SAJDem 11(1) 39–73

The Effect of HIV on the Estimation of Child Mortality Using the Children Surviving/Children Ever Born Technique

Patrick Ward[1] and Basia Zaba[2]

Abstract

The children surviving/children ever born technique is an important method of estimating child mortality levels in many developing countries. The HIV epidemic threatens the validity of the technique in a number of ways, principally through the correlation of mothers' mortality with that of their children. This paper describes the use of stable population modelling to simulate the application of the technique in populations with HIV and to assess the extent of the resulting biases. It also shows that corrected estimates can be derived given information on seroprevalence in the population on the assumption of population and epidemic stability.

INTRODUCTION

The method of estimating child mortality from reports of children ever born and children surviving (CS/CEB) is one of a number of techniques developed to estimate demographic measures indirectly in populations where vital registration systems are poor or non-existent. It uses the reports given by women, in a census or survey, of the total number of children they have borne and the number of these that have died. The proportions of children that have died, for women classified into five-year age groups, are then used to estimate standard life-table measures of mortality for the population. By

1 Oxford Policy Management, 38 St Aldates, Oxford OX1 3BN

2 Centre for Population Studies, London School of Hygiene and Tropical Medicine, 49/51 Bedford Square, London WC1B 3DP

deriving estimates based only on these aggregate proportions, the technique avoids the errors in mortality estimates which commonly result when respondents are required to recall deaths in a specified time period, or to recall exact ages or dates. For this reason, it has been extensively used to obtain estimates of child mortality in developing countries.

The epidemic of human immunodeficiency virus (HIV)[3] may undermine the validity of a number of conventional methods of estimating mortality in developing countries, including the CS/CEB technique. HIV threatens the validity of the technique in a number of ways. The technique is based on the assumption that the correlation of mortality between mothers and their children is sufficiently small to be ignored. Thus, the reports of living women on the survival of their children is an adequate representation of the survival of all children in the population. HIV may significantly violate this assumption because of the highly correlated mortality it introduces between mothers and their children, due to vertical transmission from mother to infant and the high case fatality of HIV infection in both adults and children. Reports in a cross-sectional survey on the survival of births from an earlier period may therefore fail to represent the experience of all births from that period, because of the higher mortality of HIV-positive women in the period between the births and the survey. The births of women who were HIV-positive at the time of the births will therefore form a lower proportion of the reported births than they formed of the initial set of all births. Due to vertical transmission, the births to HIV-positive mothers will also have higher mortality and so the reported proportion dead and the estimates of mortality for the population will be biased downwards.

Additionally, since seroprevalence varies with age and probability of mother to child transmission varies with viral load, HIV will introduce a source of heterogeneity into the mortality of children which is dependent on the age of their mothers at reporting. Even neglecting the bias due to the interdependence of survival, the estimates of mortality derived from the reports of a given age group of women may thus not be a valid estimate of mortality in the population as a whole. HIV will also alter age-specific mortality patterns in children and young adults. This may introduce

3 Throughout this paper, only data on HIV-1 are used for the modelling. The term HIV is used to refer to HIV-1. HIV-2 differs from HIV-1 in a number of epidemiological characteristics, including the extent of vertical transmission (De Cock et al. 1994). Although HIV-2 may also be expected to introduce some bias into the estimates, this is not examined.

additional error to the mortality estimates, since the technique is based on particular model life-tables which may be inappropriate for the age-patterns of mortality found in populations with HIV.

The effect of HIV on the validity of the technique is examined using modelling. A model HIV sexual incidence schedule is derived, by fitting an adapted fertility model to empirical seroprevalence data. Single-sex, stable populations with known mortality, fertility and HIV incidence schedules are then defined and used to calculate the proportions of dead children which women in each age group would be expected to report. These are used to estimate mortality in the populations, in a simulation of the CS/CEB technique. The errors in these estimates indicate the extent of the biases introduced by HIV. Correction factors to adjust for the biases are derived. The model is intended to represent populations in which transmission is predominantly heterosexual, many women of childbearing age are infected and vertical transmission is substantial – particularly sub-Saharan Africa, where indirect estimation is often an important source of mortality statistics.

THE EFFECT OF HIV ON ESTIMATION OF CHILD MORTALITY USING REPORTS OF CHILDREN EVER BORN AND CHILDREN SURVIVING

The CS/CEB technique was originally developed by Brass (Brass et al., 1968; Brass 1975)[4] It uses the reports given by women of the total number of children they have borne and the number of these that have died. The reports are used to calculate the proportions of children that have died, for women classified into five-year age groups, which are used to estimate standard life-table measures of mortality in the population. Provided that the family of life tables describing mortality in the population is known, then mortality estimates are calculated directly from the relationship:

$$q(z) = PD(i) * k(i) \qquad \ldots(1)$$

where $q(z)$ is the proportion dying by some integer age, z, in the life table, $PD(i)$ is the proportion of dead children for women in age group i and $k(i)$ is a multiplier derived from a given level of the model life tables. Each $q(z)$ value

4 Since the original work, the technique has been refined and elaborated by a number of authors; they also give more detailed descriptions of the method than given here. See, for example Trussel (1975), Feeney (1980), Zlotnik and Hill (1981).

can be converted to an estimate for a common age, usually *q(5)*, using an appropriate model life table. This gives a time series of estimates describing the trend in child mortality, generally over the fifteen years or so preceding the survey (Feeney 1980).

HIV threatens the validity of the technique in a number of ways. Firstly, the technique is based on the assumption that the correlation of mortality between mothers and their children is sufficiently small to be ignored. Thus the reports of living women on their children's survival are assumed to represent the survival of all children in the population. HIV may significantly violate this assumption because of the highly correlated mortality it introduces between mothers and their children, due to vertical transmission from mother to infant and the high case fatality of HIV infection in both adults and children. Such correlated mortality means that the reports given in a cross-sectional survey by women aged x on the mortality experience of their births t years ago may fail to be representative of all births t years ago. This is because women who were HIV-positive t years ago will have higher mortality in the intervening period than will women who were HIV-negative at the time. They will therefore constitute a lower proportion of the women reporting at the survey than they formed of the group of women giving birth t years ago. The children which they have borne will be correspondingly under-represented, forming a lower proportion of reported births than of the set of all births t years ago. Because of vertical transmission, these children have substantially higher mortality than the children of women who were HIV-negative t years ago. The reported proportion dead will therefore be lower than the true proportion dead and mortality in the population will be under-estimated.[5]

Secondly, the technique also assumes that the mortality experience of the children of women in a particular age group is representative of mortality in the population as a whole, either at all points in time (in a stable population) or at a particular point in the past (when mortality is changing). HIV violates this assumption by introducing heterogeneity in child mortality with respect to the age of the mothers. Since the prevalence of HIV varies with age, the proportion of children born HIV-positive may differ between the population

5 A consequence of the under-representation of HIV-positive children in the reported proportion dead is that the multipliers cease to be constant across different levels of mortality in the same life table family, even for a fixed prevalence of HIV.

as a whole and the sub-groups of children that have been born to mothers in a given age group at the time of the survey. The mortality of these sub-groups may therefore differ from that of the population as whole and the estimates of mortality derived from the reports of a given age group of women may thus not be a valid estimate of mortality in the population as a whole. Thirdly, HIV will also alter age-specific mortality patterns in children and young adults. This may introduce additional error to the mortality estimates since the multipliers, which are specific to particular families of model life-tables, may be inappropriate for the age-patterns of mortality found in populations with HIV.

ESTIMATION OF THE ERROR INTRODUCED BY HIV

In order to estimate the magnitude of the errors in the estimates introduced by HIV, the application of the technique is modelled using stable populations with a range of levels of HIV prevalence. Single-sex populations with known mortality, fertility and HIV incidence schedules are defined. A cohort of HIV-negative adult women is projected through time, decrements occurring through HIV incidence and HIV-negative mortality. Sexually infected individuals are projected using the mortality risks of HIV-positive adults and HIV prevalence is calculated for each age. For each simulation, the level of child mortality is calculated. Proportions of children who would be reported as having died by women in a cross-sectional survey is also calculated. These proportions are used to calculate the child mortality estimates that would result from application of the standard CS/CEB technique. The comparison of these estimates with the true $q(z)$ values indicates the error introduced into the technique by HIV. The relationship between the errors and prevalence is examined and the errors are entered as dependent variables into regression equations, in order to derive corrections for estimates in populations with HIV.

The calculation of the overall level of child mortality in the population and of the reported proportions dead requires the definition of age-specific incidence rates for HIV. Since empirical measures of incidence rates are not available, these are modelled using a variant of a relational fertility model (Zaba 1994). This three parameter model is fitted to a number of empirical, age-specific seroprevalence distributions, defining ranges for the parameters. These ranges are used to generate sets of age-specific incidence risks which are applied in the modelling.

