Supplementary information

Methods based on prevalence surveys

Availability of methods for use

The Workbook and EPP programs implementing this approach can be used in conjunction with UNAIDS[1] (contact: estimates@unaids.org). The advantage of EPP over the Workbook is that it makes explicit use of trends in prevalence over time, allowing for the effect of antiretroviral treatment. These programs have a user-friendly interface and are designed to be used by those without statistical or programming expertise. EPP in conjunction with Spectrum allows the estimation of the number of people living with HIV, new HIV infections, and annual HIV-related deaths, as well as the number of people in need of treatment.

The MPES method[2-4] has been developed through a collaboration between the University of Bristol and the MRC Biostatistics Unit at the University of Cambridge in the United Kingdom and is programmed in WinBUGS. It is not currently available in a user-friendly format. To some extent this reflects the fact that the understanding of the sources of the data and where possible biases or inconsistencies can arise is not automatic. This process of understanding entails a close collaboration between statisticians, epidemiologists and data providers, and an iterative nature to the work. The initial model fit is appraised, with detected conflicts leading to evaluations of the possible biases. These are modelled, perhaps using further external evidence, and the model re-appraised. In principle, the group who have developed the method are willing to consider collaborations with individual countries to implement the method (contact: daniela.deangelis@mrc-bsu.cam.ac.uk).

Methods based on reporting of HIV/AIDS diagnoses involving calculation of cumulative incidence of HIV

Cambridge method [5]

Sweeting et al. (MRC Biostatistics Unit, Cambridge, United Kingdom) describe Bayesian back-calculation using a multi-state model, and apply it to United Kingdom Health Protection Agency (HPA) data on HIV in MSM[5]. The method is re-considered by Birrell and applied to an updated dataset[6]. The data required are number of new HIV diagnoses per calendar quarter, with data on AIDS diagnosis occurring in the same calendar quarter (late diagnosis). CD4 counts around diagnosis for a subset of the diagnoses are strongly recommended but not essential, and data should be stratified by risk group.

Undiagnosed HIV prevalence is modelled at the population level using a uni-directional multi-state model. The disease states, defined by CD4 count, are: early disease; intermediate disease; advanced disease; HIV diagnosis; and AIDS diagnosis. Prevalence in any CD4 state in any quarter is assumed to be dependent only on prevalence in the same or higher CD4 states in the previous quarter. In particular, it is assumed that the number of AIDS diagnoses in any quarter is dependent only upon the number of individuals with CD4 < 200 cells/mm^3 in the previous quarter. Late diagnosis is assumed to equate to diagnosis of AIDS. It is assumed that the rate of HIV testing depends on CD4 stage and on calendar year, and is the same across a wide CD4 count range. Patients with a CD4 count available at diagnosis are assumed to be representative of all diagnosed individuals. The

implementation by Birrell[6] assumes that AIDS cases are underreported from the year 2000 onwards by some factor which is estimated, subject to the input of strong prior information. Quarterly progression probabilities between states are assumed known from external data, and rates of diagnosis of HIV and of under-reporting of AIDS diagnoses are simultaneously estimated with incidence of HIV. Due to the lack of identifiability in distinguishing changes in diagnosis rates from changes in incidence, some estimates will be very imprecise unless data on the distribution of CD4 count around diagnosis are incorporated into the model.

Atlanta method [7]

Hall et al. (Center for Disease Control and Prevention, Atlanta, United States) describe their extended back-calculation method and use it to generate estimates of HIV incidence in the United States, and compare with estimates obtained using assays that differentiate between recent and longstanding infection[7]. The extended back-calculation method was subsequently used to obtain estimates of HIV prevalence in the United States, including estimates of undiagnosed HIV prevalence[8;9]. In addition to demographic information, the data required are the number of new HIV diagnoses per calendar year with information on whether AIDS was diagnosed within the same calendar year as HIV (disease severity).

It is assumed that the HIV testing rate depends only on calendar year, and not time since infection, and that the AIDS rate depends only on time since infection, not calendar year. The values of the testing hazard and the number of infections are assumed to be respectively constant within two specified sets of calendar periods.

