Measurement of adult mortality in populations affected by AIDS: an assessment of the orphanhood method

Ian M. Timæus and Andrew J. Nunn

Centre for Population Studies, London School of Hygiene and Tropical Medicine
MRC HIV Clinical Trials Centre, University College, London

Abstract

This paper demonstrates that orphanhood data can be used to estimate adult women’s mortality in populations experiencing an epidemic of AIDS. It develops both a correction for selection bias in reports of orphanhood and a revised procedure for estimating life table survivorship for use in populations with significant AIDS mortality. These new methods yield mortality estimates for a Ugandan population that are consistent with those obtained by prospective surveillance. Countries that lack effective death registration systems should ask about the survival of mothers in the census and surveys in order to monitor the effect of the AIDS epidemic on mortality.

Most developing countries lack effective systems of death registration. Measuring adult mortality in such populations has always been difficult (Timæus 1991a). The large-scale AIDS epidemics that have developed in much of Africa and some other countries have emphasized the public health importance of monitoring adult mortality levels and trends in the developing world. They have also made this task even more of a challenge. This paper commences by reviewing briefly both the existing limitations of the methods used to estimate adult mortality in the developing world and the additional obstacles to the production of such estimates that arise in populations with substantial mortality from AIDS.

All mainland sub-Saharan African countries and most other developing countries lack complete and accurate civil registration systems. In addition, most deaths occur outside hospital. Thus, no possibility exists of compiling comprehensive data on deaths by age routinely and continuously. A straightforward and relatively cheap alternative way of generating information on mortality is to ask questions in a national census or single-round household survey about deaths in the past year. Unfortunately, such questions have proved unreliable and often yield incomplete data (Timæus 1991a). Respondents may find it difficult to recall exactly when a death occurred and often have little idea of the ages of those who have died. Moreover, only very large surveys can estimate adult death rates with reasonable precision. The deaths of adults who lived alone are unlikely to be reported. Furthermore, not everyone is attached clearly to a single household. Thus, both omission of deaths and double-counting occur. Such ambiguities in the scope of questions about recent deaths in the household are particularly severe when a household head dies, as this event is likely to stimulate the fission and reformation of households.

* This study was initiated while Andrew Nunn was working with the Medical Research Council program on AIDS in Uganda. We are grateful to the directors of the program, Dr. D. Mulder and subsequently Dr. J. Whitworth, for facilitating this research and to all the staff of the program who were involved in the collection and processing of data used here.
It seems likely that the reliability of retrospective questions about deaths in the household will be particularly poor in populations suffering from an HIV epidemic. Little direct information on this issue yet exists. However, a stigma is attached to deaths from AIDS in many cultures: therefore, they may not get reported. AIDS tends to kill young adults and any spouse may respond by moving elsewhere. Thus, the tendency of households to disperse when someone dies is probably particularly strong when the death is from this disease. Moreover, because HIV is sexually transmitted it is common for both husbands and wives to become infected. As a result a significant proportion of the spouses of those who die may become seriously ill or die during the year following the earlier death. Neither death is likely to be picked up in a single-round enquiry.

In favourable circumstances, demographic analysis can be used to assess the completeness of reports of recent deaths (Preston 1984). Most of the techniques for this, such as the Growth Balance Equation, assume that trends in mortality have had little effect on the age distribution. They are unlikely to perform well in populations that are experiencing rapid increases in mortality concentrated in a fairly narrow range of ages. Techniques of evaluation that can cope with reversals in mortality trends do exist (e.g. Bennett and Horiuchi 1981). They can be applied only in populations where the age pattern of natural increase can be measured accurately. Thus, these methods have a poor record in regions such as Africa, where age reporting is poor and flows of international migrants are large (Timæus 1993).

Collecting information prospectively could prevent many of the reporting errors that afflict retrospectively collected direct data on adult deaths. Experience shows, however, that mounting a successful multiround survey of a representative sample of a national population is both expensive and difficult (Cleland 1996). This design multiplies the costs involved in interviewing a large sample. In addition, few multiround surveys have succeeded at maintaining high-quality field operations for the year or more needed to complete such a survey.

Studies involving the demographic surveillance of a geographically-localized population have proved more successful. Close supervision of the field operations by senior staff becomes possible. Such demographic surveillance systems have usually been set up as part of a program of research involving epidemiological and social as well as demographic studies. This greatly increases the benefits gained from demographic surveillance and makes it easier to justify the costs involved.

In response to the need for mortality data for health research, several prospective studies have been established in Africa that involve the demographic surveillance of a local population (e.g. Mulder et al. 1994; Sewankambo et al. 1994; Tollman, Herbst and Garenne 1995; Boerma et al. 1997). Those studies that collect serological data have generated valuable information on mortality and related aspects of the AIDS epidemic. The approach is not suited, however, to the measurement and monitoring of mortality in national populations. Indeed, complementary methods that are able to produce national estimates of mortality have to be deployed in order to gain the maximum benefits from intensive research studies. Without such data, it is impossible to gauge how far the health situation in areas subjected to intensive study is representative of wider populations, and thus what the import of the findings of such studies is for health policy and planning.

Several approaches to the measurement of adult mortality have been proposed that are intended to circumvent the problems that arise in collecting data on deaths directly (Timæus 1991a). One of the most important is the orphanhood method of estimating adult mortality (Brass and Hill 1973). This uses data collected by asking two simple questions, ‘Is your mother alive?’ and ‘Is your father alive?’; these questions do not require respondents to recall the dates when deaths occurred or the ages at death of deceased individuals. They can be asked in single-round enquiries. Conventional life-table measures of adult mortality can be
Measurement of adult mortality in populations affected by AIDS: the orphanhood method

Supplement 2 to Health Transition Review Volume 7, 1997

The idea of collecting information on parents to estimate adult mortality can be extended to other categories of relatives such as spouses and siblings. Asking about the deaths of first husbands and wives has seldom been a success. The widowed frequently report that they are single and the divorced that they are widowed. In particular, in much of Africa marrying someone is a lengthy process, which offers respondents opportunities to redefine their union history in socially acceptable ways. Moreover, co-infection of spouses with HIV, followed by the death of them both, will lead to substantial biases in adult mortality estimates from this source. Data about deaths of adult siblings would be less subject to this type of bias; however, methods for analysing them were only developed recently (Timæus, Zaba and Ali forthcoming) and few such data exist. A number of DHS surveys have collected full sibling histories that ask whether and when brothers and sisters died and about ages and ages at death. Such histories can be used to measure adult mortality without recourse to indirect estimation methods. Despite the difficulty of collecting this information accurately, the approach has yielded promising results (Rutenberg and Sullivan 1991; Timæus 1997).