Demographic components of the model

The mortality of HIV-negative individuals is defined by the life table resulting from a logit transformation of a single standard life table (Brass 1971; Brass 1975). Following Palloni and Heligman (1986), a two-sex, single-year 'General' UN life table with a life expectancy of 54 years is used as the standard. The level parameter of this model is varied between –0.5 and +0.4, corresponding to life expectancies of between 67 and 41 years respectively. A standard age structure is used for all model populations, given by the stable population corresponding to the standard life table and a growth rate of 0.03.

The fertility schedule of the population is defined using the relational system based on the Gompertz transformation of the Brass-Booth standard (Brass 1981; Zaba 1981). The location parameter of this system is allowed to vary between –0.5 and +0.5; the spread parameter between 0.8 and 1.2. Restrictions on their relative values are described below. A single, constant value is assigned to the fertility schedule level, since the proportions of dead children are unaffected by this parameter. HIV-positive and HIV-negative women are assumed to have the same fertility schedule.

The HIV model

The excess mortality of HIV-positive individuals is modelled using exponential functions. The number of HIV-positive individuals alive at time t, $P(t)$, in the absence of other causes of death would be given by:

$$P(t) = P(0)\exp(-\frac{t}{\mu}) \qquad \ldots(2)$$

where μ is the mean survival time from infection with HIV (in years) and $P(0)$ is the number infected at $t = 0$. HIV-positive individuals are also subject to decrement from non-HIV mortality, which is assumed to be the same as the mortality risks faced by HIV-negative individuals in the population.

Separate exponential survivorship models are defined for HIV-positive children and adults. The HIV mortality of HIV-positive children[6] is described by an exponential function with a mean survival time of 2.5 years and zero survivorship after age five, representing the high mortality of children infected

6 Throughout this paper, the term HIV-positive children refers to those children who are infected with the virus at birth. It excludes children not infected with the virus but positive for HIV antibodies due to the passive transfer of maternal antibodies, and children infected through breastfeeding at a later age.

peri-natally, rather than those infected later through breastfeeding (Marston et al. 2005; Newell et al. 2004a; Ionnidis et al. 2004). HIV mortality of adults is described by an exponential function with a mean of 8 years, consistent with an excess mortality about 11 times that of uninfected persons (Porter and Zaba 2004; Zaba et al., 2007). Aside from the distinction between children and adults, excess mortality from HIV is not varied according to the age of the infected individuals nor the duration of infection.

The model allows for HIV infection through heterosexual sex and by vertical transmission from mother to child, since the majority of cases of infection in sub-Saharan Africa are accounted for by these two routes (Newell et al. 2004b). Empirical, female age-specific incidence rates for sexual infection are rarely available; when available, they are based on small numbers of events. However, a number of studies have published age-sex-specific seroprevalence data. These estimates are more stable and, given certain assumptions, can be used to infer seroincidence rates. Since sexual activity is a risk factor for both fertility and heterosexual transmission of HIV, a variant of the relational fertility model is used as a basis for modelling the incidence of sexual transmission (Zaba 1994). This provides a series of age-specific risks[7] which can be applied to a cohort to generate model seroprevalence schedules which are compared to empirical age-specific prevalence schedules.

The model has a level parameter, which sets the level of the incidence risks, and two parameters which determine the shape of the distribution. The first of these principally affects the location of the distribution along the age scale; the second largely alters the spread of the distribution. The model age-specific incidence risks are given by:

$$i(x)=T^*[I(x+1)-I(x)] \qquad \ldots(3)$$

where $i(x)$ is the risk of incidence over the age interval x to $x+1$; T is the level parameter of the incidence schedule and $I(x)$ is a cumulant calculated as $I(x) = exp(-exp(-G(x))$ where:

$$G(x) = \lambda + \sigma * G_s(x) \qquad \ldots(4)$$

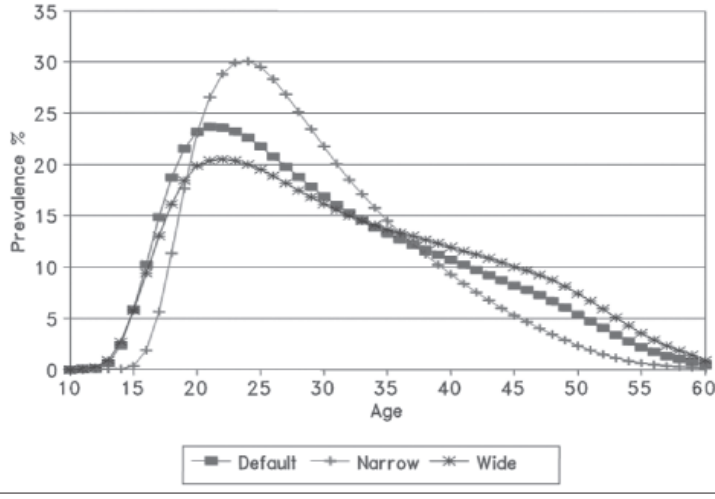
7 The fertility model gives a set of rates; in its application as an incidence model they are used as risks. Since they are fitted to empirical data, this presents no problem provided that it is done consistently.

where λ and σ are the location and spread parameters of the model. $G_s(x)$ is, in the fertility model, $-ln\ (-ln\ (F_s(x)))$, where $F_s(x)$ is cumulated fertility to age x in a normalised standard fertility schedule. However, in this case, the $G_s(x)$ values are modified in order to moderate the rapid decline in risk which occurs to women in their forties when the standard values are used (see appendix A). This schedule is then fitted to a number of empirical age-specific seroprevalence distributions, to define appropriate ranges for the model parameters. Details of these procedures, and of the fitted parameters, are also given in Appendix A.

Figure 1 illustrates the seroprevalence distribution resulting from three diverse model incidence schedules which have an exponential increase in the level parameter over twelve years.

Vertical transmission of HIV from mother to child is included in the model as a single parameter – the proportion of births to HIV-positive women that result in infants that are HIV infected at birth. This proportion is assumed

Figure 1 Age-specific female seroprevalence distributions resulting from three model incidence schedules undergoing an exponential increase in the level of incidence



The default incidence schedule has a location parameter of 0.6 and a spread parameter of 0.9; for the narrow schedule they take the values of 0.8 and 1.3 respectively, and for the wide schedule 0.4 and 0.8. For all three incidence schedules, the figure shows age-specific seroprevalence at year 12, following an exponential increase in the level parameter, at a rate of 41.6 per cent per year from a level in year zero of 0.024; this gives an adult prevalence of 15 per cent at year 12 in the default schedule.

constant across age of mother, duration of infection and progression of the disease. Since the model does not explicitly incorporate the transmission of HIV through breastfeeding, the proportion of infants HIV-positive at birth is set at 0.30, a value consistent with estimates of vertical transmission which exclude transmission through breastfeeding (Nicoll et al. 1994; Newell et al. 2004a). This avoids the need to explicitly model this pathway, and omits the transmission of HIV through breastfeeding by mothers infected post-partum. The mortality effects of HIV transmission through breastfeeding are appreciably lower than transmission intra-partum, and fewer children are infected this way (Fox et al. 2008).

Calculation of the reported proportions dead and mortality levels in the stable population models

The effect of HIV on child mortality estimates is examined using stable population models in which the level of incidence is constant over time. A cohort of HIV-negative births is projected according to:

$$u_x = u_{x-1} * (1 - i_{x-1}) * \frac{l^-_x}{l^-_{x-1}} \quad \ldots(5)$$

and the resulting HIV-positive cohort is projected according to:

$$h_x = h_{x-1} * e^{-1/\mu} + u_{x-1} * i_{x-1}) * \frac{l^-_x}{l^-_{x-1}} \quad \ldots(6)$$

where h_x, u_x denote the proportions of a birth cohort alive and respectively HIV-positive and HIV-negative at age x, i_x denotes the risk of sexual infection with HIV at age x (for $x>10$) and l^-_x represents the survivorship to age x in the HIV-negative life table. Prevalence in childbearing women (15–49) has been assumed equal to prevalence in the general adult population as is usual in the interpretation of HIV surveillance data (UNAIDS 2002). Prevalence in pregnant women aged 15 to 49, is calculated using the standard age structure.

The cohort of children who are HIV-positive at birth is projected according to the survivorship detailed above, where the proportion of HIV-positive births in the population, B^+, is given by:

$$B^+ = \sum_{x=10}^{x=50} C(x)p(x)f(x)v \Big/ \sum_{x=10}^{x=50} C(x)f(x) \quad \ldots(7)$$

where $C(x)$ is the standard age distribution, $f(x)$ and $p(x)$ are the fertility rate

and the prevalence at age x and v is the proportion of births to HIV-positive women which result in HIV-infected infants.