The method consists of a discrete-time probability model with parameters: number of infections per year; AIDS diagnosis hazards; HIV testing hazards. To account for

missing data on HIV diagnoses, people who develop AIDS during the year are assumed to be tested for HIV and classified as a new HIV diagnosis (with AIDS) in the year. The AIDS diagnosis hazards used are completely specified rather than estimated, using values obtained from a Markov model. The HIV testing hazards are estimated, and are assumed to depend only on calendar year. The unknown parameters in the back-calculation model are estimated using an expectation maximization algorithm. This alternates between calculating an expanded version of the observed dataset which is consistent with the specified model structure and with parameter values in the current iteration (expectation step), and re-estimating the parameter values (maximization step). While the observed dataset contains the number of diagnoses by year of diagnosis and disease severity, the expanded dataset contains the number of diagnoses by year of infection, disease severity, and year of diagnosis. There are more complicated versions of the model which allow any one or possibly several of the assumptions to be relaxed. For example, the model can be further extended to allow for: estimation of a testing hazard that depends on both calendar time and time since infection; partial or complete estimation of the AIDS hazards; or varying shapes for the incidence curve.

Ottawa / Sydney method [10]

Wand et al. (National Centre in HIV Epidemiology and Clinical Research, Sydney, Australia) describe and adopt a methodology developed by colleagues in the Public Health Agency of Canada. Their extended back-calculation method is used to reconstruct the HIV epidemic in Australia in the MSM, PWID, and heterosexual exposure categories[10]. The same method was also used in different provinces in Canada, to determine the national HIV incidence and prevalence[11]. The data required are HIV diagnoses, with additional data on whether the HIV infection is recent or not using either enhanced surveillance (evidence of a prior negative test, or

a diagnosis of seroconversion illness, or an indeterminate western blot within one year of HIV diagnosis), or using laboratory techniques. This methodology does not require a test for a biomarker such as CD4 count. The method also uses data on AIDS diagnoses in years before effective treatment was available.

As with the Atlanta method (and effectively the Cambridge method), the hazard of testing for HIV is modelled as a two-component process: testing while asymptomatic, and testing due to clinical symptoms in later stages of HIV progression. The testing rate in asymptomatic people is assumed to be constant for each person, with the constant probability allowed to differ between individuals, unlike in the other methods. The testing rate with symptoms is assumed to follow a distribution similar to progression to CD4 count below 200 cells/mm^3 without treatment. Two sub-models based on these assumptions are constructed: they are mathematically connected to form the combined progression rate distribution. The HIV incidence curve is then re-constructed by combining two back-projection estimated HIV incidence curves from AIDS diagnostic data (up to 1994, prior to which effective antiretroviral treatment was not available) and HIV diagnostic data using the combined progression rate distribution.

Paris method [12]

Ndawinz et al. (INSERM U943, Paris, France) describe an extended back-calculation method, which is used to estimate both the incidence of HIV infection in France and the time-dependent intervals of time from infection to diagnosis in different transmission categories [12]. If HIV and AIDS case surveillance has been in place for some time, the method can also be used to estimate the HIV prevalence and the number of undiagnosed people with HIV. The data required are times of HIV diagnosis, risk category and clinical status at diagnosis divided into three categories:

primary infection, AIDS and other clinical statuses. Multiple imputation is used for missing data.

Individuals diagnosed in primary infection are assumed to be those who seek an HIV test following HIV exposure, whereas those diagnosed in later stages of the disease are assumed to be tested for other reasons. For those diagnosed during primary infection, the time between infection and diagnosis is assumed to be uniformly distributed over the first six months after infection. For those diagnosed outside of primary infection, the time from infection to diagnosis is assumed to depend on both the rate of natural progression to AIDS, and the rate of pre-AIDS HIV testing. The median time from infection to AIDS has been estimated from cohort studies and is assumed to be ten years [13], and the rate of pre-AIDS HIV testing is assumed to depend on two unknown parameters that represent uptake of both routine testing and symptom-driven testing. The expectation-maximization-smoothing (EMS) algorithm is used to estimate the annual HIV incidence, and the Newton-Raphson method [14] is used to estimate the two unknown parameters of the distribution of the pre-AIDS HIV-testing rate. Incorporating the clinical status at diagnosis into the back-calculation method allows for heterogeneity in HIV testing behaviour, and for detecting changes in the testing rate over time.