Data on orphanhood have been collected far more frequently than data on the deaths of other adult relatives. Their quality has varied (Timæus 1991a): in East Africa, in particular, it is common for too few respondents to report that their parents are dead. It appears that a proportion of them instead answer the questions in terms of their foster-parents or step-parents. In many applications, however, comparison of successive sets of orphanhood data and of orphanhood-based and other estimates of adult mortality suggests that the method has worked well. Information on orphanhood has made a large contribution to our existing knowledge of adult mortality in Africa (Timæus 1993). If for no other reason than the inertia of data collection agencies, it also seems likely that it will continue to be collected. It therefore seems important to assess whether this source of mortality data continues to be useful in the face of the AIDS epidemic.

An inherent limitation of the orphanhood method is that data on parents’ survival can only be collected from those of their offspring who themselves remain alive. In general, the selection bias that arises from this is small (Palloni, Massagli and Marcotte 1984). In populations affected by AIDS, however, the problem is likely to be more severe. HIV-positive women can transmit the virus to their children during pregnancy, at delivery, or shortly after birth. Thus, the children of women who are at highest risk of dying are also likely to suffer higher mortality than the population in general. Because women are likely to have been infected by or to infect their husbands, orphanhood-based estimates of adult men’s mortality in populations affected by HIV may also be biased downwards, though to a lesser degree than those for women. A second major HIV-related bias in orphanhood estimates is that the data are usually converted into conventional life table indices of mortality using a set of coefficients based on very different age patterns of mortality in adulthood from those found in populations experiencing an AIDS epidemic.

A number of investigations have been conducted of the effect of the AIDS epidemic on orphanhood (e.g. Prebble 1990; Palloni and Lee 1992; Gregson, Garnett and Anderson 1994). The aim of this paper is to answer the inverse question: can information on orphanhood be used to measure adult mortality in populations affected by AIDS? Our investigation of this issue makes use of data from a longitudinal study of a population in rural Uganda that has collected information on serostatus, seroconversion, fertility, and mortality. This study is described briefly in the next section of the paper. The following three sections of the paper are theoretical. The first of them presents a mathematical model of the determinants of orphanhood in populations with significant AIDS mortality. Then, using values for the
parameters based on the Ugandan study, this model is deployed to quantify the main HIV-related biases in orphanhood estimates of adult women’s mortality and to develop methods that correct for them. Finally, these methods are applied to orphanhood data collected in the Ugandan study. The mortality estimates that result are evaluated by comparing them with direct measures obtained prospectively by surveillance of the population.

The MRC study in Masaka district

The area of study is a rural subcounty of Masaka district in southwest Uganda. It is situated approximately 20 miles from Masaka town and ten miles from the trans-African highway. The study covers 15 neighbouring villages with a population of about 10,000. Most of the population are peasant farmers; they grow bananas as a subsistence crop and cultivate coffee for sale. The predominant tribal group, the Baganda, constitute about 70 per cent of the population. Substantial immigration of Rwandese occurred up to the 1970s.

Beginning in late 1989, the study villages were mapped and a de jure census and socio-economic questionnaire was administered to an adult member of each household, preferably its head. Within four weeks of these interviews, a medical team visited each household. All residents were invited to participate in a survey that included a brief medical history, a physical examination, and the collection of a blood sample. Absentees and refusals were revisited to encourage them to participate.

Blood specimens were transported at weekly intervals to the laboratory of the Uganda Virus Research Institute in Entebbe where they were tested for antibodies to HIV-1. All sera were tested using two EIA systems, Recombigen HIV-1 EIA (Cambridge Biotech Corporation, Worcester MA) and Wellcozyme HIV-1 Recombinant (Wellcome Diagnostics, Dartford, England) with Western blotting using Novopath HIV Immunoblot (Bio-Rad laboratories, Watford, England) when indicated (Nunn et al. 1993, 1994). None of the field workers was aware of the HIV status of study participants. Trained counsellors made results available to all those who requested them (Seeley et al. 1991).

At annual intervals (Rounds 2-6, 1990-95), the census team revisited each household to ascertain the vital status of those who were resident at the previous survey and to enumerate those who had joined the household through either birth or in-migration. As in Round 1, the medical team collected a blood sample from everyone willing to provide one. Children aged less than 13 years were not bled after Round 3.

At the beginning of Round 3, monthly registration of births and deaths was introduced to supplement the data obtained from the annual surveys. Questions were asked in Rounds 2 and 3 about whether the mothers and fathers of children aged less than 15 years were alive and the dates when deceased parents died (Kamali et al. 1996). The results presented here are based on the orphanhood data collected in Round 3. The information was obtained successfully for more than 98 per cent of children.

