

3 Model Design

This section details the considerations that went into the design of the model. Our intention is to document what was left out, and why, as well as what went in and why.

At the most general level, the model will be defined by (i) the information required to define the *state* of the population at any given point in time and (ii) a *dynamic* that will generate the state of the population at any future point in time from information on the population at any past point in time.

We adopt the discrete time formulation of population projection rather than the continuous time approach of integral equations or partial differential equations. This approach is thoroughly familiar to demographers and not unfamiliar to epidemiologists who do modeling work. Precision more than sufficient for all practical work may be attained with sufficiently fine time intervals. In any case, since closed form solutions are not expected, and most unlikely to be obtained, a continuous formulation would require finite approximations for numerical calculation. Formulation in discrete terms removes the necessity for this.

Given the time scale of the phenomena and the nature of available data, one year time intervals are both necessary and sufficient. Longer intervals would not provide sufficient detail for describing epidemic trajectories and age-specific patterns, and shorter intervals would add little if anything of value.

The dynamic defining the model will therefore be a set of *state transformation equations* expressing the state of the population at the end of any year in terms of (i) the state of the population at the beginning of the year and (ii) a set of parameters and forces governing movement of persons between the various statuses comprising the state of the population. The dynamic over longer periods is defined by iterative application of the state transformation equations to an *initial state*. Applying the state transformation to the initial state distribution, taken without loss of generality to refer to time 0, produces the state of the population at time 1, applying the equations to the state at time 1 gives the state at time 2, and so on for as many iterations as desired. Information on times other than 0, 1, ... (e.g. mid-year estimates) may be obtained by interpolation. For most purposes simple linear interpolation will suffice, though curvilinear interpolation may be applied if necessary.

Defining population state

Defining what constitutes the population's state is the first step in developing the model. We are concerned in this section not only to describe but also to rationalize and explain the decision we have made about what is included and what is excluded. Given our objective of a minimally complex model, the state breakdown should be kept as simple as possible.

HIV/AIDS status

It will of course be necessary to distinguish persons infected with HIV from persons not infected. The former group may be further divided into those who are HIV positive but have not yet developed full blown AIDS and those who have developed full blown AIDS. Epimodel makes this distinction, as do numerous other models. As a general principle, finer disaggregation

improves model specification, since the dynamic may be more finely specified, though perhaps with diminishing returns if empirical data are not widely available for the categories specified.

We decided against distinguishing between those who have full blown AIDS and those who are HIV positive but apparently healthy. For the epidemics with which we are primarily concerned, there is evidence that indicates that HIV positive persons have higher mortality risks than HIV negative persons well before they contract full blown AIDS (Boerma et al, 1998; Nunn et al, 1997). If HIV positive persons are disaggregated in this way, taking account of excess mortality among HIV positive persons who have not yet developed AIDS would require separate mortality schedules for the two groups. This is a significant, though certainly not a prohibitive complication, and can just as well be taken care of computationally by specifying mortality rates for the HIV infected which are specific for duration since infection (Whitworth, 1999). Given the extremely limited empirical data on mortality risks for HIV infected persons, disaggregated or not, and the relatively short period of time between contracting full blown AIDS and death, we are inclined against making this disaggregation. In further support of this decision we note that numbers of persons with full blown AIDS could be estimated retrospectively from the time series of AIDS deaths. We therefore distinguish infected (HIV positive) and uninfected persons only.

Age

Our objective is to develop a minimally complex model that incorporates age, and we note that age is important because of the extreme variation of all relevant parameters, demographic and epidemiological, over the course of individual lives. Childbearing for women rarely occurs before age 15 or after age 50, the standard limits of “the fertile age span.” For men, ages at paternity are less well known, but 15 to 70 may be taken as reasonable limits. Even in very high mortality populations, there is an extreme concentration of mortality risks in infancy and early childhood, and in low mortality populations age differences are far more extreme, with very low risks in the late childhood and young adult ages and very high risks at older ages. For the heterosexual transmission epidemics which are our primary focus, childbearing and new HIV infections both derive from sexual activity and are therefore similarly concentrated in the reproductive age span.

The requirement that the model be formulated for one year time intervals implies that all age and duration data be expressed in single year groups. Single year detail may not be required for other purposes, indeed it might be preferable in some respects to work with broader age groups. Data may be available for five year (or even cruder) age groups, requiring interpolation procedures to generate the necessary single year detail, and outputs will rarely require single year detail. The problem with mixing single year time intervals and any but single year age (or duration) groups may be simply illustrated by noting the example of persons aged 0-4 at the beginning of a given year. Five years in the future, the survivors of these persons will be aged 5-9, *i.e.*, in the next five year age group. A single year in the future, however, they will be aged 1-5, which age group includes both persons in the 0-4 group and persons in the 5-9 group. While it would be possible to work out a procedure for allocating persons 1-5 to these two age groups, the allocation would require undesirable *ad hoc* assumptions. On balance, it is preferable to shoulder the weight of single year age detail.