The true probability of death by age z in the population as a whole is denoted $q(z)^t$, to distinguish it from the estimate of this measure, $q(z)^e$. It is given by mortality in the two cohorts, weighted by the proportion born HIV-positive:

$$q(z)^t = B^+ q(z)^{HIV+} + (1 - B^+)\, q(z)^- \quad \ldots(8)$$

where $q(z)^{HIV+}$ and $q(z)^-$ are the probability of death by age z for children who are HIV-positive at birth and children who are HIV-negative at birth respectively. The mortality of all children born to mothers who are HIV-positive at the time of the birth, $q(z)^+$, may also be defined:

$$q(z)^+ = v * q(z)^{HIV+} + (1 - v)z * q(z)^- \quad \ldots(9)$$

The reported proportion of dead children is also calculated. A group of women aged x at the time of reporting may be divided into those women who are HIV-negative and those who are HIV-positive at that time. The overall reported proportion dead is a prevalence-weighted average of the proportions reported by these two groups. The number of deceased children reported by HIV-negative women aged x, denoted by $d(x)^-$ is given by:

$$d(x)^- = \sum_{t=0}^{t=x-\alpha-1} f(x-t) * Q(t)^- \quad \ldots(10)$$

where $f(x–t)$ denotes the fertility rate centred in the interval $x–t–1$; $Q(t)^- = 1– l(t+0.5)^-$, where $l(t+0.5)^-$ is the probability of surviving to the middle of the age interval t to $t+1$ for children born to HIV-negative mothers and α is the lowest age at which childbearing occurs

The mortality of the children of women who are HIV-positive at the time of the interview is dependent upon the length of time for which the women have been infected. Children born prior to seroconversion will experience the mortality risks of births to HIV-negative mothers; those born after seroconversion will experience the risks of births to HIV-positive mothers. Consider HIV-positive women aged x at the time of the survey, reporting on the survival of births born t years previously. The proportion of such births reported as dead would be a weighted average of the probability of death by age t for births to HIV-positive women and the probability of death by age t

for births to HIV-negative women. They are weighted by the proportion of living HIV-positive women aged x that were HIV-positive at age $x–t$ and its complement, respectively. Representing this proportion by $w_{x,t}$, the reported number dead for HIV-positive women is given by:

$$d(x)^{+} = \sum_{t=0}^{t=x-\alpha-1} f(x-t) * [Q(t)^{+} * w_{x,t} + Q(t)^{-} * (1 - w_{x,t})] \quad \ldots(11)$$

where terms are defined as in Eq. (10) except that $Q(t)^{+} = 1–l(t+0.5)^{+}$, where $l(t+0.5)^{+}$ is the probability of surviving to the middle of the age interval t to $t+1$ for children born to HIV-positive mothers. The weights, $w_{x,t}$, are given by:

$$w_{x,t} = \frac{h_{(x-t-0.5)} * e(-\frac{t+0.5}{\mu}) * \frac{l_x^-}{l_{(x-t-0.5)}^-}}{h_x} \quad \ldots(12)$$

where h_x and $h_{(x–t–0.5)}$ are the proportions alive and HIV-positive at age x and $x–t–0.5$ respectively, and l^-_x and $l^-_{(x–t–0.5)}$ are the proportions surviving to the same ages in the HIV-negative lifetable. The number reported dead by all women aged x, $d(x)^r$ is then given by:

$$d(x)^r = p(x)\,d(x)^{+} + (1 - p(x))\,d(x)^{-} \quad \ldots(13)$$

where $p(x)$ is the prevalence at exact age x. The quantity $d(x)^r$ reflects the reduced representation of the births of HIV-positive women in the reports given in a cross-sectional survey, compared to the true proportion that they constitute of all births in the population. The reported proportion of dead children for women classified into the standard five-year age groups, $PD(i)^r$, is therefore:

$$PD(i)^r = \sum_{x=10+5i}^{14+5i} d(x)^r\, A(x) \Big/ \sum_{x=10+5i}^{14+5i} F(x)\, A(x) \quad \ldots(14)$$

where $A(x)$ is the series of weights based on the standard age distribution and $F(x)$ is the cumulation of fertility rates to exact age x. The multiplication of these proportions by their appropriate multipliers[8] gives the estimates of

8 The multipliers used are calculated from regression coefficients derived from simulations of HIV-free populations, using the same model fertility and mortality schedules as are used in the modelling of populations with HIV. Any errors in the estimates can thereby be ascribed to the effect of HIV.

child mortality, $q(z)^e$, derived by applying the standard CS/CEB technique. The difference between these estimates and the true level of mortality is given by:

$$n(z) = q(z)^t - q(z)^e \qquad \ldots(15)$$

RESULTS: THE MAGNITUDE OF ERRORS IN THE ESTIMATES

Table 1 and Figure 2 show the percentage error in the estimates of $q(z)^t$ for a range of adult prevalence levels, generated by varying the level parameter of the HIV infection model while other parameters are held constant at their default values. The errors become substantial at levels of prevalence which are known to occur in a number of populations. Levels of adult prevalence above 35 per cent, for example, which have been observed in some urban African populations (Fylkenses et al. 1997), would result in every estimate being substantially in error. Table 1 also shows the adult prevalence at which the error in the estimates exceeds 5 per cent. This ranges between 2 and 11 per cent prevalence, the younger age groups of mothers reaching significant levels of error only when incidence levels are higher.

In general, the error is larger at higher prevalence since the biases discussed earlier become more substantial. The plot of percentage error against prevalence is asymptotic for most age groups, however, flattening as the higher prevalences are reached. This is a consequence of the relationship between prevalence and the weights, $w_{x,t}$, defined in Equation (12). The weights are the proportion of those HIV-positive at age x that were HIV-positive at $x–t–0.5$, and reflect the balance between incidence before age $x–t$ and incidence after that age. Higher levels of incidence reduce the number of HIV-negative individuals at risk of infection at later ages. Women infected before age $x–t$ therefore form a larger proportion of all those HIV-positive at the time of interview. For all except younger ages, this effect is sufficient to increase the value of the weights, $w_{x,t}$, at higher prevalence. For larger values of t, these weights show an increasing gradient against prevalence, for ages above the early twenties. This means that the gradient of $PD(i)^r$ plotted against prevalence increases as prevalence increases, for all except the first two age groups. As the gradient comes to equal that of $q(z)^t$ against prevalence, which is linear, the percentage error in the estimates flattens out.

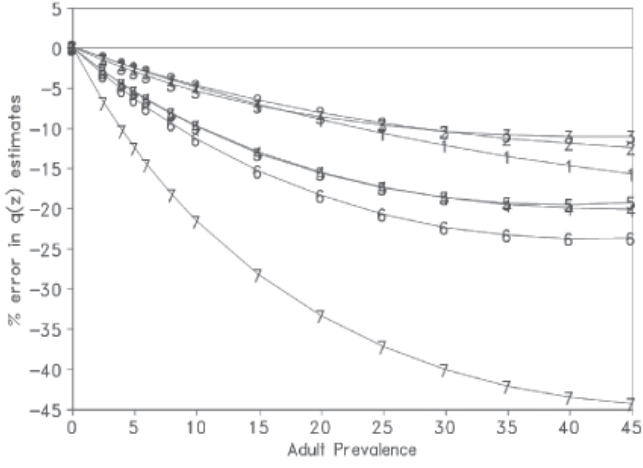
Older age groups generally have larger errors at any particular prevalence. This would be expected since cumulative HIV infection and HIV mortality is higher in the older age groups, increasing the reporting bias. However,

Table 1 Percentage error in the *q(z)* estimates by age group of mother for a range of levels of adult prevalence and the maximum prevalence at which the error is less than 5 per cent; all model parameters except the level of the incidence schedule held at default values

Adult Prevalence 15–49	*Age Group of Mother, i*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
	15–19	*20–24*	*25–29*	*30–34*	*35–39*	*40–44*	*45–49*
2.5	–1.2	–1.2	–1.5	–2.9	–2.9	–3.3	–6.8
5.0	–2.5	–2.5	–3.0	–5.5	–5.5	–6.3	–12.5
10.0	–4.9	–4.7	–5.3	–9.7	–9.8	–11.3	–21.5
20.0	–8.9	–8.0	–8.6	–15.5	–15.6	–18.4	–33.3
30.0	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
40.0	–14.6	–11.9	–11.0	–19.9	–19.5	–23.8	–43.5
45.0	–15.7	–12.4	–11.0	–20.1	–19.2	–23.6	–44.3
Max Prev.[a]	10.3	10.9	9.4	4.8	4.8	3.8	1.9

[a] Maximum adult prevalence at which error in the estimate is less than 5 per cent.

Figure 2 Percentage error in the *q*(*z*) estimates against adult female seroprevalence for the default HIV incidence model



The age group of the women from whom the estimates are derived is indicated by the number of each series.

since the measure used is percentage error, differences between age groups reflect the effect of increased incidence on both the bias and on $q(z)^t$, the denominator. The similar slopes of these two measures against prevalence results in the similarity in percentage error of the second and third and the fourth and fifth age groups, respectively, despite the absolute values being different. The larger percentage error in the estimates derived from the first age group compared to those of the second and third at higher prevalences results from the relatively small increase in the denominator, $q(z)^t$, for this age group, since the difference between true and estimated $q(z)^t$ values is smaller than for the other two age groups at all levels of prevalence. The large error in the seventh age group results from the impact of increased incidence on $q(20)^t$ through sexual transmission.