The mortality, seroprevalence and HIV incidence rates used here are based on person-years of observation from the time individuals were enrolled in the study (the date of their first seropositive or seronegative specimen) until the date of the fifth annual resurvey (Round 6), for those known to be alive; the date of death, for those known to be dead; or, for those migrating out of the study area, their date of departure. Individuals who seroconverted are counted as seronegative until the midpoint between their last negative and first positive antibody test; thereafter they are counted as seropositive. The seroprevalence rate among adults was about eight per cent in each round of the study (Mulder et al. 1995). Nearly 12 per cent of the women of childbearing age are infected with HIV. Death rates among infected adults are an order of magnitude higher than those among the seronegative population (Mulder et al. 1994; Nunn et al. 1997).
Methods of analysis

Let the number of children born at \( a \) years before a demographic survey to women who were then aged \( y \) be \( C(a,y) \). Integrating over all ages at childbearing \( \alpha \) to \( \beta \), the total number of mothers who are still alive is (Brass and Hill 1973):

\[
N(a) = \int_{\alpha}^{\beta} C(a,y) \cdot p_{y} \, dy
\]

where \( p_{y} = l(y+a)/l(y) \), that is the life table probability of surviving from age \( y \) to \( y+a \).

To model the effect of the HIV epidemic on maternal orphanhood, we distinguish infected (seropositive) women, including those who have progressed to AIDS, from women who are uninfected (seronegative). Women who were seropositive when they gave birth \( a \) years ago have lower survivorship, \( a \cdot p_{y}^{+} \), than the equivalent seronegative women, whose survivorship is denoted \( a \cdot p_{y}^{-} \). In addition, seropositive women probably have lower fertility than other women (Gregson 1994). Thus, they have relatively few children to report on them. If \( C^{+}(a,y) \) represents the seropositive women who would have given birth \( a \) years ago at age \( y \) if it was not for their reduced fertility and \( F(a,y) \) is the age-specific ratio of the fertility of seropositive to seronegative women at that time, the total number of seropositive women who gave birth \( a \) years ago and remain alive is:

\[
N^{+}(a) = \int_{\alpha}^{\beta} F(a,y)C^{+}(a,y) \cdot a \cdot p_{y}^{+} \, dy
\]

However, the number of living seropositive women who would have given birth \( a \) years ago if their fertility was the same as that of other women is:

\[
N^{+}(a) = \int_{\alpha}^{\beta} C^{+}(a,y) \cdot a \cdot p_{y}^{+} \, dy
\]

Those women who were not infected when they gave birth may become so later. If \( C^{-}(a,y) \) is the number of uninfected women who gave birth \( a \) years ago at age \( y \), the total number of mothers who gave birth at that time remaining alive and uninfected is:

\[
N^{-}(a) = \int_{\alpha}^{\beta} C^{-}(a,y) \cdot a \cdot p_{y}^{-} \cdot a \cdot v_{y} \, dy
\]

where \( a \cdot v_{y} \) is the probability of remaining uninfected between ages \( y \) and \( y+a \):

\[
a \cdot v_{y} = e^{-\int_{y}^{y+a} \mu(z) \, dz}
\]

and \( \mu(z) \) is the instantaneous rate of infection at age \( z \). Mothers who become infected between ages \( y \) and \( y+a \) subsequently have an increased risk of dying. We assume that this risk is the same as that of other infected women of their age. Thus, the number of mothers who gave birth \( a \) years ago and have become infected since but remain alive is:
The proportion still alive of all mothers who were not infected when they gave birth is:

\[ S^i(a) = \frac{N^-(a) + N^i(a)}{\int C^-(a,y)dy} \]

In total, the proportion still alive of women who would have given birth \( a \) years ago in the absence of any impact of HIV infection on fertility is:

\[ S(a) = \frac{N^+(a) + N^-(a) + N^i(a)}{\int C^+(a,y) + C^-(a,y)dy} \quad (1) \]

Note that \( C^+(a,y) + C^-(a,y) = C(a,y) \).

A proportion, \( h \), of the children of seropositive mothers have the virus transmitted to them perinatally. In general, this study assumes that \( h = 25 \) per cent but it also explores the effect of a vertical transmission rate of 35 per cent (Ryder and Temmerman 1991; Lepage et al. 1993; Newell forthcoming). Few of these children live long (Newell forthcoming). If we make the somewhat pessimistic assumption that all of them die in the first five years of life, then \((1-h)l(a)\) of them remain alive at age five and over.\(^1\) As discussed in the previous section of the paper, we assume that, apart from vertical transmission, the mortality of orphans and children with living parents is the same. Thus, the proportion of respondents aged \( a \) who report that their mother is alive is:

\[ S^i(a) = \frac{(1-h)N^+(a) + N^-(a) + N^i(a)}{\int (1-h)F(a,y)C^+(a,y) + C^-(a,y)dy} \quad \text{for } a \geq 5 \quad (2) \]

To assess the biases in orphanhood estimates of mortality, we compare estimates of \( S(a) \) and \( S^i(a) \) calculated from Equations 1 and 2 with each other and with life table survivorship, using values for \( p^+_y, p^-_y \) and \( l(a) \) obtained in the Masaka district study (Mulder et al. 1994; Nunn et al. 1997). The same study provides information on the serostatus and ages of women giving birth. Much of the initial impact of the HIV epidemic on age-specific mortality and age structure respectively has already been felt. Although further changes will occur, current demographic patterns can be used to model the approximate effect of the epidemic on patterns of orphanhood in the longer run.

Because the counts of births by age for seropositive women have large sampling errors, we assume that, although seropositive women have lower fertility than uninfected women, their age pattern of fertility is the same. Thus, \( C^+(a,y) \) and \( C^-(a,y) \) are calculated by

\(^1\) This assumption will exaggerate rather than underestimate biases in the orphanhood data.
smoothing the age-specific fertility distribution for all women and applying it to the populations of seropositive and seronegative women respectively. The indirectly standardized fertility rate for seropositive women in the study in Masaka district is 27 per cent lower than the rate for seronegative women but has a wide confidence interval (Carpenter et al. 1997).

