Estimation and projection of HIV infection and AIDS cases

James Chin

Division of Epidemiology, University of California, Berkeley, School of Public Health, Department of Biomedical and Environmental Health Sciences, 456 Kentucky Avenue, Berkeley CA 94707, USA

HOW many HIV infections and AIDS cases have occurred in the past, are currently present, and are likely to occur in the future? Reliable answers to these questions are needed to evaluate the spread of HIV and AIDS and to help plan public health, health care, and social welfare programmes vital for an appropriate response to this ever-expanding pandemic. However, since the early 1980s, when AIDS was first recognized, the pandemic's future trends and ultimate dimensions have been shrouded in uncertainty. This uncertainty persists because of the difficulties in measuring the prevalence, and especially the incidence, of AIDS cases and HIV infections in a given population with any substantial degree of precision. As a result, investigators have used many methods and models in attempts to understand the dynamics and interrelationships of the major determinants of HIV transmission and develop reliable estimates and projections of HIV and AIDS. The methods and models used need to be examined critically and understood before their outputs are accepted and used for public health programmes or policy decisions. This paper provides an overview of the uses and limitations of the methods available for HIV and AIDS estimates and projections. These methods are listed in Table 1.

Public health surveillance data[1–12]

Virtually all countries collect data on AIDS cases, primarily through public health surveillance systems that rely on passive notification of case reports. Reports of AIDS cases from industrial countries of Europe, North America and Oceania have been, up to 1993, based primarily on the 1987 Centers for Disease Control/World Health Organization (WHO) case definition[1]; those from Africa have generally been based on nationally adapted versions of the WHO clinical (Bangui) definition[11]; several countries in Latin America have used the Pan American Health Organization's Caracas definition[6]; and those from other countries involve a combination of these definitions. In addition to the problem posed by varying definitions of AIDS cases, the data are flawed by significant problems related to the wide variations found in the accuracy, completeness and timeliness of different national AIDS reporting systems. However, reported AIDS case data can provide the basis for estimating the actual number of cases that may have occurred. Researchers have prepared such estimates by adjusting reported AIDS data to correct for delays, inaccuracies and incompleteness. However, these adjustments require specific studies to determine the precise degree of correction needed for each limiting factor, and many estimators assume erroneously that these factors do not change with time. Hence, the reliability of AIDS estimates derived after such adjustments varies.

Detection of HIV infection became possible with the development of HIV antibody tests, which were first available in 1985. Most public health surveillance data on HIV infection have been obtained from a variety of uncoordinated sources, and much of the data were gathered from samples of selected population groups. Thus, data obtained from different areas or studies are usually not

Table 1. Methods for estimating and projecting of HIV infections and AIDS cases

Extrapolation from public health surveillance HIV/AIDS data
Delphi survey method
HIV and AIDS models
Type I models: statistical extrapolation from reported AIDS case data
Type II models: simple projection models
Type III models: complex deterministic/compartmental models
Combination of methods
Scenario/modelling approach

comparable. Current serological tests for detecting HIV antibodies are reliable for measuring HIV prevalence in populations: collectively they have shown that HIV infection is not randomly distributed in a given population. Relatively high prevalence rates (from a few per cent to more than 50) can be found among homosexual or heterosexual individuals who have had multiple sexual partners and those people who share drug injection equipment. The prevalence of HIV infection in most general populations of industrial countries is low and may range from almost zero to one per several thousand.

Researchers have estimated HIV prevalence most often by using the results of serological surveys and extrapolating the data to specific populations. Major problems with this method are the limited number of seroprevalence studies that are representative of specific populations or subgroups, and the wide variability in estimates of the size of important subgroups, such as female prostitutes, injecting drug users, and patients seen in sexually transmitted disease clinics. Nevertheless, investigators have derived reasonable working estimates of the prevalence, general distribution, and trends of HIV infection for many countries by analysis of all available HIV serosurvey data.

