Estimation and projection of adult AIDS cases: a simple epidemiological model

J. Chin & S.K. Lwanga

Many HIV/AIDS (acquired immunodeficiency syndrome) models have been developed to help our understanding of the dynamics and interrelationships of the determinants of HIV (human immunodeficiency virus) spread and/or to develop reliable estimates of the eventual extent of such spread. These models range from very simple to very complex. WHO has developed a simple model for short-term projections of AIDS, details of which are presented here along with results obtained using the model to estimate and project AIDS cases for the USA, sub-Saharan Africa, and south/south-east Asia.

WHO has also developed, based on the model described in this paper, a computer program (Epi Model), which will enable the user to easily change the values of any of the variables required by the WHO model.

Introduction

Since the early 1980s, when AIDS (acquired immunodeficiency syndrome) was first recognized, there has been uncertainty about the future trends and ultimate dimensions of this pandemic. This uncertainty persists because of the difficulties in measuring, with any substantial degree of precision, the prevalence and more particularly the incidence of AIDS cases and HIV (human immunodeficiency virus) infections in any given population. As a result, many varieties of HIV/AIDS models have been developed in an attempt to understand the dynamics and interrelationships of the major determinants of HIV transmission and/or to develop reliable estimates of the HIV/AIDS pandemic.

This paper describes the general approaches used in HIV/AIDS models, and presents the results obtained by a WHO model developed for estimation and short-term projection of adult AIDS cases. As models of paediatric AIDS require estimates of HIV-infected women in addition to assumptions and estimates that are different from those used in modelling adult AIDS, they are not included here.

Types of HIV/AIDS models

HIV/AIDS models can be broadly classified into three types which include the simplest to the most complex. The first two types are more empirical and the last type has also been referred to as explanatory or deterministic.

- Type I are models that use reported AIDS case data for short-term (2–3 years) projections of AIDS cases; such projections have been made by several groups who have applied statistical extrapolation techniques to the observed temporal curve of reported AIDS cases (1–3). These models assume that after adjustment for inherent reporting delays (and in some models, further adjustments for incomplete reporting), past trends in reported cases will continue for the next few years in a pattern similar to that already observed. Such extrapolation methods are relatively easy to apply but they consider no biological or epidemiological data other than reported AIDS cases. Thus, they are limited and can be used for projections of AIDS cases only in areas where AIDS case-reporting is relatively reliable and complete. A variation of this type of model “fits” a mathematical curve to the available data on reported AIDS cases. Several of these latter models have been described (4).

- Type II models use data on estimated HIV infections in addition to progression rates from infection to AIDS to calculate the number of past AIDS cases and to provide short-term (3–5 years)
project a type of AIDS cases. WHO has developed a Type II model for areas where AIDS case-reporting is known to be largely incomplete and unreliable. A variation of this approach uses annual reports of AIDS cases to estimate yearly HIV infections through use of a “back calculation” method based on annual progression rates. The derived estimates of annual HIV infections are then used to project AIDS cases over the next 2–3 years (5). However, this latter approach requires reliable and substantially complete AIDS case reports; furthermore, it cannot be used to estimate recent HIV incidence, and thus is of limited utility for most developing countries. Types I and II models have been used to make short-term projections of AIDS, but neither is capable of projecting HIV infections.

- Type III models are more complex and incorporate biological and behavioural variables which describe HIV transmission and its natural history to simulate the infection and disease process (6–8). Many Type III explanatory, mathematical or deterministic models of varying sophistication have been developed for describing the future course of HIV infection and AIDS in different population groups (9–11). If Type III models can be shown to be valid and the data sets used in them are accurate, they would be capable of making reliable projections of both HIV infections and AIDS cases. However, the major problem with these more complex models is that they require detailed data or reliable estimates of the many potential determinants of HIV transmission and natural history which are currently not available. At present, their greatest utility may be in testing hypotheses and evaluating the effectiveness of different HIV prevention and/or control measures.

Comprehensive reviews of the statistical and mathematical approaches used in HIV/AIDS modelling, and descriptions of the more conventional methods used to model the course of the HIV/AIDS pandemic and the data required have been published (12–15).