Our main assumption is that seropositive women have 20 per cent lower fertility than seronegative women but we assess the sensitivity of our results to this assumption by also modelling the effect of a 40 per cent reduction in fertility (Ryder et al. 1991; Sewankambo et al. 1994; Serwadda et al. 1997). Thus, \( F(a, y) \) simplifies to \( F = 0.8 \) or \( F = 0.6 \) for all \( a \) and \( y \), \( N'(a) \) becomes \( F.N'(a) \), and Equation 2 simplifies to:

\[
S'(a) = \int \frac{(1-h)F.C^+(a, y) + C^-(a, y)dy}{(1-h)F.N^+(a) + N^+(a) + N^-(a) + N'(a)} \quad \text{for } a \geq 5
\] (3)

To model \( N^-(a) \) and \( N'(a) \) one needs data on the incidence of HIV infection by age. Incidence rates are available for a subset of the cohort under study in Masaka district (Kengeya-Kayondo et al. 1996). Applied to a synthetic cohort, they yield a net lifetime risk of infection for women of 29 per cent. In contrast, the current mortality and seroprevalence data suggest that 37 per cent of adult women in the Masaka district study will die of AIDS. This discrepancy emphasizes that the demography and epidemiology of HIV in this area have yet to stabilize. It may reflect a recent decline in the incidence of HIV that has yet to be reflected in prevalence (Mulder et al. 1995). In order to match the incidence rates to the current prevalence of infection, therefore, they were scaled up to give a net lifetime risk of infection of 40 per cent. 2 This yields a prevalence-incidence mean survival time from infection of 6.7 years.

The integrals in Equations 1 and 3 are evaluated by summing data for five-year age group of mothers for respondents at five-year age intervals, \( a = 5, 10, \ldots 30 \), before interpolating to obtain results for age groups of respondents. 3 Most mothers of respondents

\[ \frac{1}{5} \left[ U_i^x - q_i^- \right] \]

where \( U_i^x = 1 - y_i^- \) and \( q_i^- = 1 - q_i^- \). Probabilities of dying before becoming infected are calculated using the equivalent equation. The proportion of respondents in each five-year age group with living mothers, \( S_i^x \), is calculated by interpolating between the point estimates for five-yearly intervals assuming that births grew at a constant rate, \( r \), of 2 per cent across each age group:

\[
S_i^x = \frac{l(x)S(x) + e^{-2.5r}(l(x)+l(x+5))}{l(x) + e^{-2.5r}(l(x)+l(x+5))} \frac{e^{-5r}(l(x+5)S(x+5))}{l(x) + e^{-2.5r}(l(x)+l(x+5)) + e^{-5r}(l(x+5))}
\]

2 The incidence rates have large sampling errors and were smoothed using a relational Gompertz model against a standard fertility schedule that was stretched to age 60 to allow for the infection of women aged 50-59 (Paget and Timæus 1994). The resulting single-year incidence schedule is used to calculate five-year cohort-period probabilities of HIV infection.

3 To implement the discrete models, the probabilities that the members of a cohort become infected in each five-year age interval allowing for prior mortality from diseases other than AIDS, \( sU^x \), are calculated as:

\[
S_i^x = \frac{1}{5} \left[ U_i^x - q_i^- \right] \]
aged 30 are 50 to 65 years old. Few of them will die of AIDS in the future. Thus, the relationship between life table survivorship and orphanhood in the age group 25-29 years should indicate the nature of this relationship in older age groups.

There are problems with using data from the Masaka district study to implement our model of orphanhood. The epidemiology and natural history of HIV in this area may be atypical. The data are based on a fairly small number of person-years of observation for the measurement of demographic statistics. Moreover, while trends in the prevalence of HIV infection in the study population have been muted for the last few years, the age distribution of women giving birth has been changing and will continue to do so. Nevertheless, it seems useful to model biases in orphanhood data using demographic parameters from a real population rather than parameters derived from an epidemiological model of the HIV epidemic that itself cannot be validated, especially with respect to AIDS mortality.

HIV-related selection bias

The first column of results in Table 1 shows estimates by age of respondent of the proportions alive of mothers who have never contracted HIV. These women are subject to the background mortality of the population and the proportions were calculated using data on the age structure and mortality of the seronegative population in Masaka district. The second column shows estimates of the proportions of people born to seronegative women who still have living mothers. The increase in orphanhood, compared with the previous column, reflects the mortality of women who became infected with HIV after they gave birth to the index child. The third column of simulation results refers to women who were already infected when they gave birth. Nearly half of them die while the index child is aged less than five years. The fourth column shows the overall proportions still alive of women who would have given birth if serostatus had no impact on fertility. These data equate to the proportions of living mothers that would be obtained in a survey if infected and uninfected women were equally likely to be reported on. In contrast, the last but one column shows the estimated proportions of mothers who would actually be reported to be alive in a survey, allowing for both the higher mortality of vertically infected children and the reduced fertility of infected women. Note that these estimates differ from those that would have been obtained by collecting data in Masaka district in the early 1990s. They represent the prevalence of orphanhood that would result from the current demography of the population: orphanhood in the early 1990s reflected demographic patterns during the previous 30 years.

Table 1

<table>
<thead>
<tr>
<th>Age group (x to x+5)</th>
<th>0-4</th>
<th>5-9</th>
<th>10-14</th>
<th>15-19</th>
<th>20-24</th>
<th>25-29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother subject to background mortality ((S'_x))</td>
<td>99.2</td>
<td>97.2</td>
<td>95.1</td>
<td>92.7</td>
<td>89.6</td>
<td>85.1</td>
</tr>
<tr>
<td>Mother not infected at birth ((S'_x))</td>
<td>98.6</td>
<td>94.7</td>
<td>89.5</td>
<td>83.9</td>
<td>78.0</td>
<td>71.4</td>
</tr>
<tr>
<td>Mother infected at birth ((S'_x))</td>
<td>81.6</td>
<td>44.2</td>
<td>23.2</td>
<td>11.6</td>
<td>5.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Total if no process of selection ((S'_x))</td>
<td>96.1</td>
<td>87.6</td>
<td>80.2</td>
<td>73.8</td>
<td>67.9</td>
<td>61.8</td>
</tr>
</tbody>
</table>

Note: These estimates differ from those that would have been obtained by collecting data in Masaka district in the early 1990s. They represent the prevalence of orphanhood that would result from the current demography of the population: orphanhood in the early 1990s reflected demographic patterns during the previous 30 years.
Total that would be reported\(b\) & 97.0 & 90.2 & 83.6 & 77.5 & 71.6 & 65.3 \\
\(\int \xi S_x^*\) & - & 87.6 & 78.7 & 73.0 & 67.4 & 61.5 \\
Adjusted total reports \(\int \xi S_x^*\) & - & 87.6 & 78.7 & 73.0 & 67.4 & 61.5 \\

\(a\) Assuming that seropositive and seronegative women have identical fertility and that their children have identical mortality. \\
\(b\) Assuming that the fertility of women infected with HIV is reduced by 20 per cent and a 25 per cent vertical transmission rate.