The sensitivity of the model to parameters other than the level parameter of the incidence schedule is illustrated in Appendix B. The errors in the estimates are calculated for different values of each parameter, while other parameters are held constant. The analyses show that the sensitivity of the errors varies according to the age group considered. Of the HIV incidence model parameters, the location and spread have a substantial effect on a number of age groups. Vertical transmission appreciably affects most age groups. The mean survival time of HIV-positive adults has little effect on the errors, but substantially affects adult prevalence. Of the demographic parameters, the level and slope of the underlying mortality model both have a substantial effect on the errors of all age groups.

In general, the errors in the estimates derived from the third, fourth and fifth age groups are most stable to variation in the model parameters. The corrections for these age groups can therefore be expected to be more robust to any errors in the specification of the model parameters.

CORRECTING THE ESTIMATES IN POPULATIONS WITH HIV

The possibility of deriving corrections for the errors introduced by HIV is examined by generating a set of 600 simulated stable populations. Each population is assigned values for those parameters which are allowed to vary between simulations. These values are selected randomly from within the range defined for that parameter. The level parameter of the incidence model varies between 0 and 2.52, corresponding to variation in stable adult female prevalence between 0 and 45 per cent. The level parameter of the underlying mortality schedule is varied between –0.5 and +0.4. The shape parameters

of the fertility schedule are allowed to vary, subject to the same restrictions as those used in deriving the regression coefficients for the multipliers, *k(i)*. The shape parameters of the incidence schedule may also vary, within the two sets of ranges defined in the fitting of the incidence model. The resulting four subsets of simulations ensure that all possible combinations of shape parameters are represented; they are detailed in Table 2. The level of vertical transmission, the mean survival times of HIV-positive adults and children and the slope of the underlying mortality schedule are not varied.

Table 2 Subsets of simulations defined by shape parameter ranges.

Sub-set	*Characteristics*	*Range of Values Allowed (min, max)*				*No. of Cases, N = 600*
		Fertility		*HIV Incidence*		
		Location	*Spread*	*Location*	*Spread*	
1	Core Ranges	–0.5, +0.2	0.8, 1.1	0.4, 0.7	0.7, 0.9	450
2	Extreme fertility	+0.2, +0.5	1.1, 1.2	0.4, 0.7	0.7, 0.9	50
3	Extreme HIV	–0.5, +0.2	0.8, 1.1	0.7, 0.8	0.9, 1.3	50
4	Extreme fertility and HIV	+0.2, +0.5	1.1, 1.2	0.7, 0.8	0.9, 1.3	50

Corrections for the estimates are based on the difference between true and estimated *q(z)* values, *n(z)*. This is preferable to deriving multipliers for populations with HIV based on the ratio of $q(z)^t$ to the reported proportion dead, $PD(i)^r$. For life tables which include mortality caused by HIV, this ratio can be shown to vary substantially with underlying mortality levels, even for a constant prevalence of HIV.[9] The difference between estimated and true *q(z)* values, *n(z)*, offers a better basis for correction in the estimates since it varies somewhat less with underlying mortality than do the multipliers and can be used additively rather than multiplicatively.

The difference, *n(z)*, is regressed onto a set of independent variables

9 This occurs despite proportionality of $q(z)^t$ values across all levels of underlying mortality, for $z<25$, which is a sufficient condition for constancy of the multipliers within a family of life tables in populations unaffected by HIV (Palloni and Heligman 1986). The variability of the multipliers is a consequence of the under-representation of the children of HIV-positive mothers in the reported proportions dead. Their mortality, $q(z)^+$, is therefore weighted more heavily in $q(z)^t$ than in $PD(i)^r$. For any given increase in non-HIV mortality, $q(z)^+$ will increase proportionately less than $q(z)^-$ because of the component of HIV-induced mortality in the former. The slope of $PD(i)^r$ against the level parameter of HIV-negative mortality is therefore steeper than that of $q(z)^t$ and the ratio between the two varies as HIV-negative mortality varies.

specifying the nature of the HIV epidemic. This could provide a basis to derive accurate estimates of child mortality in populations affected by HIV. A standard estimate could be calculated in the usual way and a corrected estimate derived using:

$$q(z)^t = q(z)^e + n(z) \quad \ldots(16)$$

where $n(z)$ is the predicted correction factor, $q(z)^e$ is the estimate derived from the conventional application of the technique and $q(z)^t$ is the corrected estimate.

Two regressions are used. The first takes the form:

$$n(z) = aPREV + b(PREV)^2 \quad \ldots(17)$$

where PREV is prevalence in women aged 15–49, expressed as a proportion. The results are given in Table 3.

Table 3 Regression coefficients for $n(z) = aPREV + b(PREV)^2$
Calculated estimates of $n(z)$ used in $q(z)^t = q(z)^e + n(z)$

Age Group i	*Age z of q(z)*	*a*	*b*	R^2
1	1	0.0501	–0.0169[a]	0.9019
2	2	0.0645	–0.0389	0.9384
3	3	0.0895	–0.0863	0.9876
4	5	0.1813	–0.1420	0.9949
5	10	0.1988	–0.1779	0.9913
6	15	0.2534	–0.1941	0.9778
7	20	0.5739	–0.3552	0.9697

All terms significant at .01 level except [a], p = 0.02.

The R^2 terms indicate that a substantial proportion of the variance in the $n(z)$ values from zero can be explained by a regression using information derived only from prevalence in adult women. As expected from the sensitivity analyses, the $n(z)$ values for the third, fourth and fifth age groups of mothers have the largest proportion of variance explained by the regression.

Much of the remaining variance in the errors can be ascribed to variation in

the shape of the incidence schedule. The shape of the prevalence distribution will reflect age-specific incidence patterns. Of a range of variables which reflect variation in the shape of the prevalence distribution, prevalence in women aged 15–19 was most consistently selected by a stepwise regression procedure. This variable, expressed as a proportion and denoted PREV15, was therefore entered into an extended regression; the results are given in Table 4. It can be seen that the extended regression increases the proportion of variation explained, most substantially for the first, second and seventh age groups.

Table 4 Regression coefficients for $n(z) = aPREV + b(PREV)^2 + cPREV15$
Calculated estimates of $n(z)$ used in $q(z)^t = q(z)^e + n(z)$

Age Group, i	*Age z of q(z)*	*a*	*b*	*c*	R^2
1	1	0.1134	–0.0226	–0.1112	0.9911
2	2	0.1202	–0.0438	–0.0978	0.9918
3	3	0.1107	–0.0882	–0.0373	0.9934
4	5	0.1720	–0.1412	+0.0163	0.9952
5	10	0.1749	–0.1758	+0.0420	0.9927
6	15	0.1504	–0.1849	+0.1807	0.9920
7	20	0.2552	–0.3269	+0.5594	0.9928

All terms significant at .01 level.

If the regressions are to be used to derive corrected estimates, the outcome of interest is the error in the adjusted $q(z)^t$ estimates rather than the R^2 values. The predicted corrections are applied to the simulated populations in order to assess this error. The predicted $n(z)$ values are calculated for each age group using both sets of regression coefficients. These $n(z)$ values are used to calculate estimates of corrected $q(z)^t$ values and the errors in these corrected values are calculated. Scatter plots of the errors against prevalence show that they are larger at higher prevalences and are used to find the maximum prevalence at which errors in the estimates are always less than ±5 per cent. These are conservative measures of the maximum prevalence at which the corrections can be applied, since the average error at this prevalence will be much lower. Table 5 shows these prevalences, together with the mean of the signed errors, the maximum of the absolute values of the errors and the

root mean square errors, across all simulated populations. For comparison, similar statistics are also given for the uncorrected estimates that result from applying the standard multipliers, *k(i)*, without adjusting the estimates using *n(z)*.