The state of the population at any point in time will therefore be defined by, at least, age in single years. Age distributions in demography nearly always end with an open-ended age group, such as 85+ or 100+, since it is not worth while for most purposes to provide detail to the very extreme ages reported, often inaccurately. While the open-ended group might be set lower on the grounds that AIDS generally kills well before old age, there is little to be gained by this and the old age detail will sometimes be needed – e.g. for comparability with projection results made for purposes other than assessing the demographic impact of AIDS. It is worth noting that for

medium term projections the structure of the 65+ age group will be affected by depletions of its younger members due to HIV-related mortality after about 20 to 25 years (particularly for males) so that projecting that age group using survival proportions for a stable population open age group will not be appropriate.

Duration of infection

The risk of dying of HIV-related causes is strongly conditioned by duration of infection, though there is evidence that these risks vary by age at infection as well. To incorporate duration-specific risks of dying of HIV-related causes it is necessary to disaggregate the infected population by single years of time elapsed since infection. HIV positive persons are subject to the same risks of death from non-AIDS causes as HIV negative persons, however, and to incorporate these age-specific risks it is necessary to disaggregate the HIV positive population by age as well as duration. Once we make this disaggregation, incorporating an “age at infection” survival effect for the HIV positive does not imply any extra partitioning of the population.

Sex

Including a sex breakdown in the model will ensure comparability with standard demographic projection procedures, but there are many sex differentials which need to be recognized and incorporated into the model for more fundamental reasons. Male-female differences in mortality risks from causes other than AIDS are pervasive but the magnitudes are often modest. More important in this context are male-female differences in ages at sexual initiation and cessation of sexual activity, which lead to different age profiles for HIV incidence and prevalence. Incorporating sex differences will provide more “hooks” with which to relate model results to empirical data, and where data are available for only one sex (e.g. prevalence data based on pregnant women) will ensure that the right comparisons are made. By making incidence in the female population dependant on prevalence among males and vice versa, and introducing baseline conditions which can be made different for males and females, we can ensure that our model incorporates a realistic sex-specific dynamic for the epidemic.

Sexual activity status

We may need to allow for the impact of HIV/AIDS on fertility as well as its impact on mortality. There is evidence that HIV positive women have lower risks of childbearing than HIV negative women (Zaba and Gregson, 1998). In terms of overall impact on population growth, the fertility impact is unlikely to prove as important as the mortality impact in short to medium term projections. But it will be a critical determinant of the rate of recruitment into the HIV negative “at risk” cohorts within 20 to 25 years of the start of a significant epidemic.

Several considerations lead us to include sexual activity status despite the additional complexity this introduces. First, for heterosexually transmitted epidemics, HIV infection and childbearing both result from sexual activity, leading to a positive correlation between HIV infection and childbearing at younger ages, even if HIV positive women have lower risks of childbearing. Introducing a dichotomy between sexually active and sexually inactive persons controls for this positive correlation between HIV infection and childbearing. Sexually inactive women have zero risk of childbearing and, in a heterosexually transmitted epidemic, zero risk of infection.

Second, since most of the empirical data on HIV prevalence which we will be using to calibrate our inputs comes from surveillance of pregnant women, who are by definition sexually active, the sexual activity status classification makes it easy to deal realistically with prevalence measures for the whole adult population, for the sexually active, and for pregnant women.

A third consideration is that a logically correct and internally consistent determination of the number of births to HIV positive women is needed to represent vertical transmission and orphanhood. Vertical transmission will become the main determinant of child mortality trends in populations with large heterosexual epidemics. Orphanhood is a demographic outcome measure in its own right and another piece of empirical evidence fairly widely available from censuses and demographic surveys.

Partitioning the population into the sexually active and inactive immediately raises questions about defining movement from one population to the other. Whereas first sex can be seen as an unambiguous transition from “inactive” to “active” it is more difficult to conceptualize and define movement in the other direction. We have, however, decided to leave this possibility in the model, since abstinence from sexual activity at older ages may become an increasingly important response to the threat of HIV, and may also have important fertility impacts. To deal with the problem of lack of empirical evidence about the scale and pattern of movement between these categories, we will define the transformation equations in such a way that one possible extreme assumption will be that everyone at ages (say) 13+ is “sexually active”, and define fertility rates and infection risks accordingly.