In the mid- to late-1980s, with only limited HIV data available, investigators tended to overestimate HIV prevalence for many areas and populations. In the United States for example, during the period 1985–1992 official estimates of cumulative HIV infection remained consistently in the range 1.0–1.5 million. As of early 1995, the best estimate of cumulative HIV infections was still between 1.0 and 1.5 million, which clearly indicates that all previous estimates were too high. In the United Kingdom in the mid-1980s HIV prevalence was estimated at 80,000. Subsequent reviews and analyses of the accumulating HIV data by expert committees resulted in a consistent downward revision of this estimate to a maximum of about 30,000. In 1987, WHO's estimate of global HIV infection ranged from 5 to 10 million. With the collection and analysis of more recent HIV data (primarily from sub-Saharan Africa) WHO's revised 1990 estimate of global HIV infection was reduced to the lower level of the 1987 estimate, i.e. about 5–6 million. Since the late 1980s, WHO's global and regional HIV estimates have been intentionally conservative.

The Delphi survey method[13–15]

The Delphi survey method was developed in an attempt to improve the reliability of the judgments needed in relatively uncertain situations and to provide a means of quantifying such judgements. This method was originally systematized in the 1960s for use in developing management policies. It was first applied to the health field in the 1970s. Essentially, the Delphi method obtains educated guesses from selected experts in a reiterative fashion, and then uses the average and range of the Delphi responses as projections. Major advantages of the Delphi method are speed and low cost. However, it is difficult to select truly knowledgeable experts to estimate the number of HIV infections and AIDS cases. Furthermore, estimates and projections made by the Delphi method may have extremely wide ranges.

In 1988 WHO used a Delphi survey to obtain projections of HIV prevalence. Delphi survey respondents projected that there would be 15–20 million cumulative adult HIV infections worldwide by the year 2000. Yet shortly after that survey, HIV serological data for sub-Saharan Africa and southeast Asia showed marked increases in HIV seroprevalence. These increases, estimated at about 5 million new infections within two to three years after the survey, suggested that the 1988 Delphi projections were too low. In 1991, based on observed HIV trends, WHO projected that a cumulative total of 30 to 40 million HIV infections would occur in men, women, and children by the year 2000.

In mid-1992 a Harvard group, AIDS in the World (AIW)[15], also carried out a Delphi survey to project cumulative global HIV infections by the year 2000. While AIW's low Delphi total was only slightly higher than WHO's conservative projection in 1991, its high Delphi projection, more than 100 million cumulative HIV infections by the year 2000, was three to four times higher than their low Delphi projection. With the exception of AIW's high Delphi projection (for which no comparable WHO projection was made) the independent estimates and projections of HIV infections and AIDS cases were reasonably close to WHO's projections. AIW did not select either the low or high Delphi projection as their best estimate for the year 2000, but most references to its Delphi projections focus on their high Delphi projection.

How plausible are AIW's high Delphi projections? If cumulative infections in North America (primarily the United States) were estimated to be about 1.2 million as of 1993, then almost 1 million new HIV infections would have to occur each year in North America to reach AIW's high Delphi projection of more than 8 million. Similarly, in sub-Saharan Africa, more than 3 million new HIV infections would have to take place annually to reach the projected total of more than 33 million by the year 2000. For other regions, the AIW's projections require new HIV infections of about 1 million per year in Latin America, close to 1 million per year in the Caribbean, and almost 6 million per year in southeast Asia to reach 45 million by the year 2000 (Table 2). None of these annual incidences are plausible based on observed HIV data and trends, which suggest that with the exception of southeast Asia, which is still at an early stage

Table 2. AIDS in the world: HIV estimates and Delphi projections (millions)

Region	Cumulative estimate Jan 1992	Cumulative estimate Jan 1993	High Delphi projection by the year 2000	Annual incidence needed from 1993 to 2000
North America	1.167	1 261	8 150	0 984
Western Europe	0.718	0.834	2.331	0.214
Sub-Saharan Africa	7.803	10.985	33 609	3.232
Southeast Asia	0.675	3.567	45.059	5.927
Latin America	0.995	1.050	8.554	1.072
Caribbean	0.310	0.461	6.962	0.929
Total	11.799	18.311	104.665	12.336

in its HIV epidemic, the highest annual HIV incidence in most other regions may have already occurred.

HIV and AIDS models

Researchers have developed many types of HIV and AIDS models in attempts to understand the dynamics of HIV spread and to develop reliable projections of the epidemic's eventual extent. The major components of all models include the basic assumptions and input parameters used, the value or range of values for the input parameters, and the mathematical interrelationships between the parameters used in the model. The results produced by any model should not be used to formulate programmes or make policy decisions unless those making such decisions fully understand and agree with the basic assumptions, the values used for the input parameters and the mathematical interrelations between the input variables.