Bordeaux method [15]

The method described by Sommen et. al. (INSERM U897, Bordeaux, France) is based on a Markov model which, unlike the other methods in this section, models treatment uptake. The method is illustrated using HIV/AIDS surveillance data on MSM in France. The data required are HIV and AIDS diagnosis data. The method is described in a context where HIV diagnosis data is only available for the most recent

years, but can be adapted for situations where HIV diagnosis monitoring has been in place for longer.

Disease progression rates (with and without treatment) are assumed known from previous studies. Treatment in the population they consider is assumed to have been available from 1995 onwards with a certain specified constant yearly uptake in people with asymptomatic HIV and a higher uptake in people with symptomatic HIV. The effect of treatment is also assumed known from other studies. The median time to AIDS for people treated before 1995 is assumed to be 13.5 years. For the period 1996-2005, the incubation period for treated patients is assumed to be 36.3 years, based on cohort data. Mortality in people with asymptomatic HIV is assumed to be the same as in the general population, and mortality in people with symptomatic HIV is assumed to be three times higher.

In the Markov model, transitions are assumed to occur from asymptomatic HIV to symptomatic HIV and then to AIDS within each of the following states: undiagnosed, diagnosed but untreated, and treated. Individuals who have not developed AIDS may also move from undiagnosed to diagnosed to treated. The transition probabilities between the various states change according to calendar time. The main way in which this method differs from others in this section is that it accounts for the effect of treatment and therefore makes use of data on diagnosis of AIDS in people who are already diagnosed with HIV.

Availability of methods for use

WinBUGS and JAGS code for implementation of the Cambridge method is to be made available upon publication of Birrell [6], a paper is in preparation.

The current implementation of the back-calculation model used in the Atlanta method is not suitable for general usage at this point since there is no 'user-friendly' interface to serve as a guide through the process. The authors are currently willing to discuss how to implement their extended back-calculation model but emphasise the importance of ensuring that the data used as the input to the model satisfy the various model assumptions and reflect the actual HIV diagnoses that have occurred.

User-friendly software written in R and additional documentation for the Ottawa / Sydney method are available from Ping Yan (ping_yan@phac-aspc.gc.ca).

The Paris method has been implemented in the C programming language, and is not currently available in a user-friendly format. In principle, the authors are willing to consider collaborations with individual countries to implement the method. Contact: Virginie Supervie (virginie.supervie@ccde.chups.jussieu.fr).

The Bordeaux method has been implemented using a program written in Fortran. In principle, the authors would consider adapting to a more user friendly programming format with appropriate support. Contact: Ahmadou Alioum (alioum.ahmadou@isped.u-bordeaux2.fr).

Methods using number of reported simultaneous HIV/AIDS diagnoses[16;17]

London method 1

This requires the CD4 count at HIV/AIDS diagnosis. For each CD4 count stratum, the number of people with undiagnosed HIV can be estimated by dividing the number of simultaneous HIV/AIDS diagnoses in that stratum by the CD4-specific AIDS rate. Summing across all strata gives the total number with undiagnosed HIV. For high

CD4 count strata in which the AIDS rate is low the estimates will be associated with considerable uncertainty.

London method 2

This method assumes that CD4 count in the undiagnosed population can be approximated by the CD4 count at diagnosis in patients presenting for care with asymptomatic HIV. Consequently data may also be required on whether patients are asymptomatic at HIV diagnosis, if this information is not available from appropriate cohort studies. For each CD4 count stratum, the number of people with a simultaneous HIV/AIDS diagnosis is estimated by using the total number of people with a simultaneous HIV/AIDS diagnosis across all CD4 count strata and the assumed distribution of CD4 counts in the undiagnosed population. This estimated number with simultaneous HIV/AIDS diagnoses in each CD4 count stratum is then divided by the CD4-specific AIDS rate to give an estimate of the number of people with undiagnosed HIV: summing this across all CD4 count strata as before gives an estimate of the total number of people with undiagnosed HIV.