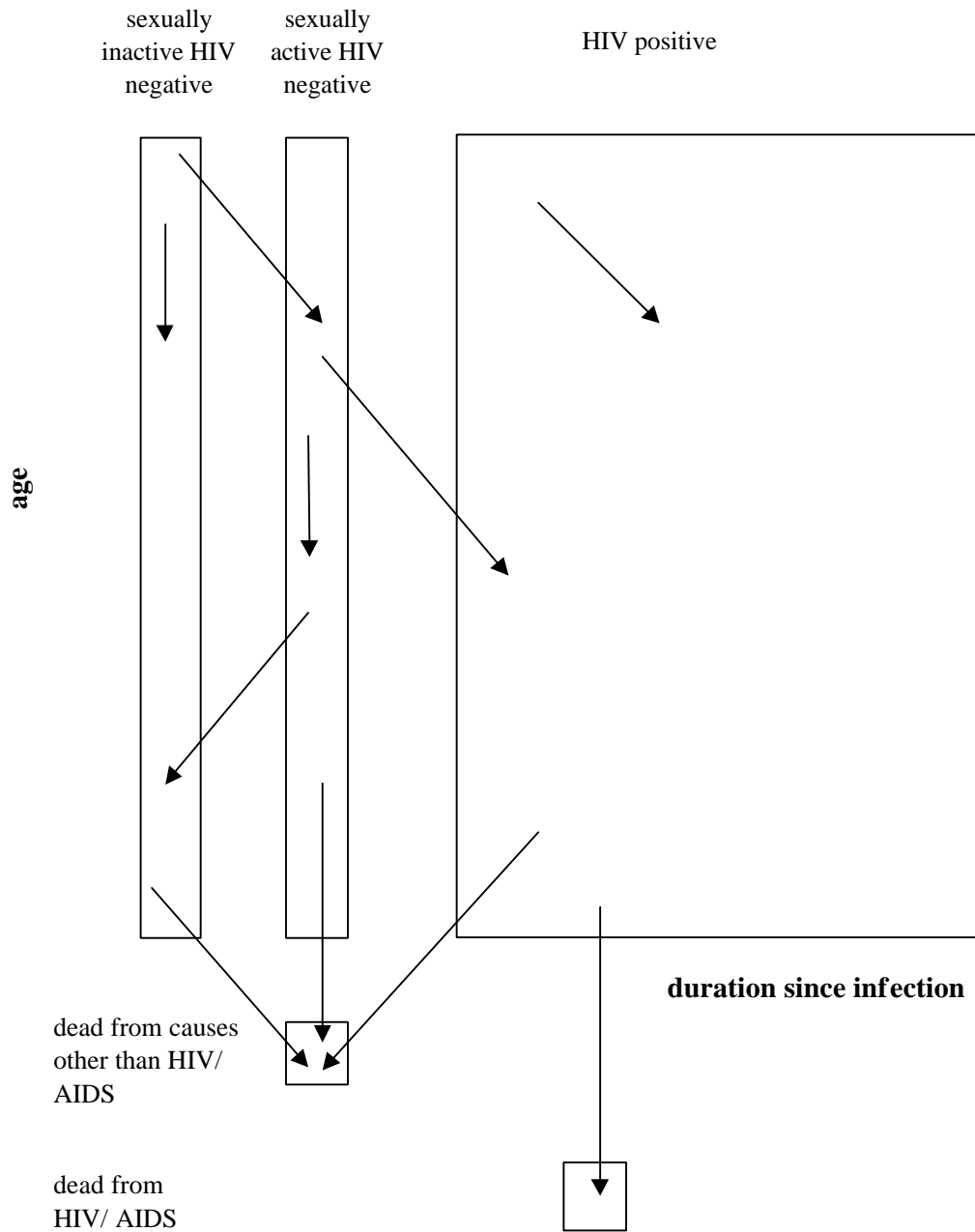
Defining a section of the population as “non-susceptible” (the sexually inactive) and specifically considering two-way transfers between “susceptibles” and “non-susceptibles” follows in the epidemiological modeling tradition described by Muench (1959) as the “reversible catalytic curve”, one of a general class of discrete time, deterministic models in wide use to this day (Daley and Gani, 1998). There are two major differences between the common use of the “non-susceptible” state in epidemiological modeling, and the sexually inactive sub-population in our model. The first is that in most models of infectious diseases people become “non-susceptible” as a result of “being cured” of the infection – they are then neither a danger to others nor can they become re-infected themselves – our model does not allow for the possibility of being cured of HIV infection. The second is that being in the non-susceptible sexually inactive state in our model carries consequences for reproduction – most epidemiological models of infection do not have specific representations for fertility.

Orphanhood

The model does not include an explicit classification of the population under age 15 by survival of parents. Although this is important information from the point of view of policy makers, the calculations involved are sufficiently different from the cyclical annual updating involved in the basic projection process as to merit separate consideration both from the point of view of formal definition and from a programming perspective. The output arrays of the model do carry enough information for orphanhood calculations to be made in a supplementary output utility program.

Figure 1

Data structure - array representation in a time slice



arrows represent transfer possibilities to next time slice

Representing the population state

The state of the population at any point in time is described by the arrays schematically represented in figure 1.

- The distribution of the uninfected, sexually inactive population by single year of age, whether for males or females, is represented by a one-dimensional array, represented by the tall thin rectangle at far left.
- The distribution of the uninfected, sexually active population is represented by a second one-dimensional array, represented by the second tall thin rectangle immediately to the right.
- The distribution of the infected population by age and duration (time elapsed since infection) is represented by a two-dimensional array, represented by the large rectangle on the right.
- The scalar values for persons who have ever died of HIV-related and other causes are represented by the two small squares at the bottom.

One set of these arrays is required for females, a second set for males. Infected persons are not disaggregated by sexual activity status and duration of infection is undefined for the uninfected.

Note that the scalar values for persons dead are stocks of deceased persons, rather than numbers of deaths, defined relative to the initial population distribution. At any point in time, the numbers in the “dead of non-HIV-related causes” and “dead of HIV-related causes” are the cumulating of all deaths in either category from the beginning of the projection. The distinction merits emphasis because demographers will tend to regard anything labeled “death” as a flow rather than a stock.

