

# 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 (i) the state of the population at the end of any year in terms of (ii) the state of the population at the beginning of the year and (iii) a set of parameters controlling the way in which the state distribution at the beginning of any year is transformed into the state distribution at the beginning of the next year. The dynamic over longer periods is defined by iterative application of the state transformation equations to an *initial state distribution*. 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.

In this application, we have decided to forgo the distinction 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 rarely occurs before age 15 or after age 50, the standard limits of “the fertile age span.” 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, and indeed it might be preferable in many respects to work with broader age groups. Data will often be available only 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 for most purposes not worth while 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 AIDS 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 AIDS is strongly conditioned by duration of infection, though there is evidence that these risks vary by age of infection as well. To incorporate duration-specific risks of dying of AIDS 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**

It is not immediately clear that the model must recognize sex. From the earliest work of Alfred J. Lotka, many models in demography are based on the female population only. 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 difference in ages of infection, which are substantial, though possibly not so large as to dictate the conclusion. Incorporating sex differences will provide more “hooks” with which to relate model results to empirical data, and will also ensure comparability with standard demographic projection procedures. On these considerations, and the modest marginal complexity incurred, we conclude that the model should incorporate disaggregation by sex.

### **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. Another important fertility effect, is that in heterosexually transmitted epidemics, HIV infection and childbearing both result from sexual activity. This results in a positive correlation between HIV infection and childbearing at younger ages, even if HIV positive women have lower risks of childbearing.

A simple way to allow for this spurious positive correlation between HIV infection and childbearing is to introduce a dichotomy between sexually active and sexually inactive persons. Sexually inactive women have zero risk of childbearing and, in a heterosexually transmitted epidemic, zero risk of infection. 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, maintaining this classification in our model will allow us to deal easily and realistically with prevalence measures for the whole adult population, for the sexually active, and for pregnant women.

A further consideration which has prompted the decision to go for a more realistic representation of the population at risk of infection and birth, is the fact that a logically correct and internally

consistent determination of the number of births to HIV positive women is needed to represent vertical transmission (which will become the main determinant of child mortality trends in populations with large heterosexual epidemics) and orphanhood (which 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 much more difficult to conceptualise 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 modelling 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).