The most important conclusion to be drawn from Table 1 is that one should expect fairly small biases in reports on the survival of parents in populations affected by AIDS. According to the Masaka district based simulation, the errors in the reported proportions of mothers alive rise from about one per cent for respondents aged less than five years to nearly four per cent in the three oldest age groups considered. Thus, the simulated reports are much closer to unbiased figures for the population than to what would be reported in the absence of the HIV epidemic. This should also be true of populations with epidemics that are more severe or less severe than that in Masaka. The reason why the selection bias is fairly small is that reduced fertility and vertical transmission are unlikely to select out more than half of the reports on infected women: the simulated results shown in Table 1 assume that 40 per cent of such reports are missing. From a lifetime perspective, our data on incidence and mortality imply that women who become infected with HIV bear about two-thirds of the children that they would have had otherwise. They have about 60 per cent of these births before becoming infected.

Simplification of Equations 1 and 2 yields a useful approximation for the selection biases in data on orphanhood. Assume that all the women who are seropositive when they give birth die in the following five years, so that \(N^*(a)=0\) for respondents aged at least five years. This assumption makes the numerators of Equations 1 and 2 identical so that, dividing the former by the latter:

\[
\frac{S(a)}{S^*(a)} = \frac{(1-h)\int F(a, y)C^+(a, y)dy + \int C^-(a, y)dy}{\int C^+(a, y)dy + \int C^-(a, y)dy} \\
\tag{4}
\]

Assume also that the seroprevalence rate, \(P\), among women seen in antenatal clinics is a measure of the proportion of all women giving birth who are infected, so that:

\[
P = \frac{\int F(a, y)C^+(a, y)dy}{\int F(a, y)C^+(a, y)dy + \int C^-(a, y)dy} \\
\tag{5}
\]

Dividing Equation 4 though by the denominator of Equation 5, one obtains:

\[
\frac{S(a)}{S^*(a)} = \frac{(1-h)P + (1-P)}{P + (1-P)} = \frac{1-hP}{1+\frac{1-F}{F}P} \\
\tag{6}
\]
Thus, the bias in the reported proportion of respondents with living mothers increases with prevalence at a rate determined by \( F \) and \( h \). As one would expect, the higher vertical transmission, and the lower the fertility of infected women, the greater the bias in reports on the survival of mothers. In practice, estimates of seroprevalence are more widely available than estimates of vertical transmission or of the reduction in the fertility of the infected. Substituting our assumptions that the fertility of infected women is 20 per cent lower than that of uninfected women and that the vertical transmission rate is 25 per cent into Equation 6, produces a simple correction factor for selection bias:

\[
S_x' = \frac{1 - 0.25P}{1 + 0.25P} \frac{S_x^*}{S_x} 
\]

(7)

To allow for the survival for more than five years of a significant proportion of mothers who were seropositive when they gave birth, the adjustment can be halved for respondents aged 5 to 9 so that \( S_x' = (1 - 0.25P) \frac{S_x^*}{S_x} \). These correction factors can be revised using Equation 6 when more is known about \( F \) and \( h \).

The final column of Table 1 shows the adjusted proportions of mothers alive, \( S_x' \), that result from correction of the data in the previous column using Equation 7. They are close to \( S_x \), the unbiased values calculated using Equation 1. Table 2 investigates whether the correction factor of 1-0.5 is adequate results in populations where the reduction in the fertility of the seropositive and the vertical transmission rate is unknown. Even in the presence of HIV-related mortality, nearly all the mothers of young respondents are alive. Thus, both the absolute and the relative errors in the data on survival of mothers seem small even though they imply substantially different mortality from that actually prevailing. To convey the significance of the errors more accurately, therefore, they are presented in terms of the ratio of the estimated to actual odds of mothers surviving:

\[
\frac{S_x'}{1 - S_x'} \frac{1 - S_x}{S_x}
\]

Table 2

<table>
<thead>
<tr>
<th>Age group</th>
<th>Baseline assumptions</th>
<th>Baseline assumptions</th>
<th>40% reduction in fertility of the seropositive</th>
<th>35% vertical transmission of HIV</th>
<th>Higher non-HIV mortality ( (e_{10}=46.9) )</th>
<th>Twice the prevalence of HIV</th>
</tr>
</thead>
</table>

* If a population-based estimate of seroprevalence among women of childbearing age, \( P^* \), is available, rather than one based on antenatal data, one must assume that the proportion of births to seropositive women aged \( y \) can be estimated from \( P^* \), so that \( C'(a,y) = P^*C(a,y) \). Dividing Equation 1 by Equation 3, one obtains:

\[
\frac{S(a)}{S'(a)} = 1 - (1 - h)F \frac{P^*}{S_x}
\]

With \( F=0.8 \) and \( h=0.25 \), the correction factor is:

\[
S_x' = (1 - 0.4P^*) \frac{S_x}{S_x}
\]
Table 2: Ratios and Adjusted Ratios of Orphanhood Data

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Ratio</th>
<th>Adjusted Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-9</td>
<td>1.298</td>
<td>0.989</td>
</tr>
<tr>
<td>10-14</td>
<td>1.257</td>
<td>0.914</td>
</tr>
<tr>
<td>15-19</td>
<td>1.223</td>
<td>0.960</td>
</tr>
<tr>
<td>20-24</td>
<td>1.192</td>
<td>0.980</td>
</tr>
<tr>
<td>25-29</td>
<td>1.164</td>
<td>0.989</td>
</tr>
</tbody>
</table>

* Baseline assumptions: 11.8% seroprevalence among women aged 15-49 years, 20% lower fertility among the seropositive, 25% vertical transmission, background life expectancy at age 10 \(e_{10}\) of 54.2 years.