Notation for the state of the population at any given time t is as follows:

$X_S(a,t)$	number of uninfected, non-sexually active persons of sex S (female, male), age a (0, 1, ..., 99) at time t ;
$Y_S(a,t)$	number of uninfected, sexually active persons of sex S (female, male), age a (0, 1, ..., 99) at time t ;
$Z_S(a,d,t)$	number of infected, persons of sex S (female, male), age a (0, 1, ..., 99), with duration since infection d years, at time t .
$D_S(t)$	number of persons of sex S (female, male) who have died by time t , of causes unrelated to HIV / AIDS
$A_S(t)$	number of persons of sex S (female, male) who have died by time t of HIV / AIDS related causes

The population aged zero at time t is derived from births, which are calculated by applying age specific fertility rates for the sexually active to the sexually active population (see below). Because we allow for vertical transmission, individuals may be born into the infected population. We assume that none of these will survive to sexual maturity, however, so that vertical transmission has no direct impact on prevalence in the adult population. Note that for those infected perinatally, the duration since infection (d) will by definition be equal to the age (a).

Transformation equations: 1 Expression in terms of flows

It is useful to think of the transformation equations in several stages. At the first stage, the equations are written as population identities involving numbers of persons in different categories at times t and $t + 1$ and numbers of persons moving between different categories, including persons born and persons dying.

Notation for flows

Numbers of persons moving between categories are denoted by catenating the upper case letters used to represent the categories, *e.g.*, $\overline{XY}(t)$ denotes the number of persons who move from state X to state Y in the time interval t to $t+1$.

Notation for permissible movements follows. The subscript S assumes the values “F” for female or “m” for male, the age index a runs from age 0 to 99. Possible movements are listed below.

$\overline{XD}_S(a,t)$	number of sexually inactive persons of sex S , age a at time t , who die in the interval t to $t + 1$.
$\overline{XY}_S(a,t)$	number of sexually inactive persons of sex S , age a at time t , who become sexually active in the interval t to $t + 1$.
$\overline{YD}_S(a,t)$	number of sexually active persons of sex S , age a at time t , who die in the interval t to $t + 1$.
$\overline{YX}_S(a,t)$	number of sexually active persons of sex S , age a at time t , who become sexually inactive in the interval t to $t + 1$.
$\overline{YZ}_S(a,t)$	number of sexually active persons of sex S , age a at time t , who become infected in the interval t to $t + 1$.
$\overline{ZD}_S(a,d,t)$	number of infected persons of sex S , age a at time t , infected d years ago, who die from causes unrelated to HIV / AIDS in the interval t to $t + 1$.
$\overline{ZA}_S(a,d,t)$	number of infected persons of sex S , age a at time t , infected d years ago, who die of HIV / AIDS related causes in the interval t to $t + 1$.
$B_S^U(t)$	number of uninfected births of sex S in the interval t to $t+1$, whether to uninfected or infected women
$BD_S^U(t)$	number of the $B_S^U(t)$ uninfected births between time t and time $t + 1$ who die before time $t + 1$ (of non-HIV-related causes)
$B_S^I(t)$	number of infected births of sex S in the interval t to $t+1$ (born to infected women)
$BD_S^I(t)$	number of the $B_S^I(t)$ infected births between time t and time $t + 1$ who die before time $t + 1$ of non-HIV-related causes

$BA_S^I(t)$ number of the $B_S^I(t)$ infected births of sex S between time t and time $t + 1$ who die before time $t + 1$ of HIV-related causes

Transformation equations in terms of flows

In terms of these flows, the transformation equations for persons in the initial population are as follows (births are treated separately below).

$$X_S(a+1, t+1) = X_S(a, t) - \overline{XD}_S(a, t) - \overline{XY}_S(a, t) + \overline{YX}_S(a, t), \quad a > 0 \quad (1)$$

$$Y_S(a+1, t+1) = Y_S(a, t) - \overline{YD}_S(a, t) - \overline{YX}_S(a, t) - \overline{YZ}_S(a, t) + \overline{XY}_S(a, t), \quad a > 0 \quad (2)$$

$$Z_S(a+1, 0, t+1) = \overline{YZ}_S(a, t), \quad a > 0 \quad (3)$$

$$Z_S(a+1, d+1, t+1) = Z_S(a, d, t) - \overline{ZD}_S(a, d, t) - \overline{ZA}_S(a, d, t), \quad a > 0, d > 0 \quad (4)$$

$$X_S(0, t+1) = B_S^U(t) - BD_S^U(t) \quad (5)$$

$$Z_S(0, 0, t+1) = B_S^I(t) - [BD_S^U(t) + BA_S^I(t)] \quad (6)$$

In formulating these balancing equations we have specifically chosen to disregard the possibility of different transformation events befalling an individual in a single year – i.e. we have assumed that the probabilities of a person becoming sexually active and infected within the space of a single year, or becoming infected and dying of HIV related causes within the same year are effectively zero.