HIV and AIDS models can be broadly classified into three types that range from the simplest to the most complex. The first two types are empirical, extrapolation models, while models of the last type have been described as explanatory or deterministic. In general, the simpler extrapolation models are designed for short-term projections and the more complex models for evaluating determinants of the HIV and AIDS pandemic and for hypothesis testing.

Type I models[16–18]

Type I models use reported AIDS case data to make short-term (two to three years) projections of AIDS cases. Such short-term projections have been made by statistical extrapolation and regression techniques applied to the observed temporal curve of the reported AIDS cases. These models assume that after adjusting for reporting delays (and in some models for incomplete reporting), the trends of reported cases during the next few years will remain essentially similar to those observed in the recent past. A profusion of such models has been described, but a major problem with these models is that many mathematical curves can fit the data equally well. Not only is there no way to choose the best among the various curves, but the fitted curves may lead to widely divergent projections, particularly if they are used to make projections for periods longer than two to three years. Type I models should be used only to project AIDS cases two to three years ahead, and only for populations for which AIDS case reporting is relatively timely, reliable and complete.

Type II models[19,20]

Type II models use the data on estimated HIV infections and progression rates from infection to AIDS to calculate the number of past AIDS cases and to provide short-term (three to five years) projections of AIDS cases. A variation of this approach is the back calculation method, which uses AIDS case reports in conjunction with annual progression rates from infection to AIDS to estimate the number of annual HIV infections that have occurred. This method requires relatively accurate, complete and timely reporting of AIDS cases. However, because few people develop AIDS within two to three years of infection, this method cannot be used to estimate HIV infections that occurred during the prior two to three years even in areas where AIDS case reporting is considered reliable.

A simplified variation of this method uses an estimated ratio of HIV infections to AIDS cases to calculate total HIV infections. Like back calculation, this ratio method requires reliable estimates of AIDS cases, which are not available in most developing countries. In addition, users of the ratio method must realize that the ratio of HIV infection to AIDS cases changes rapidly with time. The ratio falls from many thousands to one during the first few years of extensive HIV spread in a population to less than ten to one after the first decade. This decline occurs whether HIV incidence is increasing or decreasing because a substantial proportion of HIV-infected individuals progress to AIDS as long as 10 to 20 years after contracting HIV infection.

Type III models[21–26]

Type III models are more complex and incorporate biological and behavioural variables that describe the transmission and natural history of HIV infection so as to simulate the entire disease process. Many mathematically sophisticated type III models have also been developed to project the future course of HIV and AIDS epidemics in different areas and populations. These complex models usually use differential equations to describe biological and behavioural variables known to be important in the transmission and progression of HIV disease in their simulation of the process from infection to progression to AIDS and death.

These deterministic or simulation models include: the mathematical model(s) of Anderson *et al.*[21]; the demographic model developed by Bulatao of the World Bank[25]; a Monte Carlo type model ('Simulaids') developed by Auvert *et al.*[22]; (1990); and an extremely detailed model – the interagency working group model – developed principally by Stanley and Seitz[26]. These models require, in varying degrees, extensive data sets on virtually all the demographic, biological and behavioural variables considered to be important in the epidemiology and natural history of HIV infection. The major problem with these models, aside from their complexity, is that most of the precise, detailed data sets they require to make projections are not available, even in those countries with the best data collection systems.

Most complex type III models have used hypothetical populations to obtain projections. In an exercise for which the population characteristics and a defined set of variables and input values were provided, a group of models produced widely varied, long-term demographic projections for HIV and AIDS. Projections for the percentage of the adult population infected with HIV in the year 2010 varied from a high of almost 40% in one model to a low of less than 3% in another. The greatest value of complex mathematical models may well be in their use to test hypotheses rather than for estimation and projection of HIV and AIDS prevalence.