The WHO model

Several applications of the model developed by WHO to estimate current and future AIDS cases have been described in previous publications (16–18). Briefly, an HIV point prevalence estimate is used in combination with (1) the estimated year when HIV transmission became widespread, and (2) the HIV infection curve during the epidemic period. These data are then used to calculate annual cohorts of HIV-infected adults. AIDS cases are then determined by multiplying each annual cohort of HIV-infected persons by the progression rates from infection to clinical AIDS. A more complete description of how the model works and the estimates required is given below.

(a) The year extensive spread of HIV infections started. Available data indicate that extensive HIV spread (defined as when at least 1% HIV seroprevalence is found in a high-risk “group”) did not begin in human populations until the late 1970s or early 1980s (19). Thus, in the absence of specific data to indicate otherwise, 1980 ± 2 years can be used as the most likely starting point for extensive spread of HIV infections in North America, sub-Saharan Africa, and western European countries. For south and south-east (S/SE) Asian countries, extensive spread of HIV infections was not documented until 1988 (20).

(b) HIV point prevalence. HIV point prevalence in any specific population may be estimated from available HIV serological data such as the “family of serosurveys” in the USA (21), and from computerized HIV databases (22). WHO has made global and regional estimates of HIV prevalence: for industrialized countries, estimates developed by national experts and/or national AIDS programmes were used; for African, Asian and Pacific countries, data bases of both published and unpublished serological data were utilized (23).

(c) The year HIV prevalence was measured (the reference year). Specifications from (a) and (c) define the “age” of the epidemic.

(d) Where on the HIV epidemic (incidence) curve is the reference year? Epidemiological data and observations may suggest that HIV incidence is generally increasing or decreasing. To illustrate the use of the WHO model, the reference year can initially be placed at that point on the HIV epidemic curve that coincides with the year of peak incidence (Fig. 1).

(e) The shape of the HIV incidence curve. A fundamental assumption in the WHO model is that, within any population group, cumulative HIV infections ultimately follow a sigmoid curve. Such a curve is characteristic of a single source epidemic with person-to-person transmission. The model further assumes that the distribution of HIV infection over time in any population will be skewed with a long right tail. Of the numerous curves that could satisfy these assumptions, we selected a simple gamma function:

\[ t^{(p-1)} e^{-t}/(p - 1)! \]

to describe the HIV incidence at time \( t \). Parameter
Fig. 1. The WHO projection model for adult AIDS cases: annual and cumulative (cum) HIV incidences and HIV prevalence.

<table>
<thead>
<tr>
<th>Year</th>
<th>Cumulative incidence</th>
<th>HIV prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.005</td>
</tr>
<tr>
<td>2</td>
<td>0.005</td>
<td>0.030</td>
</tr>
<tr>
<td>3</td>
<td>0.030</td>
<td>0.090</td>
</tr>
<tr>
<td>4</td>
<td>0.090</td>
<td>0.150</td>
</tr>
<tr>
<td>5</td>
<td>0.150</td>
<td>0.220</td>
</tr>
<tr>
<td>6</td>
<td>0.220</td>
<td>0.290</td>
</tr>
<tr>
<td>7</td>
<td>0.290</td>
<td>0.360</td>
</tr>
<tr>
<td>8</td>
<td>0.360</td>
<td>0.430</td>
</tr>
<tr>
<td>9</td>
<td>0.430</td>
<td>0.500</td>
</tr>
<tr>
<td>10</td>
<td>0.500</td>
<td>0.540</td>
</tr>
<tr>
<td>11</td>
<td>0.540</td>
<td>0.580</td>
</tr>
<tr>
<td>12</td>
<td>0.580</td>
<td>0.620</td>
</tr>
<tr>
<td>13</td>
<td>0.620</td>
<td>0.660</td>
</tr>
<tr>
<td>14</td>
<td>0.660</td>
<td>0.700</td>
</tr>
<tr>
<td>15</td>
<td>0.700</td>
<td>0.740</td>
</tr>
<tr>
<td>16</td>
<td>0.740</td>
<td>0.780</td>
</tr>
<tr>
<td>17</td>
<td>0.780</td>
<td>0.820</td>
</tr>
<tr>
<td>18</td>
<td>0.820</td>
<td>0.860</td>
</tr>
<tr>
<td>19</td>
<td>0.860</td>
<td>0.900</td>
</tr>
<tr>
<td>20</td>
<td>0.900</td>
<td></td>
</tr>
</tbody>
</table>

- \( p \) defines the steepness of the HIV epidemic curve. A value of \( p = 5 \) is used since this gamma distribution for HIV infections provided the best empirical "fit" to the reported AIDS-case curves in countries with reliable case-reporting systems.