Transformation equations: 2 Calculating flows

The size of the flows defined in the preceding section is determined by “forces” of movement between categories, birth rates, and related parameters. This section defines these quantities and gives formulas for calculating flows from them. The essential concepts of “multiple decrement” or “competing risk” theory that are needed to partition transfers out of a risk group are described in an Appendix.

Forces of movement, birth rates, sex ratio at birth, and vertical transmission

Sexually inactive, uninfected persons present in the population at any time t may die or become sexually active before time $t + 1$.

$m_S(a, t)$ is the force of mortality from causes other than HIV / AIDS for individuals of sex S (female, male), age a (0, 1, ..., 99) at time t

$m_S^B(t)$ is the force of mortality from non-HIV-related causes for births of sex S between time t and time $t + 1$ individuals of sex S (female, male)

$u_S(a, t)$ is the force of becoming sexually active for sexually inactive, HIV negative individuals of sex S (female, male), age a (0, 1, ..., 99) at time t

$t_S(a, t)$ the force of becoming sexually inactive for sexually active, HIV negative individuals of sex S (female, male), age a (0, 1, ..., 99) at time t

$I_S(a,t)$	the force of infection experienced by sexually active individuals of sex S (female, male), age a (0, 1, ..., 99) at time t
$a(a,d,t)$	is the (duration specific) force of mortality from HIV / AIDS experienced by individuals aged a (0, 1, ..., 99), who have been infected for d years at time t
$a^B(t)$	is the force of mortality from HIV-related causes for births between time t and time $t + 1$
$b_Y(a,t)$	birth rate for sexually active, uninfected females aged a at time t , defined as the number of sexually active, uninfected women age a in completed years at time t divided into the number of births these women have over the following year
$b_Z(a,t)$	birth rate for infected females aged a at time t , defined as the number of infected women age a in completed years at time t divided into the number of births these women have over the following year
s_S	proportion of all births of sex S (female, male)
$z(t)$	vertical transmission ratio, defined as the proportion of infected births during the t -th year among all births to infected females during this year

Expressions for flows

Sexually inactive and uninfected persons aged a and time t may die or become sexually active by time $t + 1$. The numbers of persons comprising these two flows are given by the following multiple decrement formulas.

$$\overline{XD}_S(a,t) = X_S(a,t) \frac{m_S(a,t)}{m_S(a,t) + u_S(a,t)} [1 - \exp\{-[m_S(a,t) + u_S(a,t)]\}] \quad (7)$$

$$\overline{XY}_S(a,t) = X_S(a,t) \frac{u_S(a,t)}{m_S(a,t) + u_S(a,t)} [1 - \exp\{-[m_S(a,t) + u_S(a,t)]\}] \quad (8)$$

Note that equations (5) and (6) are identical on the right hand side but for the numerator of the middle term, which assumes the values of the summands in the denominator (and in the exponential term) in the two formulas. In practice we assume $u_S(a,t)$ is zero for all $a < 10$, so that there are no persons in the uninfected sexually active category aged 10 or under.

Sexually active, uninfected persons aged a at time t may die, become sexually inactive, or become infected before time t . The numbers of persons comprising these three flows are given by the following multiple decrement formulas.

$$\overline{YD}_S(a,t) = Y_S(a,t) \frac{m_S(a,t)}{m_S(a,t) + t_S(a,t) + I_S(a,t)} [1 - \exp\{-[m_S(a,t) + t_S(a,t) + I_S(a,t)]\}] \quad (9)$$

$$\overline{YX}_S(a,t) = Y_S(a,t) \frac{t_S(a,t)}{m_S(a,t) + t_S(a,t) + I_S(a,t)} [1 - \exp\{-[m_S(a,t) + t_S(a,t) + I_S(a,t)]\}] \quad (10)$$

$$\overline{Y_Z}_S(a,t) = Y_S(a,t) \frac{I_S(a,t)}{m_S(a,t) + t_S(a,t) + I_S(a,t)} [1 - \exp\{-[m_S(a,t) + t_S(a,t) + I_S(a,t)]\}] \quad (11)$$

Note again that the right hand sides of formulas (7), (8) and (9) are identical but for the numerator of the middle term, which assumes one of the summands occurring in the denominator and the exponential term.

Infected persons aged a and of duration d years since infection at time t may die of non-HIV-related causes or of HIV-related causes. The numbers of persons comprising these two flows are given by the following multiple decrement formulas.

$$\overline{ZD}_S(a,d,t) = Z_S(a,d,t) \frac{m_S(a,t)}{m_S(a,t) + a_S(d,t)} [1 - \exp\{-[m_S(a,t) + a_S(d,t)]\}] \quad (12)$$

$$\overline{ZA}_S(a,d,t) = Z_S(a,d,t) \frac{a_S(d,t)}{m_S(a,t) + a_S(d,t)} [1 - \exp\{-[m_S(a,t) + a_S(d,t)]\}] \quad (13)$$

Note that we effectively assume that $a(9,9,t) = \infty$ for all t , to ensure that all persons born HIV positive die before age 10, so that they do not enter the sexually active population.