A scenario/modelling approach

Given that predictions by complex models appear to be of doubtful validity, and because no criteria exist to select a good or best HIV and AIDS model (especially for projecting future HIV incidence and prevalence), a simpler scenario/modelling approach for estimation and projection of HIV infections and AIDS cases was developed by the author when he was at WHO during the late 1980s. The scenario/modelling approach can be used in individual countries and for selected populations within countries to provide working estimates and short-term projections of HIV-related morbidity and mortality for purposes of policy development and public health planning.

A scenario is defined as an outline of any series of events, real or imagined. Plausible HIV scenarios can be constructed from available HIV data and observed trends. Because future AIDS cases and deaths depend on the number and pattern of HIV infection and progression rates from infection to severe disease and subsequent death, AIDS cases and deaths can be calculated for different HIV prevalences and patterns (HIV scenarios) based on simple assumptions about the occurrence of these events.

The general methods used in the scenario/modelling approach to develop working estimates and projections of HIV and AIDS are as follows:

* The most recent patterns, prevalence and trends of HIV infection in specific population groups are estimated from an epidemiological analysis of all available HIV data. The HIV database compiled by the Center for International Research, US Bureau of the Census, is a valuable data source. Unpublished national HIV data, when available, should also be used to supplement these data.
* Based on these data and other epidemiological observations, different HIV scenarios can be estimated and projected to the year 2000 or beyond for major geographical and epidemiological regions. Optimistic or best-case scenarios should be constructed to derive conservative estimates and projections for public health planning.
* EPIMODEL, a relatively simple type II AIDS model developed by the author when he was with the WHO's Global Programme on AIDS can be used to estimate and project AIDS cases and deaths based on the HIV scenarios.

The reliability of HIV and AIDS projections obtained by this scenario/modelling approach depends primarily on two factors: the accuracy of the estimated number of HIV-infected people, and the proportion of HIV-infected people who ultimately develop AIDS.

The estimates and projections obtained from this scenario/modelling approach need to be modified as additional data on HIV prevalence and progression indicate that changes are warranted. However, health and social welfare systems should use conservative estimates and projections derived by this method to plan for and develop the resources needed to respond to the minimum number of HIV-related disease and deaths that can be expected during the remainder of this decade.

Summary and conclusions

The answer to the opening question 'How many HIV infections and AIDS cases have occurred, are currently present, and will occur in the future?' is that we really do not know, but we can estimate and project these numbers with varying degrees of uncertainty. Our ability to estimate past and current numbers is improving as public health surveillance systems collect more data.

Although the accuracy of such estimates varies greatly depending upon the quantity and quality of the surveillance data, plausible minimum, maximum and mid-range working estimates can be developed for use by public health and health care programmes.

Projection of future patterns and prevalence of HIV and AIDS is more difficult, but such projections or scenarios are essential to begin planning programmes to respond to the increasing number of premature deaths among young and middle-aged adults because of AIDS. A conservative HIV and AIDS scenario can be constructed from data on current global status and trends of regional epidemics. According to this scenario a cumulative total of about 40 million men, women and children would have been infected with HIV by the year 2000, with more than 90% of this total occurring in developing countries. By understanding the uses and limitations of all available methods and models that have been used for forecasting HIV and AIDS epidemics, such scenarios can assist policy-makers and social welfare and health care planners to prepare for the projected numbers of AIDS deaths and AIDS-related orphans.