Table 5 Maximum of the absolute values of the percentage errors in the estimates, mean of the percentage errors, the maximum prevalence at which error in the estimates is always within ±5 per cent and the root mean squared errors in the estimates of $q(z)^t$ – for uncorrected estimates and two sets of corrected estimates

Age group i	*Maximum of absolute values of percentage error in estimates across all simulations*			*Maximum prevalence at which error in estimates is always within ± 5 per cent*			*Mean of signed percentage errors across all simulations*			*Root mean square of differences*		
	Correction			*Correction*			*Correction*			*Correction*		
	No[a]	*Bsc*[b]	*Ext*[c]	*No*[a]	*Bsc*[b]	*Ext*[c]	*No*[a]	*Bsc*[b]	*Ext*[c]	*No*[a]	*Bsc*[b]	*Ext*[c]
1	43.8	20.9	5.2	3	12	44	–9.2	+0.0	–0.0	.012	.004	.001
2	32.1	15.0	5.3	3	12	33	–8.2	–0.1	–0.1	.014	.003	.001
3	22.2	6.3	4.6	4	33	All[d]	–8.7	–0.1	–0.1	.016	.002	.001
4	32.0	4.8	4.7	3	All[d]	All[d]	–15.8	–0.2	–0.2	.035	.003	.002
5	35.0	7.7	5.7	3	30	34	–15.8	–0.3	–0.3	.037	.003	.003
6	48.0	16.7	9.5	2	12	12	–19.7	–0.3	–0.3	.050	.007	.004
7	69.1	22.8	9.1	1	3	12	–34.9	–0.2	–0.5	.121	.021	.010

[a] No correction – using the standard multipliers, *k(i)*, to estimate $q(z)^t$.

[b] Correction based on the basic regression $n(z) = aPREV + b(PREV)^2$

[c] Correction based on extended regression $n(z) = aPREV + b(PREV)^2 + cPREV15$

[d] All: Error less than 5 per cent in all estimates up to the maximum prevalence used (45 per cent)

Table 5 further illustrates the effect of HIV on the accuracy of the uncorrected estimates. When a range of parameters are varied simultaneously, the errors at any particular prevalence have a wider range. The prevalence at which none of them exceed 5 per cent, for any particular age group, is therefore lower than the corresponding prevalence for the default model (given in Table 1). Since these parameters may be unknown in practical application, the figures given in Table 5 represent more realistic limits to the prevalence

at which accurate estimates can be reliably expected. The standard technique cannot be relied upon once adult prevalence exceeds about 3 per cent; estimates from the last two age groups may occasionally be biased even at lower prevalence.

The mean and maximum errors given in Table 5 illustrate the improvements in the estimates brought about by using the corrections. Both are substantially reduced in the estimates derived using the corrections calculated from the basic regression equation, compared to the errors in estimates made using the standard procedure. If a 5 per cent error in the final estimates is used as the limit of acceptability, then it can be seen that the basic corrections produce accurate estimates from the reports of the fourth age group of women at up to 45 per cent prevalence and for the third and fifth age groups up to 30 per cent prevalence. Estimates from the first, second and sixth age groups are only reliable up to 12 per cent prevalence and the seventh age group is unreliable even at very low prevalence.

If information is available on seroprevalence in women aged 15 to 20, then the extended regression increases the prevalence at which accurate estimates can be made. Accurate estimates may be derived from the reports of the third and fourth age groups in populations with a prevalence of up to 45 per cent. In practice, this also applies to the reports of women in the first, second and fifth age groups, since the maximum errors in the estimates over the entire range of seroprevalence levels are only marginally above 5 per cent. Estimates derived from the reports of women in the sixth and seventh age groups cannot be corrected over so extensive a range.

DISCUSSION

The analyses have shown that HIV can be expected to introduce significant error into the estimates produced by the CS/CEB technique and that, to a certain extent, corrections can be derived using information on prevalence. These findings are subject to a number of qualifications and limitations. The sensitivity analyses showed that two parameters, in particular, are important when prevalence is high, yet their value and their variation between populations are not well established at present. These are the mean survival time of adults with HIV and the level of vertical transmission. The former affects the level of adult prevalence at any given level of incidence and thereby has an important impact on the estimates derived from the reports of older women. The corrections given for these groups are therefore

sensitive to moderate variations in its value. Substantial variations in the level of vertical transmission would also invalidate the corrections in populations with a high prevalence of HIV. The shape of the HIV-negative life table, governed by the slope parameter, was not varied in the simulations used to derive the corrections, although the sensitivity analyses showed that it had a substantial effect on the magnitude of the errors. The adjustments thus apply to populations in which the UN General model life table describes mortality in the HIV-negative population.

The most important limitation of the model is that the level of the age-specific incidence risks is constant over time, whereas most countries of East Africa experienced rising incidence rates before reaching an incidence plateau in the 1990s, and in much of Southern Africa incidence may still be rising (UNAIDS 2008). For any given level of prevalence, a stable population model will overestimate the extent to which older cohorts have been subject to infection when compared to real populations, since no real population has experienced constant incidence rates for such long periods. The model will therefore tend to overestimate the magnitude of the biases for any given level of prevalence, particularly for rapidly growing epidemics of short duration and for older women. For this reason, while they provide a first approximation, the corrections should not be relied upon to provide accurate corrections in real populations currently undergoing an epidemic of HIV. Further work to assess the effects of changing incidence levels over time is required.

In the usual application of the CS/CEB technique, the estimates of $q(z)$ are converted to refer to a common age, usually age 5. The time location of each estimate is calculated, giving a measure of the trend in child mortality over the period. In populations with HIV, the model life table used to convert estimates to $q(5)$ would need to include the effect of HIV, since it substantially affects the age pattern of mortality. It would need to reflect the overall prevalence of HIV in the population and may also need to reflect variation in the shape of the distribution of prevalence by age. If prevalence changes over time, a number of life tables reflecting the impact of different prevalence levels on the shape of the life table may be required (INDEPTH 2004; Zaba et al. 2007).

Where prevalence has been changing rapidly, heterogeneity in child mortality by age of mother is also a threat to the validity of the estimates. Differences in the proportion of children born to infected women in different age groups may mean that the mortality risks experienced by children born to a

given age-group of women are substantially different from those experienced by the population as a whole. In a situation of rising incidence we would expect that vertical transmission would occur most frequently at the ages of peak incidence, generally in the 20–24 age group, when the mother's viral load was high immediately following infection. But the reporting biases would only be manifest some 5 to 10 years later when the cohort of mothers itself became substantially depleted through AIDS mortality. This implies that during a period of rising incidence biases would tend to emerge first in those aged 30–34 at survey, before affecting older age groups.

This heterogeneity and the change in shape of the life table will also affect the standard procedure for estimating the time location of the mortality estimates. The time location procedure assumes that there has been a linear change in mortality over time (Feeney 1980); this may be violated in a population undergoing a substantial epidemic of HIV, where a long-standing secular decline in child mortality may be halted or reversed by HIV (Zaba et al. 2004). The impact of these factors is likely to depend on a number of characteristics of the epidemic, including the rate of increase in HIV prevalence levels and the time since the epidemic began. Although these issues are not addressed in this paper, they constitute substantial additional problems in applying the technique in real populations that have experienced changing incidence levels.

Other child mortality estimation techniques, such as the preceding birth technique (Bicego et al. 1989) and synthetic cohort methods (Zlotnik and Hill 1981) will also be affected by the same biases. Direct estimates of mortality from birth histories are affected by the under-reporting of child deaths due to correlated mortality, but not by the heterogeneity of mortality by mothers' age and they do not make use of model life tables. Direct estimates might therefore be more appropriate in populations affected by HIV, although work to assess biases in these estimates is also required (Mahy and Zaba 2004).

There is some evidence that HIV-positive women may be expected to have lower fertility than HIV-negative women (Carpenter et al. 1997; Gray et al. 1998; Gregson 1994). In the model, a single fertility schedule was used to describe the fertility experience of all women, both HIV-negative and HIV-positive, since the extent of any fertility reduction remains uncertain. If, however, HIV-positive women do generally have appreciably lower levels of fertility, then this will also affect the validity of the findings. The effect of such lower fertility would probably be to reduce the extent of the bias at

any given prevalence, since the contribution of the births of HIV-positive mothers to child mortality in the population as a whole would be lower.

A number of other simplifications were used in constructing the model. The fertility and the HIV-incidence parameters were allowed to vary independently of one another, despite work suggesting that a fairly close relationship between them would be expected (Zaba 1994). Only heterosexual and vertical transmission were included in the model, and transmission through breastfeeding following postnatal infection of the mother was ignored. It was assumed that the mean survival time of HIV-positive individuals had no effect on the extent of horizontal transmission. It was assumed that the shape of the distribution of incidence by age was independent of the level of incidence, neglecting any possible interaction, such as higher prevalence leading to younger incidence patterns. Simple exponential functions were used to model excess mortality from HIV, though there is evidence that progression of the disease is more complicated (UNAIDS 2002) and that a proportion of children may survive beyond the limit of five years used in the model (Marston et al. 2005). Aside from using estimates of HIV-positive mortality from developing countries, it was assumed that the level of HIV-negative mortality had no relationship with the excess mortality of HIV-positive individuals, ignoring any possible correlation introduced by the higher prevalence of infectious diseases in the general population. It was also assumed that the mortality of HIV-negative children born to HIV-positive mothers was the same as that of other HIV-negative children. Reports of HIV-positive women were included in proportion to their prevalence, ignoring any effect which illness could have on their representation in a sample survey. All of these simplifications could have some effect on the validity of the findings, although in most cases the effect would probably be small and the data with which to adjust for them are unavailable.