* Adjusted using a correction factor of 1 - 0.25\(P\).

The first two columns of ratios in Table 2 correspond to the last two columns of Table 1. According to our model, the odds that the mother of a living respondent is alive are 15 to 30 per cent higher than the odds for all women who would have given birth if serostatus did not affect fertility. The errors in the odds decrease slightly with age. Adjustment of reported data using the correction factor given in Equation 7, however, should produce nearly correct estimates. The next two columns of Table 2 shed light on how sensitive adjusted data are to the assumptions made about size of the reduction in fertility of seropositive women and the level of vertical transmission. The third column of ratios also uses data from Masaka district on the ages and mortality of infected and uninfected women to calculate \(S_x^\ast\) and \(S_x'^\ast\). However, it is assumed that the fertility of seropositive women is 40 per cent below that of seronegative women, rather than 20 per cent lower. Similarly, the next column illustrates the impact of a 35 per cent, as compared with a 25 per cent, vertical transmission rate. As one would expect, the survival of mothers tends to be overestimated even after correction by 1 - 0.5\(P\). In both scenarios, however, the errors in the odds are moderate and the adjusted estimates much more accurate than the unadjusted ones.

The derivation of the correction factor suggests that it should be valid whatever the level of background mortality or seroprevalence. The last two columns of Table 2 confirm this. A correction factor of 1 - 0.5\(P\) operates equally well in a model population with much higher background mortality than the baseline model and nearly as well in a model population in which the net lifetime risk of infection is raised to 67 per cent in order to double the prevalence of infection among women of childbearing age.

The model data support the theoretical argument that Equation 6 provides a basis for correcting orphanhood data for selection bias. They also suggest that the further simplification presented in Equation 7 is a robust one: while information on the fertility of infected women and vertical transmission rate can be used to refine the adjustment, it is not essential. This method of adjustment would perform less well for young respondents in a population where infected women survived longer than in Masaka district. Too little is known about the natural history of HIV infection in developing countries to investigate the importance of such differences thoroughly. It seems unlikely, however, that survival times vary enough across the developing world to affect greatly the conclusions drawn from Table 2.

**Bias in the estimation procedure**

The second source of bias in orphanhood-based estimates of adult mortality for populations affected by AIDS lies in the regression equations used to convert proportions of respondents with living mothers into life table probabilities of surviving from age 25. These equations take the form:

\[
P_{25} = \beta_0(n) + \beta_1(n)M + \beta_2(n)S_{n.5}
\]
where $M$ is the mean age at childbearing and is a control for variation between populations in the ages over which the mothers are exposed to the risk of dying. Deaths from AIDS are concentrated among young adults. Therefore, the epidemic produces age patterns of mortality that differ from those in the mortality models used to estimate the coefficients of the estimation equations (Timæus 1997).

### Table 3

**Ratios of the odds of estimated to actual life table survivorship ($p_{25}$).**

<table>
<thead>
<tr>
<th>Years of survivorship from age 25</th>
<th>Existing coefficients</th>
<th>Revised coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unbiased proportions</td>
<td>Unbiased proportions</td>
</tr>
<tr>
<td>10</td>
<td>1.283</td>
<td>1.002</td>
</tr>
<tr>
<td>15</td>
<td>1.174</td>
<td>1.038</td>
</tr>
<tr>
<td>20</td>
<td>1.037</td>
<td>1.035</td>
</tr>
<tr>
<td>25</td>
<td>1.037</td>
<td>1.039</td>
</tr>
<tr>
<td>30</td>
<td>1.007</td>
<td>1.008</td>
</tr>
</tbody>
</table>

The first column of results in Table 3 shows the biases in estimated life table survivorship produced by applying the existing orphanhood method of estimation to unbiased proportions of respondents with living mothers ($S_x$). The errors are again presented as odds ratios. The estimates were made from the simulated data presented in Table 1 using the regression-based procedure proposed by Timæus (1992). The results would be very similar if another of the existing estimation procedures was used. They are compared with the information on survivorship in Masaka district that was used to generate the simulated data in the first place. Thus, the ratios in the first column indicate the errors that result from the use of regression equations to estimate life table survivorship that do not take into account the unusual age pattern of mortality in populations experiencing AIDS mortality. The life table odds of surviving are overestimated substantially from data on the first two age groups. In contrast, the biases in estimates based on the reports of respondents aged more than 15 years are small.

The reason why the existing coefficients overestimate life table survivorship from 25 to 35 and 40 years is that seroprevalence among women in Masaka district peaks among women in their mid-twenties. A smaller proportion of all women giving birth are infected with HIV than of women aged 25. Therefore, the proportion of mothers who survive ten years is higher than life table survivorship from 25 years to 35 years. On the other hand, mothers as a whole are more likely to become infected in the years after they give birth than are women aged 25 as the younger mothers are still passing through the peak ages of infection. After about 15 years, the close relationship that exists in populations unaffected by HIV between the proportion of mothers alive and life table survivorship from age 25 is re-established. By

---

5 Note that these odds ratios, $R$, define the error in $\alpha$, the level parameter of any relational logit system of model life tables (Brass 1971), irrespective of the choice of a standard mortality schedule:

$$\alpha' - \alpha = -0.5 \ln R$$

6 We use a mean age of childbearing of 26.4 years calculated from the follow-up data on births by age of mother.
the time the respondents are aged 25 years or more, the bias in the existing coefficients can be ignored for all practical purposes.