Births are obtained by summing the effects of the fertility rates experienced by sexually active females, allowing for perinatal infection and appropriate breakdown by sex. Births not infected with HIV are given by

$$B_S^U(t) = s_S \left\{ \sum_{a=15}^{49} Y_F(a,t) \cdot b(a,t) + [1 - z(t)] \sum_{a=15}^{49} \sum_{d=0} Z_F(a,d,t) \cdot b_Z(a,t) \right\}. \quad (14)$$

The superscript “ U ” denotes uninfected. The single summation on the left gives uninfected births to sexually active, uninfected women, the double summation on the right uninfected births to infected women. Infected births to infected women are given by

$$B_S^I(t) = s_S \cdot z(t) \sum_{a=15}^{49} \sum_{d=0} Z_F(a,d,t) \cdot b_Z(a,t). \quad (15)$$

The superscript “ I ” denotes infected.

The birth rates $b_Z(a,t)$ are defined here as the number of births between time t and time $t + 1$ to women aged a in complete years at time t divided into the number of women aged a in completed years at the t . These differ slightly from the conventional “central” birth rates, which are defined as the number of births between time t and time $t + 1$ to women aged a in complete years divided by person years lived by women aged a in completed years between time t and time $t + 1$. The rates we use closely approximate central rates for the age group beginning at exact age $a + 0.5$ and ending at exact age $a + 1 + 0.5$, which may be approximate as the average of the central rates for ages a and $a + 1$ in complete years.

Uninfected births are exposed to the risk of non-HIV-related mortality only. The number of the $B_S^U(t)$ uninfected births of sex S between times t and $t + 1$ who die (of non-HIV-related causes) during this same time period is given by

$$BD_S^U(t) = B_S^U(t) \times \{1 - \exp(-m_S^B(t))\} \quad (16)$$

Infected births are exposed the risks of both non-HIV-related and HIV-related causes. The numbers of deaths of each type are given by the following multiple decrement formulas.

$$\overline{BD}_S^I(t) = B_S^I(t) \frac{m_S^B(t)}{m_S^B(t) + a^B(t)} \left[1 - \exp\{-[m_S^B(t) + a^B(t)]\} \right] \quad (17)$$

$$\overline{BA}_S^I(t) = B_S^I(t) \frac{a^B(t)}{m_S^B(t) + a^B(t)} \left[1 - \exp\{-[m_S^B(t) + a^B(t)]\} \right] \quad (18)$$

Note again the symmetry in the terms on the right hand side.

The multiple decrement expressions for deaths of births involve a bit of slight of hand that is worth calling attention to. A full multiple decrement treatment is complicated by the variable period of exposure. Births born near the beginning of the year are exposed to the risk of mortality for nearly a full year, whereas births born toward the end of the year are exposed for almost no time. In contrast, all persons alive at the beginning of the year are exposed to the risk of mortality for exactly one year before the end of the year.

We have adopted the expedient of treating all births as though they were born at mid-year, and therefore experience exactly on half year of exposure before the end of the year. The force of mortality for uninfected births, denoted $m_S^B(t)$ (here as above the subscript S denotes sex and assumes the values “f” and “m”), is thus calculated so that

$$\exp\{-0.5 \times m_S^B(t)\} = L_0 / \ell_0, \quad (19)$$

i.e.,

$$m_S^B(t) = -2 \ln\{L_0 / \ell_0\}, \quad (20)$$

where L_0 / ℓ_0 is taken from the (associated single decrement) life table for death from non-HIV-related causes in year t . The force of mortality for infected births, $a^B(t)$, is assumed equal to $a(0,0,t)$.

Transformation equations: 3 The force of new infection

In general, we do not constrain the age schedules of the various forces employed in the model to follow any particular functional form, leaving open the possibility of choosing empirical data or functional forms appropriate to each application. The important thing to note in the notational definitions given at the beginning of the preceding section is the form of the functional dependence of these forces: we have assumed that mortality from non-HIV related causes, sexual initiation and cessation, and infection are all age and sex-specific, and may change over time. Fertility is assumed to be age specific, and may vary over time, and can be different for healthy and infected women. HIV related mortality is assumed to depend on duration since infection and on age, and is also allowed to vary over time, but is assumed to be the same for both sexes.

The forces of infection are a different matter, for specification of a particular functional form is central to the definition of the model. We assume that the force of infection can be expressed as a sum of two components. The first component consists of a term that may be thought of as representing new infection introduced into the sexually active population from outside, e.g., by infected persons usually residing elsewhere who enter the population temporarily and infect usual residents, or by processes such as blood transfusion or re-use of needles. We allow this term to vary over age and time and denote it by $i_s(a,t)$. This term serves to initiate an epidemic and, if it remains positive, to sustain the epidemic.

The second component captures two essential elements of the model, a positive dependence of the level of new infections on the number of persons already infected, and the possibility of behavioral adaptations that reduce the force of new infection. This term is a product of the proportion of infected persons among potential sexual partners and a term with two additive components, $k_s(a)$ and $q_s(a,t)$. The first of these represents factors that may vary by age but which are assumed not to vary over time. It can be regarded as representing the minimum risk level for transmission of HIV associated with unprotected sexual intercourse which is necessary for reproduction. The second represents behavioral factors conducive to transmission of the disease that may vary over both age and time, with a decline over time tending to reduce the force of infection. Such a decline could represent the effectiveness of any combination of behavioral change and medical intervention such as increase in condom use, lower partner change rates, a decrease in the level of concurrency in partnerships, a decline in the extent of assortative mixing, and effective treatment of other STDs which can act as transmission co-factors.