CONCLUSION

The CS/CEB technique is an important method of estimating child mortality in developing countries. The analyses have shown that HIV can be expected to introduce significant error into the estimates produced by the technique. The size of these errors depends on the prevalence of HIV, on the age group of the women reporting and on the particular characteristics of the population and of the epidemic. In general, higher levels of prevalence and lower levels of underlying HIV-negative mortality result in larger errors. The

reports given by older women generally show larger errors. Under certain conditions, significant errors can be introduced into the estimates by adult prevalence as low as 3 per cent.

The results of the simulation exercises suggest that the estimates can, to a certain extent, be corrected. In addition to the usual data needed to apply the technique, the corrections require information on the seroprevalence of HIV in adult women. Using this knowledge, accurate estimates of mortality can be derived from the reports of women in the third, fourth and fifth age groups at prevalences of over 30 per cent. Estimates derived from the other age groups are less robust to variation in the model parameters. However, estimates derived from the youngest and oldest age groups are in any case the least reliable in the normal application of the technique. Improvements can be made to the accuracy of most estimates if information on seroprevalence in women aged 15 to 19 is available. Using this information, accurate estimates can be derived from the reports of women in all age groups except the final two, in populations with a seroprevalence of over 30 per cent. However estimates derived from the reports of the first two age groups are sometimes biased for other reasons (Hill 1991).

These findings are subject to a number of qualifications and limitations. Most importantly, the level of incidence in the simulations is held constant over time. Because of this, the model will tend to overestimate the magnitude of the biases and of the corrections required for any given level of prevalence, compared to current real populations in which incidence has not been stable over such a long period. Further work is required to address the implications of this limitation.

Appendix A: Fitting the HIV Incidence Model

The incidence model is based on the Brass-Booth relational Gompertz fertility model and is described in the main text. The standard $G_s(x)$ values are modified, however, before fitting the incidence model to empirical seroprevalence distributions.This is necessary because, whichever values of the transforming parameters are used, the standard values are unable to fit some of the seroprevalence distributions which have relatively 'flat' patterns with respect to age. This can be overcome by 'stretching' the standard using the approach of Paget and Timaeus[10] to derive a male standard fertility schedule. The result is that the very low risks occurring to women in their forties are moved into the fifties. This may not be a good model for incidence in the fifties, but output is only generated for women up to age fifty. The resulting modified standard $G_s(x)$ values are given in Table A1.

The model is fitted to a set of empirical female age-specific seroprevalence distributions, found by searching the U.S. Bureau of the Census database and by a review of the literature. Only data from population-based studies of the general population are used – data from specific groups such as pregnant women, patients or prostitutes are excluded. The age-specific prevalence rates generated by the model seroincidence schedules are fitted to selected data, to determine a range of values for the model parameters which can be used in the simulations. The data used are given in Table A2. The fitting of the models assumes that prevalence increases over the duration of the epidemic; specifically, the level parameter of the incidence schedule is assumed to increase exponentially between the beginning of the epidemic and the time of the study. This is necessary because cross-sectional measures of seroprevalence by age will reflect the varying experiences of different cohorts as incidence increases during the epidemic. In particular, prevalence in older women will reflect the level of incidence during an earlier period, when they were younger and at highest risk, yet age-specific incidence rates were lower than current incidence rates.

10 In the work described in this paper, a ratio of 1.25 is used to 'stretch' the age scale of the standard schedule.

Table A1 Modified standard $G_s(x)$ values and model female annual age-specific HIV incidence risks, per 1000 women

Age	$G_s(x)$	*Risks*[a]	*Age*	$G_s(x)$	*Risks*	*Age*	$G_s(x)$	*Risks*
10	–	0.00	30	0.162	14.41	50	3.374	3.51
11	–3.405	0.18	31	0.267	13.71	51	3.706	2.82
12	–2.974	1.28	32	0.373	13.07	52	4.069	2.20
13	–2.637	4.54	33	0.480	12.48	53	4.466	1.67
14	–2.332	10.07	34	0.589	11.92	54	4.901	1.24
15	–2.056	15.72	35	0.700	11.52	55	5.386	0.91
16	–1.814	20.75	36	0.815	11.03	56	5.958	0.64
17	–1.595	24.09	37	0.935	10.54	57	6.664	0.43
18	–1.397	25.68	38	1.059	10.03	58	7.668	0.24
19	–1.216	25.89	39	1.188	9.51	59	9.500	0.06
20	–1.052	24.66	40	1.323	9.11	60	–	0.00
21	–0.904	23.85	41	1.467	8.55			
22	–0.766	22.90	42	1.617	8.03			
23	–0.635	21.85	43	1.777	7.53			
24	–0.509	20.67	44	1.948	7.02			
25	–0.388	19.28	45	2.133	6.67			
26	–0.273	18.18	46	2.339	6.09			
27	–0.161	17.13	47	2.564	5.45			
28	–0.052	16.14	48	2.810	4.78			
29	0.056	15.23	49	3.080	4.09			

[a] The incidence risks resulting from the default age-specific incidence shape parameters used in the sensitivity analyses. The location and spread parameters take values of 0.6 and 0.9 respectively; the level parameter takes a value of 0.527, yielding an adult prevalence of 10 per cent.

In fitting the model, a set of age-period-specific risks are applied to successive cohorts of women:

$$h_{x,t} = (h_{x-1,t-1} * e^{-1/\mu} + u_{x-1,t-1} * i_{x-1,t-1}) * \frac{l_x^-}{l_{x-1}^-} \qquad \ldots(18)$$

and

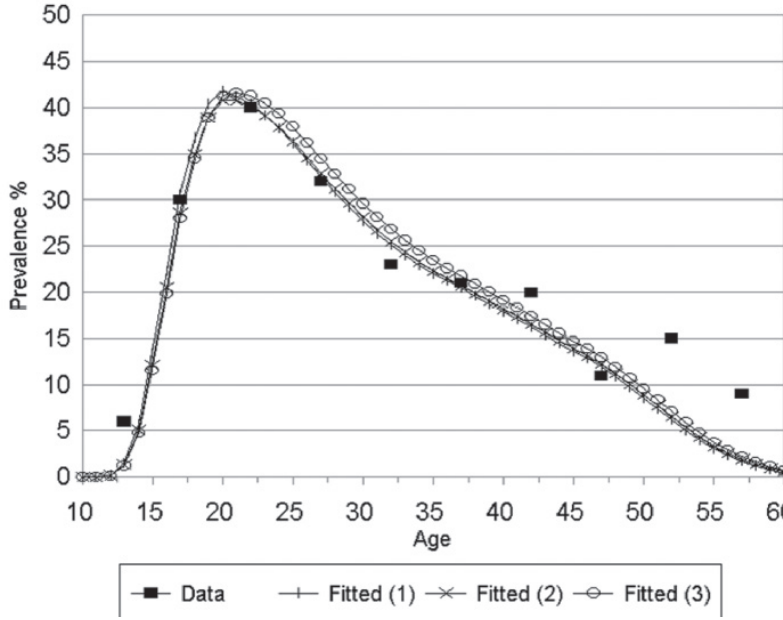
$$u_{x,t} = u_{x-1,t-1} * (1 - i_{x-1,t-1}) * \frac{l^-_x}{l^-_{x-1}} \quad \ldots(19)$$

where $h_{x,t}$, $u_{x,t}$ denote the proportions of a birth cohort alive and respectively HIV-positive and HIV-negative at age x and time t, $i_{x,t}$ denotes the risk of sexual infection with HIV at age x and time t, for x >10 and l^-_x represents the survivorship to age x in the HIV-negative life table. For each data set, the rate of increase in the level parameter required to produce the observed overall seroprevalence in women aged 15 to 50 is determined iteratively, assuming that the epidemic began in 1979 with a prevalence in the same age group of approximately 0.5 per cent. The location and spread parameters are then fitted iteratively, using a visual comparison between the empirical age-specific seroprevalence distribution and that generated by the model.

While the fitted rate of increase in the level parameter is sensitive to variations in the assumptions made about the duration of the epidemic and the initial prevalence, the shape parameters fitted to each distribution are much less sensitive. The range in the shape parameters needed to fit a particular distribution due to varying these assumptions is much narrower than the range of shape parameters fitted to the different empirical distributions. This is illustrated by Figure A1, which shows three alternative fitted seroprevalence distributions for one data set, resulting from variations in the assumed duration of the epidemic. Table A2 gives the values of the two shape parameters for all of the data sets to which the model was fitted.

Based on these fitted values, two sets of shape parameter values are defined for the modelling exercise. The first, covering the majority of distributions, has the location parameter between 0.4 and 0.7 and the spread parameter between 0.7 and 0.9. The second set, covering more extreme distributions, has the location parameter between 0.7 and 0.8 and the spread parameter between 0.9 and 1.3. In this way, the combination of a high value spread parameter with a low value location parameter is excluded, so that narrow prevalence distributions with maxima at much older ages than is observed in the empirical data are not generated. The default values of the parameters are set to 0.6 and 0.9 respectively. The age-specific incidence rates for these default values of the parameters, with a value of the level parameter chosen to give an adult prevalence of 10 per cent, are given in Table A1, together with the modified standard $G_s(x)$ values.