## **Representation of and notation for the population state**

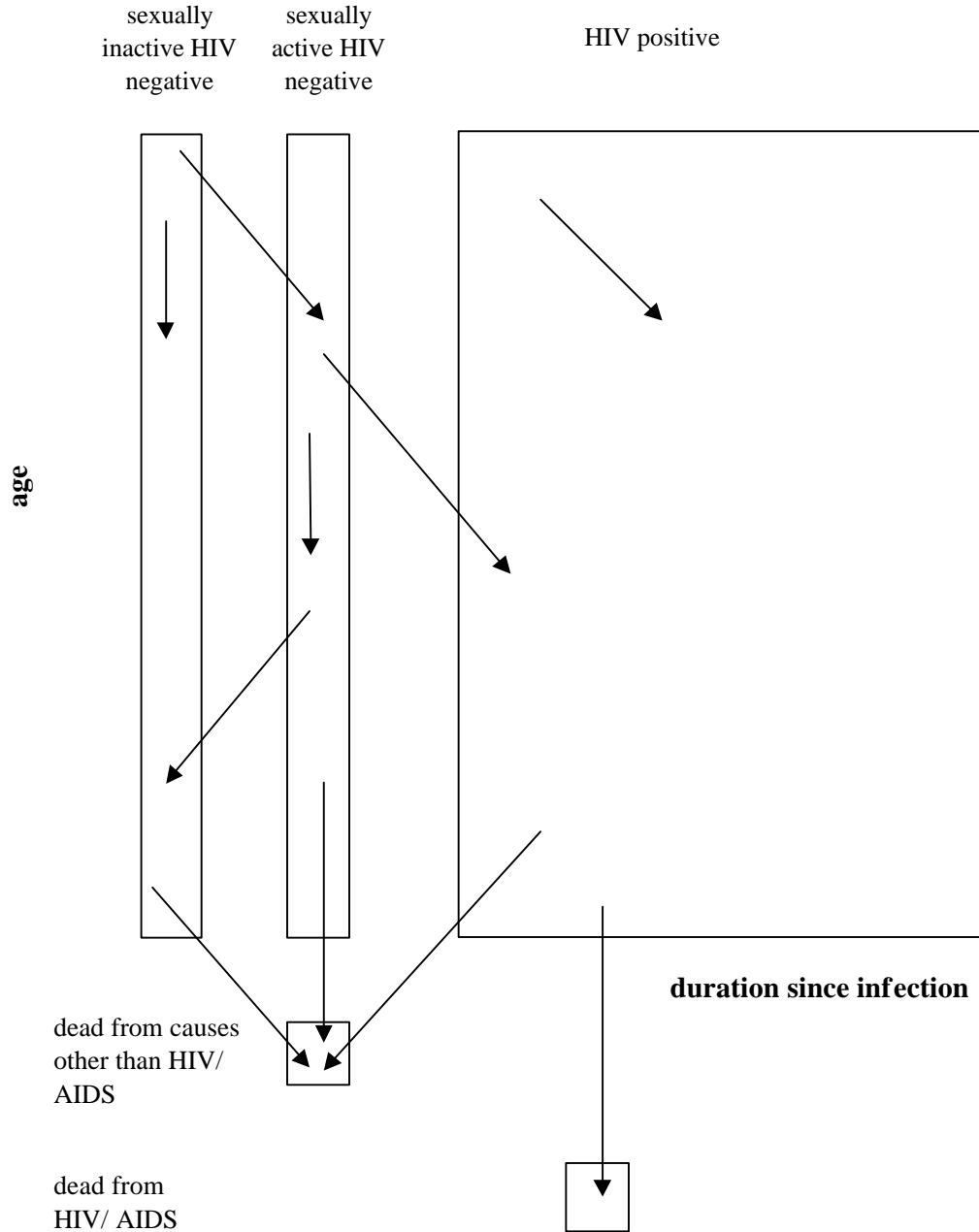
---

The state of the population at any point in time is described by a set of arrays that may be envisioned as indicated 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 rectangle. One set of these three arrays is required for females, a second set for males. Note that duration of infection is undefined for uninfected persons and that infected persons are not disaggregated by sexual activity status.

Figure 1

**Data structure - array representation in a time slice**



*arrows represent transfer possibilities to next time slice*

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

$X_S(a, t)$	denotes the number of uninfected, non-sexually active persons of sex $S$ (female, male), age $a$ (0, 1, ..., 99) at time $t$ ;
$Y_S(a, t)$	denotes the number of uninfected, sexually active persons of sex $S$ (female, male), age $a$ (0, 1, ..., 99) at time $t$ ;
$Z_S(a, d, t)$	denotes the number of infected, persons of sex $S$ (female, male), age $a$ (0, 1, ..., 99), with duration since infection $d$ years, at time $t$ .
$D_S(t)$	denotes persons of sex $S$ (female, male) who have died by time $t$ , of causes unrelated to HIV / AIDS
$A_S(t)$	denotes 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, but we assume that none of these will survive to sexual maturity, 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 ( $a$ ).

## Change of state dynamics

---

It is useful to think of the transformation equations in two stages. At the first stage, permissible movements between categories are noted. In the second stage, parameters defining the magnitude of permissible flows are defined.