(f) Annual HIV-infected cohorts. With regard to the distribution of HIV infections back to the starting year, any estimate of an HIV point prevalence will underestimate total infections that occurred (i.e., cumulative incidence), since those persons who developed AIDS and/or died would have been omitted from the most recent point prevalence estimate (Fig. 1, lower curve). The model, therefore, adjusts the point-prevalence estimate upwards by using annual progression rates from HIV to AIDS to arrive at an estimate of cumulative HIV incidence from the time widespread transmission began to the year the prevalence estimate was obtained. HIV cumulative incidence (upper curve in Fig. 1) is then partitioned into annual cohorts of HIV-infected persons by use of the gamma function given above.

(g) Progression rates from HIV infection to AIDS. Cohort studies that followed HIV-infected persons from the time of initial HIV infection showed that only about 3% developed AIDS within the first 3 years. Thereafter, there appears to be a steady 6–7% annual increase up to about year 11, the current limit of observations. Annual progression rates used in the WHO model up to year 10 were extrapolated from published cohort studies (24, 25). After 10 years, progression to AIDS is assumed to continue at the rate of 4% annually so that about 70% of the initial cohort develop AIDS within 15 years, and about 90% within 20 years (Table 1).

Model results according to different scenarios

To illustrate the WHO model, a hypothetical population with a point prevalence of 100,000 HIV-infected adults 10 years after the start of extensive spread of HIV infections is used to estimate the numbers of AIDS cases that would have occurred up to the 10th year and to project AIDS cases for the next 3–4 years. For initial modeling, the point-prevalence year (year 10) is assumed to be the year of peak incidence. The effects of stopping or continuing HIV transmission beyond the year of peak incidence can be modelled (scenario 1), as well as shifting the point-prevalence year to earlier or later points on the HIV epidemic or incidence curve (scenario 2).

Fig. 2 presents results of the WHO model for the first scenario described above—HIV transmission continued or stopped after the 10th year. When HIV transmission is continued there is a cumulative total of over 11,000 AIDS cases by year 10 and a total of about 40,000 by year 14. With HIV transmission stopped after the 10th year, the cumulative total of AIDS cases shows no difference until year 13 (Table 2); by year 14 the total is still relatively close to that obtained by continuing HIV infections—37,000 compared with 40,000.

The effects of moving the point-prevalence year to earlier or later points on the HIV incidence curve (scenario 2) is shown in Fig. 3. There is very little...
difference in projected cumulative AIDS cases when the base year is taken to be 3 years prior to the peak incidence as compared with placement of the reference year at the point that coincides with peak incidence. When the reference year is moved to 3 years after the peak incidence period, then larger numbers of AIDS cases occur both during the first 10 years and for at least 4 years thereafter. If the projection period is limited to 3–4 years beyond the point-prevalence year (i.e., the reference year), the different results obtained by moving the reference year to earlier or later points on the incidence curve are of relatively minor importance for public health purposes (Table 3).

**Practical examples**

Three examples—the USA, sub-Saharan Africa, and S/SE Asia—are described below to illustrate use of the model in making short-term projections of AIDS cases. The following specific assumptions or estimates were used.

1. **HIV prevalence.** HIV prevalence for the USA is derived from the Centers for Disease Control (CDC) estimate of 1 to 1.5 million infections as of June 1989 (26): the lower CDC estimate was used. WHO estimated that about 5 million adult HIV infections had occurred in African countries as of 1990, and by late 1990 there were about 350,000 infected adults in S/SE Asia (23).