Figure A1 Variation in fitted seroprevalence distributions with variation in assumed duration of the epidemic



Empirical data is from KABP study, Rakai, Uganda, read from a graph in Berkley et al. (1990). Fitted model seroprevalence distributions assume the duration of the epidemic to be 8, 10 and 12 years respectively, with a rate of increase in the incidence level parameter to give 17 per cent prevalence in the year of the study (rates of 0.78, 0.59 and 0.48 respectively). All use identical values for the location and spread parameters (0.6 and 0.9 respectively).

Table A2 Female age-specific seroprevalence data used in fitting the incidence model: sources, adult prevalence and fitted parameters

Source	*Country*	*Adult Prevalence*[a] *(%)*	*Incidence Parameters*	
			Location	*Spread*
Barongo et al. (1992)	Tanzania			
Urban areas		15.2	0.8	1.3
Roadside areas		8.7	0.8	1.3
Rural areas		2.8	0.8	1.3
Benoit et al. (1990)	Côte d'Ivoire			
Urban areas		3.1	0.5	0.8
Rural areas		2.2	0.4	0.8
Berkley et al. (1990)	Uganda			
KAP, 1987		13.1	0.7	0.8
National survey		16.0	0.6	0.9
KABP, Rakai		27.9	0.6	0.9
Nunn et al. (1994)	Uganda	11.8	0.5	1.0

[a] Prevalence in women aged 15 to 49, or very similar age range. If this is not given, it was estimated from the given data, if necessary using the standard stable age structure.

Appendix B: Sensitivity Analyses

The sensitivity of the model to variation in each of the parameters is examined, while other parameters are held constant. The measure of effect is the percentage error in the estimates of *q(z)* for each age group of mothers. This is given for each of three values of each parameter – low, default and high – in Tables B1, B2 and B3 . The default value is the value used when the parameter is held constant as others are varied; it is indicated in bold in the tables. The high and low values represent the extremes of the probable range of the parameter.

The level parameter of the incidence schedule is varied between zero and 2.5, corresponding to adult prevalence ranging between zero and 45 per cent. The default value is set at 1.62, corresponding to a prevalence of 30 per cent. This is high and allows the effect of the other parameters to be clearly seen; at lower levels of incidence, the sensitivity of the outcome measures to other parameters is lower. Two alternative age structures are applied. The default is the standard stable age structure. The second is the stable age structure calculated from the life table of the population as a whole, which includes mortality caused by HIV, assuming a TFR of 6.

The effect of variation in the level of the incidence schedule is discussed in the main text. Table B1 illustrates that the effect of other parameters of the HIV model can also be substantial, the magnitude often varying with the age group considered. The mean survival time of HIV-positive adults and the location parameter of the HIV incidence schedule have small effects on the errors, except for those of the oldest age groups. The spread parameter of the incidence schedule and the mean survival time of HIV-positive children have substantial effects on the error in estimates derived from the first three age groups, but small effects on the fourth, fifth and sixth. In the case of the spread parameter, this reflects the variation which it causes in prevalence in young women. In the case of the mean survival time of HIV-positive children, the average age of the children will be higher for older age groups of mothers and so the effect of a parameter which affects mortality at early years, but not after age 5, will be much smaller. The spread parameter has a substantial effect on the error in the estimate for the seventh age group, reflecting its effect on horizontal transmission in young adults and hence on $q(20)^t$. The extent of vertical transmission has a substantial effect at all ages.

The effect of the mean survival time of HIV-positive adults appears to be the smallest. However, aside from the level parameter of the incidence schedule, this is the only parameter in the HIV model which has any substantial effect on adult prevalence. The values of the parameter given in Table B1 result in variation in the prevalence between 25 and 34 per cent.

Table B1 Results of the sensitivity analyses for HIV model parameters at 30 per cent adult prevalence[a]; percentage error in *q(z)* estimates for each age group of mother.

Parameter	*Value*	*Age Group of Mother, i*						
		1	*2*	*3*	*4*	*5*	*6*	*7*
Level of Incidence Schedule	0.5	–4.9	–4.6	–5.3	–9.7	–9.8	–11.3	–21.5
	1.6	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	2.5	–15.7	–12.4	–11.0	–20.1	–19.2	–23.6	–44.3
Location of Incidence Schedule	0.4	–14.1	–12.5	–11.4	–18.0	–17.2	–19.6	–34.1
	0.6	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	0.8	–9.2	–7.9	–9.2	–19.1	–19.9	–25.4	–45.5
Spread of Incidence Schedule	0.7	–4.8	–5.9	–8.9	–19.0	–19.8	–28.1	–47.8
	0.9	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	1.3	–20.1	–17.7	–13.2	–17.6	–15.9	–18.1	–27.4
Mean Survival of HIV+ Adults	6	–10.2	–9.5	–11.1	–19.9	–20.8	–24.7	–44.7
	8	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	10	–13.5	–11.0	–9.8	–17.4	–16.3	–19.4	–35.4
Mean Survival of HIV+ Children	1.5	–17.5	–14.4	–14.0	–17.1	–17.7	–21.9	–40.0
	2.5	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	3.5	–9.1	–7.5	–7.1	–19.6	–19.2	–22.6	–40.1
Vertical Transmission (%)	0.2	–8.5	–7.4	–7.6	–14.1	–14.0	–17.4	–37.5
	0.3	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	0.4	–15.3	–12.8	–12.6	–22.1	–22.3	–26.3	–42.2

[a] Except in the first panel, the level parameter of the incidence schedule takes the default value of 1.62, resulting in an adult prevalence of exactly 30 per cent for the default values of the other parameters. In the first panel, prevalence varies as a result of variation in the level of incidence: in other panels, the exact level of adult prevalence may vary as the other parameters vary.

Table B2 illustrates the errors in the estimates which result from the given values of the mean survival time, while prevalence is held constant by varying

the level parameter of the incidence schedule. At 10 per cent prevalence, the effects are generally small. They are appreciable, however, at 30 per cent prevalence for the fourth age group and upwards. The correction of these estimates using prevalence could therefore produce errors if incidence levels are high and the true value of this parameter varies from that used to derive the corrections.

Table B2 Error in *q(z)* estimates by age group of mother for alternative values of the mean survival time of HIV-positive adults at two levels of prevalence

Mean Survival of HIV+ Adults	*Age Group of Mother, i*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
	Adult Prevalence of 10 per cent						
6	–4.7	–4.9	–6.3	–11.2	–11.5	–13.3	–27.0
8	–4.9	–4.6	–5.3	–9.7	–9.8	–11.3	–21.5
10	–5.0	–4.6	–4.7	–8.7	–8.5	–9.8	–17.9
	Adult Prevalence of 30 per cent						
6	–11.7	–10.6	–11.9	–21.5	–22.4	–27.0	–48.1
8	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
10	–12.4	–10.3	–9.4	–16.7	–15.9	–18.9	–33.9

The effects of variation in the demographic parameters of the model are shown in Table B3. The level parameter of the mortality schedule has a substantial effect, with lower errors for higher values of the parameter, corresponding to higher levels of underlying mortality. This is not only a consequence of variation in the denominator; the difference between the true and estimated *q(z)* values is also smaller at higher levels of underlying mortality. This is because the reported proportions dead have a steeper slope against underlying mortality than does $q(z)^t$, as explained in the main text. The difference between the two is therefore lower when underlying mortality is higher.

The slope of the mortality model also has a significant effect on the errors. Corrections to the estimates will necessarily be specific to a particular family of life tables. The spread and location parameters of the fertility schedule have a small effect on the errors. The effect of varying the age structure used as weights is small.

The analyses show substantial differences in the sensitivity of the errors of the different age groups. The error in the estimates of *q(z)* derived from the third, fourth and fifth age groups are relatively insensitive to variations in the model parameters, with the exception of vertical transmission, the mean survival time of adults and the level of incidence. The errors in estimates derived from the reports of women in other age groups are sensitive to a wider range of parameters and the corrections will be less robust.

