500 | 2,750 | 3,000 | 36,033 | 36,283 | 36,533 | 6.9 | 7.6 | 8.2 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Country of birth** |  |  |  |  |  |  |  |  |  |  |  |  |
| *Italy* | 79,358 | 15,237 | 94,595 | 10,416 | 11,457 | 12,499 | 105,011 | 106,052 | 107,094 | 9.9 | 10.8 | 11.7 |
| *Abroad* | 14,788 | 2,839 | 17,627 | 3,358 | 3,694 | 4,030 | 20,985 | 21,321 | 21,657 | 16.0 | 17.3 | 18.6 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Age group** |  |  |  |  |  |  |  |  |  |  |  |  |
| *< 25 years* | 3,047 | 585 | 3,632 | 1,189 | 1,308 | 1,427 | 4,821 | 4,940 | 5,059 | 24.7 | 26.5 | 28.2 |
| *25-49 years* | 59,754 | 11,473 | 71,227 | 9,392 | 10,331 | 11,270 | 80,619 | 81,558 | 82,497 | 11.6 | 12.7 | 13.7 |
| *50-59 years* | 23,609 | 4,533 | 28,141 | 1,746 | 1,921 | 2,095 | 29,887 | 30,062 | 30,237 | 5.8 | 6.4 | 6.9 |
| *≥ 60 years* | 7,736 | 1,485 | 9,222 | 1,207 | 1,328 | 1,448 | 10,429 | 10,550 | 10,670 | 11.6 | 12.6 | 13.6 |

*\*Data from the 2012 Italian National survey from all infectious disease clinics (see Camoni et al. AIDS Res Hum Retroviruses 2015;31:282-7);*

*\*\*Estimate of 19.1% out of the total diagnosed from three Italian clinical infectious disease units,8th Italian Conference on AIDS and Retroviruses 2015*

**Supplementary Table 3.** Comparison ofCD4 distribution used to estimate undiagnosed HIV people in Italy with those estimated among undiagnosed in published studies.

|  |  |  |  |
| --- | --- | --- | --- |
| **CD4** | **CD4 at diagnosis in newly diagnosed asymptomatic Italian HIV cases (2012)** |  | **Studies reporting estimates of CD4 distribution among undiagnosed** |
| **Reported data** |  | **Reported data increased by 15%** |  | **Birrell 2013, Lancet Inf Dis (England & Wales)** | **Supervie 2013, AIDS (France)** | **van Sighem 2015, Epidemics (The Netherlands)** |
| **MSM** | **Non MSM** | **All** |  | **MSM** | **Non MSM** | **All** |  | **MSM**  | **All** | **MSM** | **MSM**  |
| **<200** | 11% | 21% | 17% |   | 8% | 17% | 13% |   | 5% | 19% | 18% | 28%\* |
| **200-349** | 18% | 23% | 21% |   | 13% | 18% | 16% |   | 15% | 21% | 19% |
| **350-499** | 27% | 22% | 24% |   | 22% | 21% | 22% |   | 25% | 20% | 20% | 18% |
| **≥500** | 44% | 34% | 38% |   | 57% | 44% | 49% |   | 55% | 40% | 41% | 54% |

*\* this study reports % estimates for MSM <350 CD4*