2. **Calculation of annual HIV infected cohorts.** Epidemiological data suggest that the incidence of HIV infection in the USA was highest in the early 1980s: transmission rates among white male homosexuals had declined prior to the mid-1980s. The 1,000,000 prevalent HIV infections as of 1990, after adjustments to estimate cumulative incidence, were divided into annual cohorts of newly infected persons starting in the late 1970s; 1985 was used as the year of peak incidence. Annual HIV infections in sub-Saharan Africa were determined in a similar fashion from a point-prevalence estimate for 1990, and with 1992 projected to be the peak incidence year. In S/SE Asia, extensive spread of HIV infections did not begin until after 1987, and the

---

### Table 2: Cumulative AIDS cases with HIV transmission continued or stopped after year 10

<table>
<thead>
<tr>
<th>Year</th>
<th>Stopped</th>
<th>Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11000</td>
<td>11000</td>
</tr>
<tr>
<td>11</td>
<td>16000</td>
<td>16000</td>
</tr>
<tr>
<td>12</td>
<td>23000</td>
<td>23000</td>
</tr>
<tr>
<td>13</td>
<td>30000</td>
<td>31000</td>
</tr>
<tr>
<td>14</td>
<td>37000</td>
<td>40000</td>
</tr>
</tbody>
</table>

---

### Table 3: Cumulative AIDS cases when year 10 is placed at different points on the sigmoid curve

<table>
<thead>
<tr>
<th>Year</th>
<th>Sigmoid A: low (3 years before peak)</th>
<th>Sigmoid B: middle (At peak incidence)</th>
<th>Sigmoid C: high (3 years after peak)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>9000</td>
<td>11000</td>
<td>14000</td>
</tr>
<tr>
<td>11</td>
<td>14000</td>
<td>16000</td>
<td>20000</td>
</tr>
<tr>
<td>12</td>
<td>21000</td>
<td>23000</td>
<td>27000</td>
</tr>
<tr>
<td>13</td>
<td>28000</td>
<td>31000</td>
<td>35000</td>
</tr>
<tr>
<td>14</td>
<td>36000</td>
<td>40000</td>
<td>44000</td>
</tr>
</tbody>
</table>
increases in incidence through 1990 are believed to fall on the steep ascending portion of the incidence curve.

(3) Progression from HIV infection to AIDS and from AIDS to death. Progression rates for the USA were estimated from cohort studies of HIV-infected persons (homosexual men and males with haemophilia) as noted earlier. Progression data are not available for African or Asian populations and, therefore, the same progression rates were used.

From these basic data, the annual incidence and prevalence of adult AIDS cases were calculated for the USA, sub-Saharan Africa and S/S-E Asia. To calculate AIDS deaths in sub-Saharan Africa, cumulative survival rates were assumed to be as follows: 50% during the year of diagnosis, 20% after one year, 5% after 2 years, and none after 3 years.

Specific projections

Projections for the USA compared with CDC projections. Fig. 4 presents reported and projected annual AIDS cases for the USA compared with projections made by CDC's Type I extrapolation model. The WHO model projections "fit" the temporal curve of reported cases fairly well up to 1988, and although slightly lower, provide a good "fit" to the lower limit of the CDC projections for 1989 through 1992 (26).

Projections for sub-Saharan Africa. Based on an estimated prevalence in 1990 of 5 million HIV-infected adults, the WHO model estimated that by the end of 1990 there would be a cumulative total of close to 700000 adult AIDS cases in sub-Saharan Africa (Fig. 5). By 1994, a cumulative total of over 2 million adult AIDS cases was projected with over 1.75 million deaths. The WHO estimate of adult AIDS cases for sub-Saharan Africa by late 1990 is about 10 times greater than the reported total of AIDS cases for the same period.

Projections for south/south-east Asia. Based on 350 000 HIV-infected adults as of late 1990 and assuming no further increase in HIV infections, the model projects a minimum of over 60000 AIDS cases for this region by 1994 (Fig. 6). Up to late 1990, just over 100 AIDS cases had been reported in S/SE Asia; thus, projected adult AIDS cases will constitute an enormous increase in clinical disease for which detailed health care planning is needed.