The general formula for the force of new infections is as follows.

$$I_s(a,t) = i_s(a,t) + [k_s(a) + q_s(a,t)] \frac{\sum_a Z_{\tilde{s}}(a,t)}{\sum_a Y_{\tilde{s}}(a,t) + Z_{\tilde{s}}(a,t)} \quad (21)$$

where the symbol \tilde{s} on the right hand side of the equation denotes the opposite sex to the one for which the force of infection is being computed – that is the force of infection for females depends on male prevalence and vice versa.

This formulation was developed by a series of explorations with a similarly specified model for a non-age structured population, detailed elsewhere. In the Uganda projection which we used to test the methodology, we did not have enough empirical information to specify the age dependencies of $i_s(a,t)$, $k_s(a)$, and $q_s(a,t)$, but we did have some data on incidence over a five-year time period which allows us to obtain directly an age vector for $I_s(a,x)$ pertaining to some time x which we are unable to specify at the moment in terms of years since the notional beginning of the epidemic. We can normalise this age pattern – e.g. by setting the maximum age component to 1.0 – and denoting the normalised pattern by $f_s(a)$ we can assume a simpler form for the time dependency of $I_s(a,t)$:

$$\begin{aligned} I_s(a,t) &= i_s(t) \cdot f_s(a) + [k_s \cdot f_s(a) + q_s(t) \cdot f_s(a)] \frac{\sum_a Z_{\tilde{s}}(a,t)}{\sum_a Y_{\tilde{s}}(a,t) + Z_{\tilde{s}}(a,t)} \\ &= f_s(a) \left\{ i_s(t) + [k_s + q_s(t)] \frac{\sum_a Z_{\tilde{s}}(a,t)}{\sum_a Y_{\tilde{s}}(a,t) + Z_{\tilde{s}}(a,t)} \right\} \end{aligned} \quad (22)$$

which is equivalent to saying we believe that the age pattern of the incidence force does not change over time, only the level changes, and that the level has a component which is dependent on prevalence.

Initial values

This section discusses the setting of the initial state, that is, values of $X(a,0)$, $Y(a,0)$ and $Z(a,d,0)$. We may assume without loss of generality that $Z(a,d,0) = 0$ for all a and d . The initial values for $X(a,0)$ and $Y(a,0)$ are set in stages, by first computing the initial age distribution of the population as a whole $X(a,0) + Y(a,0)$, then computing the proportion of persons at each age who are sexually active, and finally applying the proportions active and inactive for each age to the number of persons at this age to obtain $X(a,0)$ and $Y(a,0)$.

The initial population age distribution may of course be set directly from empirical observation, though age reporting for many populations is too poor to use single year data without appropriate adjustment. For many purposes, however, empirical distributions may be well approximated by a stable age distribution (Keyfitz, 1968). A stable age distribution is defined by an age schedule of mortality and a population growth rate.

The population growth rate is conventionally denoted r . For the discrete formulation used here, the age schedule of mortality takes the form of the proportion of persons born during a given year who survive to the beginning of the i -th following year, $i = 1, 2, \dots$. In standard life table notation this is usually expressed as L_x / l_0 , where l_0 denotes the radix of the table, which we take without loss of generality to be one, and L_x denotes persons years lived between the ages x and $x+1$. Given these parameters, the stable age distribution is calculated as

$$X(a,0) + Y(a,0) = (L_x / l_0) \exp\{-r(a + 0.5)\}, \quad a = 0, 1, \dots, 99 \quad (23)$$

To compute the proportions sexually active at each age note first that all persons aged 0 are inactive, whence $X(0,0) = 1$ and $Y(0,0) = 0$. The distribution of persons at subsequent ages by activity status is defined recursively by

$$\begin{aligned} X(a+1, t+1) &= X(a, t) - \overline{XY}(a, t) + \overline{YX}(a, t), \quad a = 0, 1, \dots, 99 \\ Y(a+1, t+1) &= Y(a, t) + \overline{XY}(a, t) - \overline{YA}(a, t), \quad a = 0, 1, \dots, 99 \end{aligned} \quad (24)$$

where $\overline{XY}(a, t)$ denotes the number of persons aged a who are inactive at time t and active at time $t+1$ and $\overline{YX}(a, t)$ denotes persons aged a who are active at time t and inactive at time $t+1$.