The more severe the epidemic, the higher the ratio of mothers infected at birth to those infected subsequently and the greater the bias in the existing coefficients for data on 5-9 and 10-14 year old respondents. This is illustrated in Figure 1 for respondents aged 10-14 years and \( t_{15}p_{25} \). The data points are for prevalences of 2.9, 5.9, 11.8, 17.7, 23.6 and 35.3 per cent among women of childbearing age and two levels of background mortality, that prevailing in Masaka district and a higher mortality life table in which \( e_{10} \) is 46.9 not 54.2 years. In essence, AIDS mortality steepens the slope of the line linking the survival of mothers to the survivorship of adult women.

Figure 1
Relationship between the survival of the mothers of respondents aged 10-14 years (\( S_{10} \)) and life table survivorship from 25 to 40 years (\( t_{15}p_{25} \)) for different prevalences of HIV infection, estimates based on data from Masaka district.

![Figure 1](image_url)

Figure 1 reveals that for respondents aged 10-14 years the relationship between survival of mothers and life table survivorship from 25 to 40 years remains close to linear. This is also true of all other age groups. This means that one can estimate revised coefficients for the estimation of survivorship from orphanhood data in populations subject to significant AIDS mortality. These are presented in Table 4. For respondents aged less than 25 years the regression lines have lower intercepts and steeper slopes than in existing estimation procedures. For the 25-29 year age group, however, the regression line is almost identical to that proposed by Timæus (1992).
Table 4
Provisional revised coefficients for estimating life table survivorship from 25 to 25+n years from the proportion of the mothers alive of respondents aged n-5 to n.

<table>
<thead>
<tr>
<th>Age (n)</th>
<th>β₀(n)</th>
<th>β₁(n)</th>
<th>β₂(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>-0.3611</td>
<td>0.00125</td>
<td>1.2974</td>
</tr>
<tr>
<td>15</td>
<td>-0.4030</td>
<td>0.00222</td>
<td>1.3732</td>
</tr>
<tr>
<td>20</td>
<td>-0.2120</td>
<td>0.00372</td>
<td>1.1342</td>
</tr>
<tr>
<td>25</td>
<td>-0.2389</td>
<td>0.00586</td>
<td>1.1131</td>
</tr>
<tr>
<td>30</td>
<td>-0.2513</td>
<td>0.00885</td>
<td>1.0223</td>
</tr>
</tbody>
</table>

\[ q_{25} = β₀(n) + β₁(n)M + β₂(n)S_{n-5} \]

The regression coefficients in Table 4 are provisional. They are based entirely on data from Masaka district. In particular, the age pattern of incidence relative to that of childbearing and the distribution of survival times of the infected differ between populations. Few data exist with which to investigate how predicted survivorship varies with these characteristics of AIDS epidemics and we do not know whether the population in the Masaka district study is typical in these respects. The finding that the biases in the existing coefficients are negligible for older respondents should apply to all populations. The revised coefficients would not be ideal, however, for a population in which prevalence peaked among older women, not those in their twenties. Nevertheless, as even methods that make no allowance for the distinctive age pattern of AIDS mortality yield moderate errors, as demonstrated in the first column of Table 3, we expect the revised coefficients to be robust to variation in those epidemiological parameters of AIDS epidemics that are not investigated here.

Returning to Table 3, the second column of results shows that, when survivorship is estimated using the revised regression coefficients, very accurate results are obtained. Minor errors persist because the points linking survival of mothers and survivorship at a prevalence of 11.9 per cent do not fall exactly on the regression lines. In the third column of results, life table survivorship has been estimated from the adjusted proportions of mothers alive \( S^c \) produced by application of the correction factor given in Equation 7. As one would expect from the results in Table 2, the estimated odds of surviving still differ only slightly from the actual values for Masaka district. The remaining two columns present errors in the estimated odds for populations in which mortality is higher and orphanhood more prevalent than in the baseline model. The errors are again small.

**Orphanhood and adult mortality in Masaka district**

This section of the paper uses the results obtained so far to analyse orphanhood reports collected in the Masaka district study. Such data have been collected only from children. Estimates of adult women’s mortality based on them are compared with direct measures of the mortality of adult women calculated from data obtained prospectively. As the study collected...
information on when parents died, it is possible to calculate not only the proportion of respondents with living mothers at the time of the survey but also the proportions of women who were alive five and ten years earlier. Using these three series of proportions, one can calculate the proportion of mothers alive in synthetic cohorts for the periods 0-4 and 5-9 years before the Round 3 enumeration. These synthetic data reflect the level of orphanhood that would result if a cohort went through childhood experiencing the rates of orphanhood of 1987-92 and 1982-87 respectively.

Table 5
Reported and adjusted proportions of mothers alive and estimated survivorship of women, Masaka district.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reported proportions of mothers alive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-4</td>
<td>0.9946</td>
<td>0.9888</td>
<td>0.9837</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td>0.9755</td>
<td>0.9301</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>0.9230</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adjusted proportions of mothers alive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-4</td>
<td>0.9946</td>
<td>0.9888</td>
<td>0.9837</td>
<td>0.9888</td>
<td>0.9837</td>
</tr>
<tr>
<td>5-9</td>
<td>0.9473</td>
<td>0.9027</td>
<td>0.9417</td>
<td>0.8980</td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td></td>
<td>0.8695</td>
<td></td>
<td>0.8244</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Survivorship of women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i_{0}p_{25}$</td>
<td>0.9009</td>
<td>0.8431</td>
<td>0.8937</td>
<td>0.8270</td>
<td></td>
</tr>
<tr>
<td>$i_{15}p_{25}$</td>
<td>0.8497</td>
<td></td>
<td></td>
<td>0.7876</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 presents the reported data, together with proportions adjusted for selection bias using the correction factor given in Equation 7. The prevalence of orphanhood has been increasing rapidly in Masaka district. A higher proportion of children are orphaned in each successive cohort and more children were orphaned in 1987-92 than 1982-87. Table 5 also presents estimates of life table survivorship. These estimates were derived from data on the age groups 5-9 and 10-14 years using the coefficients presented in Table 4. They reveal that a rapid decrease in the survivorship of adult women has occurred between the two cohorts and two periods on which data are available.
Figure 2
Orphanhood and direct estimates of women’s survivorship from 25 to 35 years, Masaka district.