Fig. 4. Reported and projected annual AIDS cases, USA, 1980 to 1992.
Fig. 5. Current and projected adult AIDS cases and deaths in sub-Saharan Africa, 1982 to 1994.
Fig. 6. Estimated annual HIV infections and projected adult AIDS cases in south and south-east Asia, 1988 to 1994.
Comments

Projections or extrapolations of AIDS cases that use case data depend on the availability of reasonably reliable data on reported cases. Essentially extrapolation involves determination of a trend line that describes the relationship between time intervals (e.g., 1982, 1983, etc.) and AIDS cases diagnosed during each identified interval. The trend line or curve used may be linear, quadratic, cubic, logistic, or some other curve. The equation for the curve is fitted to the existing data, usually by least squares methodology. Once the coefficients for the curve have been determined, the curve or trend line can be extended and a projection made for any future time. While several such models have been described, there is serious difficulty in using models that fit mathematical curves to observed case data since many mathematical curves fit the data equally well. Not only is there no way of choosing the best among the various curves, but the fitted curves may lead to widely divergent projections, particularly if they are used to make projections over longer periods of time. In addition, reliable reported AIDS case data are generally available only for the USA and a number of countries in Europe. AIDS case-reporting is largely incomplete in those sub-Saharan African countries most affected by the pandemic. Therefore, methods for estimating past, current and future epidemic trends that rely on reported AIDS case data cannot be applied in these countries.

To evaluate the utility of complex exploratory models (Type III models) to project the future course of the HIV/AIDS pandemic, WHO and the United Nations Population Division convened a modelling workshop in December 1989 (27). Participants were given a uniform set of assumptions and input parameters for their models so that their long-term (25 years) projections of HIV infections and AIDS cases could be compared. The results were quite divergent and varied in some instances by more than tenfold. The results indicated quite clearly that major differences exist among current Type III models and that, at present, there is no way to validate or select the most “accurate” among the many complex mathematical models that have been developed.

Still some estimate of AIDS cases must be made for developing countries in order to plan intervention and patient care programmes. WHO has used the available data (i.e., estimates of the prevalence and distribution of HIV infection, and our present understanding of the natural history of the disease) to make current estimates and short-term projections of AIDS cases. The WHO model presented here aims at providing AIDS control programmes with a simple tool which can provide reasonable insight into likely trends and numbers of AIDS cases over the succeeding 3–4 years. The WHO model cannot provide reliable future estimates of HIV infections but, as has been shown, either stopping or continuing HIV transmission after a specific year does not greatly affect short-term projections of AIDS cases. However, major sources of potential errors in this relatively simple model must be constantly reviewed.

The greatest error could occur in estimating an HIV point prevalence. In the USA, the use of numerous data sets for analysis and extrapolation has resulted in national HIV seroprevalence estimates that range from a low of about half a million to over 1.5 million. Recent HIV seroprevalence estimates in developed countries may be in error by as much as ±50% depending on the extent of reliable serological surveys. The WHO model’s results suggest that cumulative HIV incidence in the USA as of 1990 may be slightly less than the lower range (1 million) of the CDC estimate. In many of the larger African countries, available serological survey data may enable estimates of an HIV point prevalence to be made with similar or even greater precision.

Another major source of error in producing estimates and projections with the WHO model might be the annual progression rates from HIV infection to AIDS. The rates used were obtained from cohort studies of homosexual men and males with haemophilia. The degree of error which may result from using these progression rates is not known. If progression rates in some developing countries are significantly faster (i.e., a median of 6 years versus 10 years), then our current estimates would be too low; if progression rates are longer, our current estimates would be too high. Furthermore, because of possible significant delays in progression to clinical disease in many patients in the USA as a result of extensive use of zidovudine (26), the progression rates used in the WHO model will need to be adjusted for such situations.

A frequent criticism of this model is that a great many curves can be ‘drawn’ through a single point (the point-prevalence year). While that argument is theoretically true, the epidemiological basis upon which the model is formulated rules out the majority of curves which can be drawn through the point chosen for the reference year. In addition, relatively small differences were obtained when the point prevalence or reference year was placed at different points on the HIV epidemic curve used.

We recognize the limitations of our approach. For example, the model:
— should not be used to project AIDS cases for periods longer than 4 years;
— offers no additional insights about the epidemiological features of the pandemic;  
— cannot be used to predict future HIV infections or consequences of behavioural and social changes in populations.