Changes of state are described using the notational convention  $\overline{XY}(t)$  to describe movement from state  $X$  to state  $Y$  in the time interval  $t$  to  $t+I$

$\overline{XD}_S(a, t)$	denotes the number of sexually inactive persons of sex $S$ (female, male), age $a$ (0, 1, ..., 99) at time $t$ , who die in the interval $t$ to $t + I$ .
$\overline{XY}_S(a, t)$	denotes the number of sexually inactive persons of sex $S$ (female, male), age $a$ (0, 1, ..., 99) at time $t$ , who become sexually active in the interval $t$ to $t + I$ .
$\overline{YD}_S(a, t)$	denotes the number of sexually active persons of sex $S$ (female, male), age $a$ (0, 1, ..., 99) at time $t$ , who die in the interval $t$ to $t + I$ .
$\overline{YX}_S(a, t)$	denotes the number of sexually active persons of sex $S$ (female, male), age $a$ (0, 1, ..., 99) at time $t$ , who become sexually inactive in the interval $t$ to $t + I$ .

- $\overline{YZ}_S(a,t)$  denotes the number of sexually active persons of sex  $S$  (female, male), age  $a$  (0, 1, ..., 99) at time  $t$ , who become infected in the interval  $t$  to  $t + 1$ .
- $\overline{ZD}_S(a,d,t)$  denotes the number of infected persons of sex  $S$  (female, male), age  $a$  (0, 1, ..., 99), infected  $d$  years ago at time  $t$ , who die from causes unrelated to HIV / AIDS in the interval  $t$  to  $t + 1$ .
- $\overline{ZA}_S(a,d,t)$  denotes the number of infected persons of sex  $S$  (female, male), age  $a$  (0, 1, ..., 99), infected  $d$  years ago at time  $t$ , who die of HIV / AIDS related causes in the interval  $t$  to  $t + 1$ .
- $\overline{ZZ}_S(a,d,t)$  denotes the number of infected persons of sex  $S$  (female, male), age  $a$  (0, 1, ..., 99), infected  $d$  years ago at time  $t$ , who survive the interval  $t$  to  $t + 1$ .

The accounting equations describing movements between states follow from these definitions:

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

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

$$Z_S(a+1,0,t+1) = \overline{YZ}_S(a,t) - \overline{ZD}_S(a,d,t) - \overline{ZZ}_S(a,0,t) \quad (3)$$

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

## Defining the transformation equations

---

In this section we define, explain and rationalize the transformation equations by which the state of the population at the end of any year may be calculated from the state at the beginning of the year.

The forces acting on the population are as follows:

- $m_S(a,t)$  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  $t$
- $a(a,d,t)$  is the force of mortality from HIV / AIDS experienced by individuals aged  $a$  (0, 1, ..., 99), who have been infected for  $d$  years at time  $t$
- $I_S(a,t)$  is the force of infection experienced by sexually active individuals of sex  $S$  (female, male), age  $a$  (0, 1, ..., 99) at time  $t$
- $u_S(a,t)$  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  $t$

- $t_S(a,t)$  is the force of becoming sexually inactive for sexually active, HIV negative individuals of sex  $S$  (female, male), age  $a$  (0, 1, ..., 99) at time  $t$
- $b_Y(a,t)$  is the birth rate experienced by sexually active HIV negative females aged  $a$  at time  $t$
- $b_Z(a,t)$  is the birth rate experienced by HIV positive females aged  $a$  at time  $t$

We also need to define two ratios:

$s_S$  is the proportion of all births of sex  $S$  (female, male)

$z(t)$  is the vertical transmission ratio – the proportion of births at time  $t$  to HIV positive females that are also HIV positive

With these definitions, we can define the sizes of the “change of state” components, taking full account of multiple decrement issues:

$$\overline{XD}_S(a,t) = X_S(a,t) \frac{m_S(a,t)}{m_S(a,t) + u_S(a,t)} [1 - \exp\{-[m_S(a,t) + u_S(a,t)]\}] \quad (5)$$

$$\overline{XY}_S(a,t) = X_S(a,t) \frac{u_S(a,t)}{m_S(a,t) + u_S(a,t)} [1 - \exp\{-[m_S(a,t) + u_S(a,t)]\}] \quad (6)$$

$$\overline{YD}_S(a,t) = Y_S(a,t) \frac{m_S(a,t)}{m_S(a,t) + t_S(a,t) + I_S(a,t)} [1 - \exp\{-[m_S(a,t) + t_S(a,t) + I_S(a,t)]\}] \quad (7)$$

$$\overline{YX}_S(a,t) = Y_S(a,t) \frac{t_S(a,t)}{m_S(a,t) + t_S(a,t) + I_S(a,t)} [1 - \exp\{-[m_S(a,t) + t_S(a,t) + I_S(a,t)]\}] \quad (8)$$