Figure 2 examines the time trend in the survivorship of women. It compares the orphanhood estimates with those calculated directly from the follow-up data. The time location of those orphanhood estimates based on cohort data was estimated using the standard procedure (Brass and Bamgboye 1981). The orphanhood estimates are somewhat erratic. This is to be expected. It reflects misreporting of orphanhood status, dates of death, and children’s ages and also the approximations involved in adjustment of the proportions and estimation of survivorship. In addition, chaining the cohort proportions to produce proportions for synthetic cohorts tends to magnify any errors in the original data. Despite these potential problems, the estimates document a rapid increase in women’s mortality between 1983 and 1989. They all indicate a level of survivorship intermediate to that in the total population and that in the seronegative population for 1990-95. Extrapolating the trend in the orphanhood estimates backward suggests that early in the 1980s the total population had the level of mortality now found in the seronegative population. Extrapolating forward yields an estimate of survivorship for 1993 that is only slightly higher than that calculated directly. Thus, the orphanhood method works well in this application. To a limited extent this is because the revised regression coefficients were developed using data from this population. On the other hand, the derivation of the procedure proposed for adjusting for selection bias was not tailored to Masaka district.
Discussion

This paper investigates the biases to be expected in orphanhood-based estimates of adult mortality in populations where HIV-related deaths occur in significant numbers. It presents a mathematical model of the relationship between mortality and the survival of mothers in populations affected by the AIDS epidemic. The model is operationalized using demographic data collected during five years of surveillance of a rural Ugandan population. The analysis reveals that the majority of the children of women who die from AIDS are born before their mothers become infected. As only about a quarter of these women’s remaining children are infected vertically, selection biases originating in vertical transmission lead the orphanhood method to overestimate adult survivorship by only a few percentage points. Reductions in the fertility of seropositive women can produce a comparable additional error. Apart from among the youngest respondents, this bias is a function of seroprevalence among women of childbearing age, the vertical transmission rate and the amount by which infected women’s fertility is reduced. A simple way of adjusting for the bias is proposed that can be used whenever an estimate is available of seroprevalence among either pregnant women or all women of childbearing age. Estimates based on antenatal samples are both widely available and ideal for this purpose.

A second source of bias anticipated in the orphanhood-based estimates of adult mortality is that the estimation method is based on models with different age patterns of mortality from those in populations with high AIDS mortality. Modelling based on the data from Masaka district suggests that, despite this, the existing regression equations continue to perform fairly well for data on respondents aged 15 years and over. For younger respondents, in particular, the prevalence of orphanhood increases more slowly as overall mortality rises in populations affected by AIDS than in those that are not. The Masaka district data suggest, however, that the relationship between the proportion of mothers alive and life table survivorship from 25 years remains linear. On this basis, a provisional set of revised estimation coefficients are proposed for use in populations with significant AIDS mortality. The coefficients are provisional because ideally they would be based on data from a wide range of populations with different age patterns of HIV infection. Unfortunately, few such data yet exist for the developing world. While a single set of coefficients should be adequate for use in a wide range of populations, it is unlikely that the AIDS epidemic in Masaka district is in all respects typical. Therefore, a more satisfactory revision of the coefficients will become possible as our understanding of the parameters of AIDS epidemics in developing countries improves.

Estimates of adult women’s mortality made from the data on orphanhood collected for children in Masaka district are very consistent with those calculated directly from the prospective data. Orphanhood data not only indicate the approximate level of mortality in populations affected by AIDS but can document the upward trend in adult mortality over time. Equally, the orphanhood-based results confirm the high quality of the direct information on adult deaths collected by the longitudinal study in Masaka district and the value of the information on the mortality effect of the HIV epidemic that is emerging from this study.

Information on orphanhood has not been gathered from the study population for those aged 15 years and more. However, both the proposed adjustment for selection bias and the new regression coefficients are less robust for the data on 5-9 and 10-14-year-old respondents than for those on older age groups. As the quality of orphanhood data on children is also more suspect than that of those supplied by older individuals, it seems reasonable to infer that the latter data will continue to represent a useful source of estimates of adult mortality. At present, almost no populations exist in which many mothers of adult respondents were infected when they gave birth to the respondent. Selection bias is not yet a problem in...
older age groups and mortality would be overestimated if one adjusted the reported portions of mothers alive downward unnecessarily. Equally though, this means that extensions of the orphanhood method that allow recent mortality to be measured using data supplied by adult respondents (Timæus 1991b,c) should continue to provide representative estimates of the mortality of the middle-aged until at least early next century.

The extent of the biases in estimates of adult male mortality obtained from data on the survival of fathers has not been investigated in this paper. Selection biases linked to vertical transmission will be less severe for data on fathers than those on mothers as not all couples are both infected. However, it is possible that serious bias in the regression coefficients for male mortality may not be restricted to those for respondents aged less than 15 years. Men tend to be older than women both when they become fathers and when they contract HIV infection. They also tend to do both over a wider range of ages. Nevertheless, as the basic structure of the determinants of the prevalence of orphanhood is the same for men and women, it seems likely that data on the survival of fathers will also continue to be of use for the estimation of mortality.

The orphanhood method remains a valuable way of measuring adult mortality even in populations with a relatively high prevalence of HIV infection that are undergoing rapid increases in adult mortality. Analysed appropriately, orphanhood data yield estimates that are less subject to bias in populations affected by HIV than those from most other methods of measuring adult mortality that can be used in national enquiries. Countries that lack effective death registration systems should ask questions about the survival of mothers in the census and national household surveys to monitor the effect of the AIDS epidemic on mortality. Demographers and international advisers should promote the orphanhood method with renewed vigour.

References