To compute $\overline{XY}(a, t)$ and $\overline{YX}(a, t)$ it is necessary to take account of the competing risks, movement to sexually active status and mortality on the one hand and movement to sexually inactive status and mortality on the other. The general formulas apply and we have

$$\overline{XY}(a, t) = X(a, t) \frac{\mathbf{u}(a, t)}{\mathbf{u}(a, t) + \mathbf{m}(a, t)} [1 - \exp\{-[\mathbf{u}(a, t) + \mathbf{m}(a, t)]\}] \quad (25)$$

$$\overline{YX}(a, t) = Y(a, t) \frac{\mathbf{t}(a, t)}{\mathbf{t}(a, t) + \mathbf{m}(a, t)} [1 - \exp\{-[\mathbf{t}(a, t) + \mathbf{m}(a, t)]\}] \quad (26)$$

These are sufficient for applying the recursive formulas (16). Should we wish to know the number of deaths to inactive and active persons aged a , these would be calculated as

$$\overline{XD}(a,t) = X(a,t) \frac{\mathbf{m}(a,t)}{\mathbf{u}(a,t) + \mathbf{m}(a,t)} [1 - \exp\{-[\mathbf{u}(a,t) + \mathbf{m}(a,t)]\}] \quad (27)$$

$$\overline{YD}(a,t) = Y(a,t) \frac{\mathbf{m}(a,t)}{\mathbf{t}(a,t) + \mathbf{m}(a,t)} [1 - \exp\{-[\mathbf{t}(a,t) + \mathbf{m}(a,t)]\}] \quad (28)$$

Having computed $X(a,t)$ and $Y(a,t)$ from (21-27) the proportion sexually active is computed as

$$\frac{Y(a,t)}{X(a,t) + Y(a,t)}, \quad a = 0, 1, \dots, 99 \quad (29)$$

Outputs

The main outputs of the model are the detailed population age and sex distributions for the healthy and infected at the start of each single year time period, and the transfers between categories in each single year time interval. We envisage five main categories of output: sex and age distributions (numbers and percentages); rates and ratios specific for age and sex; crude rates and ratios; synthetic cohort summary statistics for time periods; and real cohort summary statistics. Single year distributions are stored for graphical output and secondary analysis; five-year age group summaries will be prepared for optional printout.

Note that by prevalence we understand: $\frac{\text{HIV infected}}{\text{uninfected} + \text{HIV infected}}$ at a specified point in time

and by incidence we mean: $\frac{\text{new HIV infections}}{\text{uninfected at start of interval}}$ for a specified time interval

The table below lists the main outputs envisaged, important time trends will also be shown graphically.

Output category	Items
Population age distributions (numbers and percentages, broken down by sex)	total population sexually active HIV infected new infections HIV-related deaths all deaths births by infection status and age of mother HIV-related deaths cumulated over time, with no age breakdown
Age specific rates and ratios	HIV prevalence in total population HIV prevalence in sexually active HIV prevalence in pregnant women

(by sex)	HIV incidence in total population HIV incidence in sexually active proportion sexually active in total population mortality rates for total population and by HIV status fertility rates for total population and by HIV status
Crude rates and ratios	HIV prevalence HIV incidence
(by sex, and for both sexes)	death rate HIV related deaths as a proportion of all deaths birth rate – both sexes only rate of natural increase – both sexes only
Synthetic cohort statistics	life expectancy
(by sex, and for both sexes)	index of adult mortality ${}_{45}q_{15}$ index of child mortality ${}_5q_0$ lifetime risk of infection lifetime risk of HIV related death TFR (females only)
Real cohort statistics	life expectancy (= mean age at death)
(by sex, and for both sexes)	index of adult mortality ${}_{45}q_{15}$ (adult cohort) index of child mortality ${}_5q_0$ proportion of cohort ever infected (adult cohort)
adult cohort indicates cohort starting at 15 th birthday	proportion of cohort deaths which are HIV related (adult cohort) completed family size (adult cohort)

Appendix Notes: Competing risks

Where persons can leave a population in more than one way, we have to account for competing risks in assigning those who leave to decrements from different causes. The example illustrated in equations (5-9) below demonstrates this in a population in which for notational clarity we do not specify age and sex structure, or time dependence in the forces.

Consider the population $X(t)$ of individuals who are not sexually active – people leave this population by dying (annual risk m) and by becoming sexually active (annual risk u), so the instantaneous rate of change in this population is given by:

$$\frac{dX(t)}{X(t)} = -[m+u]dt \quad (5)$$

The population remaining at the end of a unit interval of time in which these annual risks can be assumed constant, can be found by integration:

$$\begin{aligned} \int_t^{t+1} \frac{dX(t)}{X(t)} &= -\int_t^{t+1} [m+u]dt \\ X(t+1) &= X(t) \exp\{-[m+u]\} \end{aligned} \quad (6)$$

and the total decrement in the time interval t to $t+1$ can also be found by integration:

$$\int_t^{t+1} [m+u] \exp\{-[m+u]t\} dt = X(t) [1 - \exp\{-[m+u]\}] \quad (7)$$

similarly, the decrements due to mortality and to becoming sexually active are, respectively:

$$\int_t^{t+1} m \exp\{-[m+u]t\} dt = X(t) \frac{m}{m+u} [1 - \exp\{-[m+u]\}] \quad (8)$$

$$\int_t^{t+1} u \exp\{-[m+u]t\} dt = X(t) \frac{u}{m+u} [1 - \exp\{-[m+u]\}] \quad (9)$$

The principles illustrated above are applied in chapter 2 to the full age specific model with time dependent forces.