1. Centers for Disease Control, Revision of the CDC surveillance definition for acquired immunodeficiency syndrome, *Morbid. Mortal. Weekly Rep.*, 1987, **36** (supplement 1S).
2. Chin, J., Public health surveillance of AIDS and HIV infections, *Bull. WHO*, 1990, **68**, 529–36.
3. Chin, J. and Mann, J. M., The global patterns and prevalence of AIDS and HIV infection, *AIDS 1988*, 1989, **2**, S247–S252.
4. Chin, J. and Mann, J. M., Global surveillance and forecasting of AIDS, *Bull. WHO*, 1989, **67**, 1–7.
5. Department of Health and the Welsh Office, Short-Term Prediction of HIV Infection and AIDS in England and Wales, Report of a Working Group (ISBN 011 321191 0), London.
6. Pan American Health Organization, Working Group on AIDS Case Definition, *Epidemiol. Bull. Pan Am. Health Organiz.* 1990, **10**, 9–11.
7. Public Health Laboratory Service, Acquired Immune Deficiency Syndrome in England and Wales to End 1993: Projections Using Data to End September 1989, Communicable Disease Report, (ISSN 0959-2547) pp. 1–12.
8. Public Health Laboratory Service, The Incidence and Prevalence of AIDS and Other Severe HIV Disease in England and Wales for 1992–1997: Projections Using Data to the End of June 1992, Communicable Disease Report 3 (ISSN 0264-1607) (Supplement 1): S1–S17.
9. Sato, P. A., Chin, J. and Mann, J. M., Review of AIDS and HIV infection: Global epidemiology and statistics, *AIDS 1989*, 1990, **3**, S301–S307.
10. United States Bureau of the Census, Center for International Research, AIDS/HIV Surveillance Database, Washington, DC, 1992.
11. World Health Organization, Report of a Workshop on AIDS in Central Africa. Bangui, Central African Republic, October 22–25, WHO/CDS/AIDS/85.1, 1985.
12. WHO, Global Programme on AIDS, Current and Future Dimensions of the HIV/AIDS Pandemic: A Capsule Summary, Geneva, 1992.
13. Chin, J., Sato, P. A. and Mann, J. M., Projections of HIV infections and AIDS cases to the year 2000, *Bull. World Health Organiz.*, 1990, **68**, 1–11.
14. Levine, A., A model for health projections using knowledgeable informants, *World Health Statist. Quart.*, 1984, **37**, 306–317.
15. Mann, J. M., Tarantola, D. J. M and Netter, T.W., (eds), *AIDS in the World*, Harvard University Press for the Global AIDS Policy Coalition, Cambridge, MA, 1992.
16. Dondero, T. J., St Louis, M., Anderson, J., Peterson, L. and Pappaioanou, M., Evaluation of the estimated number of HIV infections using a spreadsheet model and empirical data, Fifth International Conference on AIDS, Montreal, Canada, June, 1989 (Abstract M.A.O.4).
17. Healy, M. R. J. and Tillett, H. E., Short-term extrapolation of the AIDS epidemic, *J. R, Statist. Soc.*, 1988, **151**, 50–65.
18. Morgan, W. M. and Curran, J. W., Acquired immunodeficiency syndrome: Current and future trends, *Public Health Rep.*, 1986, **101**, 459–465.
19. Brookmeyer, R. and Gail, M. H., A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J. Am. Statis. Assoc.*, 1988, **83**, 301–308.
20. Chin, J. and Lwanga, S. K., Estimation and projection of adult AIDS cases: A simple epidemiological model, *Bull. WHO*, 1991, **69**, 399–406.
21. Anderson, R. M., May, R. M., Boily, M. C., Garnett, G. P. and Rowley, J. T., The spread of HIV-1 in Africa: Sexual contact patterns and the predicted demographic impact of AIDS, *Nature*, 1991, **352**, 581–589.
22. Auvert, B., Moore, M., Bertrand, W. E., Beauchet, A., Aegerter, P., Lusamba, D., Diong, K. T., and Klinowski. J., Dynamics of HIV infection and AIDS in central African cities, *Int. J. Epidemiol.*, 1990, **19**, 417–428.
23. Way, P.O. and Stanecki, K., The demographic impact of an AIDS epidemic on an African country: Application of the IWGAIDS model, Center for International Research Staff Paper No. 58, US Bureau of the Census, Washington, DC, 1991.
24. United Nations/World Health Organization, The AIDS epidemic and its demographic consequences, Proceedings of the United Nations/World Health Organization Workshop on Modelling the Demographic Impact of the AIDS Epidemic in Pattern II countries: Progress to Date and Policies for the Future, New York, 15–16, December 1989, United Nations Publication Sales No. E 91.XIII.5ST/ESA/SER.A/119, New York, 1991.
25. Bulatao, R. A., Projecting the demographic impact of the HIV epidemic using standard parameters, (Author/JC's note: Randy Bulatao participated in the above workshop and had his paper published in the above proceedings: pp. 90–104), 1991.
26. Stanley, E. A., Seitz, S. T., Way, P. O., Johnson, P.D. and Curry, T.F., The interagency working group (IWG) model for the heterosexual spread of HIV and the demographic impact of the AIDS epidemic, (Similar note as above: pp 119–136), 1991.