Availability of methods for use

The methods are straightforward to implement. CD4-specific AIDS incidences can be obtained from cohort studies (and do not need to be estimated separately in the country, although this is an option), as can the distribution of first CD4 count in patients newly-diagnosed with asymptomatic HIV. These methods are yet to be fully developed and have thus far been presented in conference poster format only.
Contact: Rebecca Lodwick (r.lodwick@ucl.ac.uk).

References

1. UNAIDS. Epidemiological software and tools (2009). http://www.unaids.org/en/KnowledgeCentre/HIVData/Epidemiology/EPI_software2009.asp

2. 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; 171:541-567

3. Presanis AM, De Angelis D, Spiegelhalter DJ, Seaman S, Goubar A, Ades AE. Conflicting evidence in a Bayesian synthesis of surveillance data to estimate human immunodeficiency virus prevalence. Journal of the Royal Statistical Society Series A-Statistics in Society 2008; 171:915-937.

4. Presanis AM, Gill ON, Chadborn TR, Hill C, Hope V, Logan L et al. Insights into the rise in HIV infections, 2001 to 2008: a Bayesian synthesis of prevalence evidence. AIDS 2010; 24:2849-2858.

5. Sweeting MJ, De Angelis D, Aalen OO. Bayesian back-calculation using a multi-state model with application to HIV. Statistics in Medicine 2005; 24(24):3991-4007.

6. Birrell P, De Angelis D, Chadborn TR, Sweeting MJ. Determining trends in incidence using Bayesian backcalculation: the HIV epidemic among homosexual men in England and Wales. 2010. (unpublished)

7. Hall HI, Song RG, Rhodes P, Prejean J, An Q, Lee LM et al. Estimation of HIV incidence in the United States. Journal of the American Medical Association 2008; 300(5):520-529.

8. Campsmith ML, Rhodes P, Hall HI, Green T. HIV Prevalence Estimates-United States, 2006 (Reprinted from MMWR, vol 57, pg 1073-1076, 2008). Journal of the American Medical Association 2009; 301(1):27-29.

9. Campsmith ML, Rhodes PH, Hall HI, Green TA. Undiagnosed HIV Prevalence Among Adults and Adolescents in the United States at the End of 2006. Journal of Acquired Immune Deficiency Syndromes 2010; 53(5):619-624.

10. Wand H, Yan P, Wilson D, McDonald A, Middleton M, Kaldor J et al. Increasing HIV transmission through male homosexual and heterosexual contact in Australia: results from an extended back-projection approach. HIV Medicine 2010; 11(6):395-403.

11. Yang q, Boulos D, Yan P, Zhang F, Remis RS, Schanzer D et al. Estimates of the number of prevalent and incident human immunodeficiency virus (HIV) infections in Canada, 2008. Can J Public Health 2010l 101(6):486-490.

12. Ndawinz JDA, Costagliola D, Supervie V. Recent Increase in the Incidence of HIV Infection in France. 17th Conference on Retroviruses and Opportunistic Infections. San Francisco, California, USA, 16th-19th February 2010.

13. Becker NG, Lewis JJC, Li Z, McDonald A. Age-specific back-projection of HIV diagnosis data. Statistics in Medicine 2003; 22:2177-2190.

14. Kreyszig E. Advanced Engineering Mathematics. 7th ed. John Wiley & Sons, New York, 1993, pp. 929.

15. Sommen C, Alioum A, 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(11):1554-1568.

16. Lodwick RK, Sabin CA, Phillips AN. Estimation of the number of undiagnosed HIV-positive people within a region based on surveillance of simultaneous HIV/AIDS diagnoses. 12th European AIDS Conference. Cologne, Germany, 11th-14th November 2009.

17. Lodwick RK, Sabin CA, Phillips AN. A method to estimate the number of people in a country or region with HIV who are undiagnosed and in need of ART. 10th International Congress on Drug Therapy in HIV Infection. Glasgow, UK, 7th-11th November 2010.