Table B3 Results of the sensitivity analyses for demographic model parameters at 30 per cent adult prevalence [a]; percentage error in *q(z)* estimates for each age group of mother

Parameter		*Age Group of Mother, i*						
		1	*2*	*3*	*4*	*5*	*6*	*7*
Level (alpha) of HIV-Mortality Schedule	–0.5	–24.0	–19.4	–18.3	–31.0	–32.0	–38.2	–60.5
	0.0	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	0.4	–7.2	–5.9	–6.3	–10.9	–10.4	–12.7	–24.6
Slope (beta) of HIV-Mortality Schedule	0.7	–11.3	–8.3	–8.5	–13.6	–12.8	–15.9	–30.7
	1.0	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	1.3	–16.1	–14.0	–13.4	–24.6	–25.2	–29.7	–49.7
Location of Fertility Schedule	–0.5	–11.5	–9.7	–9.4	–17.7	–17.5	–21.4	–39.3
	0.0	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	0.5	–12.3	–10.5	–10.7	–18.3	–18.3	–21.5	–39.6
Spread of Fertility Schedule	0.8	–13.9	–11.8	–11.2	–18.4	–17.7	–20.8	–39.0
	1.0	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	1.2	–10.9	–9.4	–9.5	–18.3	–19.0	–23.3	–40.9
Age Structure	**DF**	–12.1	–10.4	–10.4	–18.6	–18.6	–22.3	–40.0
	Stbl	–13.4	–11.5	–10.9	–18.7	–18.2	–21.6	–39.3

[a] See note on Table B1.

References

Barongo, L., Borgdorff, W., Mosha, F.F., Nicoll, A., Grosskurth, H., Senkoro, K.P., Newell, J.N., Changalucha, J., Klokke, A.H., Killewo, J.Z., Velema, J.P., Hayes, R.J., Dunn, D.T., Muller, L.A.S. and Rugemalila, J.B. 1992. "The epidemiology of HIV-1 infection in urban areas, roadside settlements and rural villages in Mwanza Region, Tanzania", *AIDS* **6**:1521–1528.

Benoit, S., Gershy-Damet, G.M., Coulibaly, A., Koffi, K., Sangare, V.S., Koffi, D., Houdier, R., Josseran, R., Guelain, J., Aoussi, E., Odehouri, K., Ehouman, A. and Coulibaly, N. 1990. "Seroprevalence of HIV infection in the general population of the Côte d'Ivoire, West Africa", *Journal of Acquired Immune Deficiency Syndromes* **3**(12):1193–1196.

Berkley, S., Naamara, W., Okware, S., Downing, R., Konde-Lule, J., Wawer, M., Musagaara, M. and Musgrave, S. 1990. "AIDS and HIV infection in Uganda – are more women than men infected?" *AIDS* **4**:1237–1242.

Bicego, G., Augustin, A., Musgrave, S., Allman, J. and Kelly, P. 1989. "Evaluation of a simplified method for estimation of early childhood mortality in small populations", *Int J Epidemiol* **18**(4 Suppl 2):S20–32.

Brass, W., Coale, A.J., Demeny, P., Heisel, D.F., Lorimer, F., Romanuick, A. and Van de Walle, E. 1968. *The Demography of Tropical Africa*. Princeton: Princeton University.

Brass, W. 1971. "On the scale of mortality", in Brass, W. (ed.). *Biological Aspects of Demography*. London: Taylor and Francis.

Brass, W. 1975. *Methods for estimating fertility and mortality from limited and defective data*. Carolina: Laboratory for Population Statistics.

Brass, W. 1981. "The use of the Gompertz relational system to estimate fertility" in International Population Conference, Manila, 3. Liege: International Union for the Scientific Study of Population.

Carpenter, L., Nakiyingi, J., Ruberantwari, A., Malamba, S., Kamali, A. and Whitworth, J. 1997. "Estimates of the impact of HIV infection on fertility in a rural Ugandan population cohort", *Health Transition Review*, supplement to vol. 7, 113–126.

De Cock, K., Zadi, F., Adjorlolo, G., Diallo, M., Sassan-Mokoro, M., Ekpini, E., Sibailly, T., Doorly, R., Batter, V., Brattegaard, K. and Gayle, H. 1994. "Retrospective study of maternal HIV-1 and HIV-2 infections and child survival in Abidjan, Côte d'Ivoire", *British Medical Journal* **308**: 441–443.

Feeney, G. 1980. "Estimating infant mortality trends from child survivorship data", *Population Studies* **34**(1):109–127.

Fox, M.P., Brooks, D., Kuhn, L., Aldrovandi, G., Sinkala, M., Kankasa, C., Mwiya, M., Horsburgh, R. and Thea, D. 2008. "Reduced Mortality associated with Breast-Feeding–acquired HIV Infection and Breast-Feeding among HIV-Infected children in Zambia", *J AIDS* **48**:90–96.

Fylkesnes, K., Mubanga-Musonda, M., Kasumba, K., et al. 1997. "The HIV epidemic in Zambia: sociodemographic prevalence patterns and indications of trends among childbearing women", *AIDS* **11**:339–345.

Gray, R., Wawer, M., Serwadda, D., et al. 1998. "Population-based study of fertility in women with HIV-1 infection in Uganda", *Lancet* **251**:98–103.

Gregson, S. 1994. "Will HIV become a major determinant of fertility in sub-Saharan Africa?" *The Journal of Development Studies* **30**(3):650–679.

Hill, K. 1991. "Approaches to the measurement of childhood mortality: a comparative review", *Population Index* **57**(3):368–382.

INDEPTH 2004. *Model life tables for Sub-Saharan Africa*. Ashgate publishing, UK.

Ioannidis, J.P., Tatsioni, A. and Abrams, E.J. 2004. "Maternal viral load and rate of disease progression among vertically HIV-1 infected children: an international meta-analysis", *AIDS* **18**:99–108.

Mahy, M. and Zaba, B. "Measuring Child Mortality in AIDS affected countries", UN Population Bulletin, 2004. Electronic publication: http://www.un.org/esa/population/publications/adultmort/Adultmortality.htm.

Marston, M., Zaba, B., Solomon, J., Brahmbhatt, H. and Bagenda, D. 2005. "Estimating the net effect of HIV on child mortality in African populations affected by generalised HIV epidemics", *J.AIDS* Feb 1;**38**(2):219–27.

Newell, M.L., Coovadia, H., Cortina-Borja, M., Rollins, N., Gailard, P. and Dabis, F. 2004a. "Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: a pooled analysis", *Lancet* 2004; **364**:1236–1243.

Newell, M.L., Brahmbhatt, H. and Ghys, P.D. 2004b. "Child mortality and HIV infection in Africa: a review", *AIDS* **18**(suppl 2):S27–S34.

Nicoll, A., Timaeus, I., Kigadye, R-M., Walraven, G. and Killewo, J. 1994. "The impact of HIV-1 infection on mortality in children under 5 years of age in sub-Saharan Africa: a demographic and epidemiological analysis", *AIDS* **8**:995–1005.

Nunn, A.J., Kengeya-Kayondo, J.F., Malamba, S., Seeley, J.A. and Mulder, D.W. 1994. "Risk factors for HIV-1 infection in adults in a rural Ugandan community: a population study", *AIDS* **8**:81–86.

Palloni, A. and Heligman, L. 1986. "Re-estimation of structural parameters to obtain estimates of mortality in developing countries", *Pop Bull United Nations* **18**:10–33.

Paget, W.J. and Timaeus, I.M. 1990. "A Relational Gompertz Model of Male Fertility: Development and Application to Time Location Parameters". Centre for Population Studies Research Paper, 90–2, London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Porter, K. and Zaba, B. 2004. "Empiricial evidence for impact of HIV on adult mortality in the developing world: data from serological studies", *AIDS* **18:** Sup 2, S9–S18.

Trussel, T.J. 1975. "A re-estimation of the multiplying factors for the Brass technique for determining childhood survivorship rates", *Population Studies* **29**(1):97–107.

UNAIDS Reference Group on Estimates, Modelling and Projections. 2002. "Improved methods and assumptions for estimating the HIV/AIDS epidemic and its impact", *AIDS* **16**:W1–W14.

UNAIDS. "2007 AIDS epidemic update: regional summary for Sub-Saharan Africa. WHO March, 2008". UNAIDS/08.08E / JC1526E Geneva, Switzerland. On line version http://data.unaids.org/pub/Report/2008/jc1526_epibriefs_ssafrica_en.pdf

Zaba, B. 1981. "Use of the relational Gompertz model in analysing fertility collected in retrospective surveys". Centre for Population Studies Working Paper no 81–2. London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Zaba, B. 1994. "The demographic impact of AIDS: some stable population simulation results". Centre for Population Studies Research Paper, 94–2, London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Zaba, B., Marston, M. & Floyd, S. "The effect of HIV on Child Mortality Trends in sub-Saharan Africa". UN Population Bulletin, 2004. Electronic publication: http://www.un.org/esa/population/publications/adultmort/Adultmortality.htm

Zaba, B., Marston, M., Crampin, A., Isingo, R., Biraro, S., Barninghaus, T., Lopman, B., Lutalo, T. and Todd, J. 2007. "Age-specific mortality patterns of HIV infected persons: a comparative analysis of community study data from five African countries", *AIDS* **21** Sup 6: S87–S96.

Zlotnik, H. and Hill, K. 1981. "The Use of Hypothetical Cohorts in Estimating Demographic Parameters under Conditions of Changing Fertility and Mortality", *Demography* **18**(1):103–122.