In addition, it must be recognized that in any large area or population, the spread of HIV infection and the subsequent appearance of AIDS cases are usually the consequence of several epidemics, i.e., in either different risk “groups” or different geographical areas. Each epidemic—whether it be among persons with multiple sex partners, or injecting drug users, in urban or in rural areas—has its own starting point and intensity of spread (“force of infection”). Each epidemic should be modelled separately if sufficient epidemiological data are available.

WHO has developed a “user friendly” computer program (Epi Model) to estimate and project AIDS cases based on the model described in this paper. Epi Model will enable the user to change any of the variables used; in addition, it incorporates progression from AIDS to death so that both AIDS deaths and prevalent AIDS cases can be calculated. WHO is also preparing country-specific HIV databases along with guidelines to update and maintain these databases. These materials will be distributed to national AIDS control programmes so that they may derive their own HIV point-prevalence estimate to run Epi Model. A paediatric version of the model is also currently under development. It is being designed to determine the effects of (a) different age-specific HIV infection rates in women; (b) different age-specific fertility rates; and (c) different transmission rates from an infected mother to her fetus or infant. Such a model is needed to estimate more precisely the growing annual numbers of paediatric AIDS cases as well as potential AIDS-related orphans (28).

The WHO model described here is a simple method for estimation of past and current adult AIDS cases, and for short-term projection of AIDS cases. The model can be used to obtain the information needed for deciding future health care policies and plans for populations that lack reliable data on behavioural and biological variables as well as AIDS cases.

Résumé

Estimations et projections des cas de SIDA chez l’adulte: un modèle épidémio logique simple

De nombreux modèles VIH/SIDA (syndrome d’immunodéficience acquise) ont été élaborés afin d’aider à comprendre la dynamique et les interrelations des déterminants de l’extension du VIH (virus de l’immunodéficience humaine) et/ou de procéder à des estimations fiables de son extension future. Ces modèles vont du plus simple au plus complexe. En général, les modèles simples sont conçus pour les projections à court terme, et les plus complexes pour évaluer les déterminants de la pandémie VIH/SIDA et tester des hypothèses. L’OMS a mis au point un modèle simple pour les projections du SIDA à court terme; on trouvera dans le présent article les détails de ce modèle ainsi que les résultats des estimations et des projections du nombre de cas de SIDA qu’il a permis d’obtenir pour les États-Unis d’Amérique, l’Afrique subsaharienne et l’Asie du Sud et du Sud-Est.

Pour utiliser le modèle OMS, il faut avoir une estimation de la prévalence du VIH une année donnée, une estimation de la date à laquelle l’extension épidémique du VIH a commencé et une courbe des cas d’infection estimés pendant la période épidémique. A partir de ces données, on calcule les cohortes annuelles d’adultes infectés par le VIH. On peut ensuite calculer les projections à court terme (jusqu’à cinq ans) des cas de SIDA en multipliant chaque cohorte annuelle de sujets infectés par les taux annuels de passage de l’infection au SIDA avéré. Le modèle OMS ne peut fournir des estimations fiables des infections futures par le VIH, mais les projections à court terme des cas de SIDA sont pratiquement indépendantes des infections futures puisqu’environ 90% des cas de SIDA attendus au cours des trois à quatre prochaines années apparaîtront chez des personnes actuellement infectées par le virus.

Les projections des cas de SIDA pour les États-Unis obtenues grâce à ce modèle concordent avec la limite inférieure des projections des cas de SIDA calculée par les Centers for Disease Control (États-Unis d’Amérique) au moyen d’un modèle différent. Le modèle OMS prévoyait pour la fin 1990 près de 700 000 cas adultes de SIDA en Afrique subsaharienne, chiffre environ 10 fois supérieur au nombre total de cas notifiés pour cette période. Pour l’Asie du Sud et du Sud-Est, il prévoit un minimum de plus de 60 000 cas de SIDA en 1994.

L’OMS a mis au point, d’après le modèle décrit dans cet article, un programme d’ordinateur (Epi Model), qui permettra à l’utilisateur de modifier facilement la valeur de l’une ou l’autre des variables requises. Le programme Epi Model sera mis à la disposition des épidémiologistes du SIDA pour préparer des estimations et des projections des cas de SIDA dans les populations pour lesquelles on manque de données fiables sur les variables comportementales et biologiques ainsi que sur les nombres de cas de SIDA.
References