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

$$\overline{ZD}_S(a,t) = Z_S(a,d,t) \frac{m_S(a,t)}{m_S(a,t) + a_S(d,t)} [1 - \exp\{-[m_S(a,t) + a_S(d,t)]\}] \quad (10)$$

$$\overline{ZD}_S(a,t) = Z_S(a,d,t) \frac{a_S(d,t)}{m_S(a,t) + a_S(d,t)} [1 - \exp\{-[m_S(a,t) + a_S(d,t)]\}] \quad (11)$$

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

Finally, the births into the population are obtained by summing the effects of the fertility rates experienced by sexually active females, allowing for perinatal infection and appropriate breakdown by sex:

$$X_S(0, t+1) = s_S \left\{ \sum_{a=15}^{49} Y_F(a, t) \cdot \mathbf{b}(a, t) + [1 - z(t)] \sum_{a=15}^{49} \sum_{d=0} Z_F(a, d, t) \cdot \mathbf{b}_Z(a, t) \right\} \quad (13)$$

$$Z_S(0, 0, t+1) = s_S \cdot z(t) \sum_{a=15}^{49} \sum_{d=0} Z_F(a, d, t) \cdot \mathbf{b}_Z(a, t) \quad (14)$$

## 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 all model trajectories begin prior to the beginning of an epidemic, so 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 [*question of reference(s)*]. 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

$$X(a, 0) + Y(a, 0) = (L_x / l_0) \exp\{-r(a + 0.5)\}, \quad a = 0, 1, \dots, 99 \quad (15)$$

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

$$\begin{aligned} X(a+1, t+1) &= X(a, t) - \overline{XY}(a, t) + \overline{YX}(a, t), \quad a = 0, 1, \dots, 99 \\ Y(a+1, t+1) &= Y(a, t) + \overline{XY}(a, t) - \overline{YA}(a, t), \quad a = 0, 1, \dots, 99 \end{aligned} \quad (16)$$

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 the number of 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

$$\overline{XY}(a, t) = X(a, t) \frac{\mathbf{u}(a, t)}{\mathbf{u}(a, t) + \mathbf{m}(a, t)} [1 - \exp\{-[\mathbf{u}(a, t) + \mathbf{m}(a, t)]\}] \quad (17)$$

$$\overline{YX}(a, t) = Y(a, t) \frac{\mathbf{t}(a, t)}{\mathbf{t}(a, t) + \mathbf{m}(a, t)} [1 - \exp\{-[\mathbf{t}(a, t) + \mathbf{m}(a, t)]\}] \quad (18)$$

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

$$\overline{XD}(a, t) = X(a, t) \frac{\mathbf{m}(a, t)}{\mathbf{u}(a, t) + \mathbf{m}(a, t)} [1 - \exp\{-[\mathbf{u}(a, t) + \mathbf{m}(a, t)]\}] \quad (19)$$

$$\overline{YD}(a, t) = Y(a, t) \frac{\mathbf{m}(a, t)}{\mathbf{t}(a, t) + \mathbf{m}(a, t)} [1 - \exp\{-[\mathbf{t}(a, t) + \mathbf{m}(a, t)]\}] \quad (20)$$

Having computed  $X(a, t)$  and  $Y(a, t)$  from (15-20) the proportion sexually active is computed simply as

$$\frac{Y(a, t)}{X(a, t) + Y(a, t)}, \quad a = 0, 1, \dots, 99 \quad (21)$$