DEMOGRAPHIC RESEARCH

Research Article

More on the cohort-component model of population projection in the context of HIV/AIDS: A Leslie matrix representation and new estimates

Jason R. Thomas

Samuel J. Clark

Table of Contents

More on the cohort-component model of population projection in the context of HIV/AIDS: A Leslie matrix representation and new estimates

Jason R. Thomas[1]

Samuel J. Clark[2]

Abstract

This article presents an extension of the cohort-component model of population projection (CCMPP) first formulated by Heuveline (2003) that is capable of modeling a population affected by HIV. Heuveline proposes a maximum likelihood approach to estimate the age profile of HIV incidence that produced the HIV epidemics in East Africa during the 1990s. We extend this work by developing the Leslie matrix representation of the CCMPP, which greatly facilitates the implementation of the model for parameter estimation and projection. The Leslie matrix also contains information about the stable tendencies of the corresponding population, such as the stable age distribution and time to stability. Another contribution of this work is that we update the sources of data used to estimate the parameters, and use these data to estimate a modified version of the CCMPP that includes (estimated) parameters governing the survival experience of the infected population. A further application of the model to a small population with high HIV prevalence in rural South Africa is presented as an additional demonstration. This work lays the foundation for development of more robust and flexible Bayesian estimation methods that will greatly enhance the utility of this and similar models.

---

[1] Center for Demography of Health and Aging, University of Wisconsin-Madison.
E-mail: jthomas@ssc.wisc.edu.
[2] Department of Sociology, University of Washington. Institute of Behavioral Science (IBS), University of Colorado at Boulder. MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand. E-mail: work@samclark.net.

1.   Introduction

The cohort-component model of population projection (CCMPP) is perhaps *the* iconic method in demography, see for example Bowley (1924); Cannan (1895); Whelpton (1936); Leslie (1945); Pritchett (1891); Pearl and Reed (1920); Dorn (1950). This classic method moves incrementally forward in time a population defined by age according to a specified life table and set of age-specific fertility rates, taking into account net migration at each age. In its basic form it is straightforward and easy to implement, which has allowed it to become one of the essential tools used by governments and planning organizations to help them understand the likely future size and composition of a population, and how that may change under different assumptions or as a result of interventions of various types.

Fundamentally the CCMPP relates the age structure of a population to fertility, mortality, and migration, with the current age structure being the result of fertility, mortality and migration at each age in the past. Most commonly a future age structure of the population is 'predicted' given a time series of age-specific fertility, mortality, and net migration. The model can also be used to *estimate* trends in a subset of the four components given the others, and it is a use of this type that occupies us here.

The HIV epidemic affecting Africa and other parts of the developing world poses significant challenges to demographers concerned with either measuring the current state of an affected population or predicting its future. Because HIV affects both fertility and mortality in important ways (Sewankambo et al. 1994; Nunn et al. 1997; Carpenter et al. 1997; Gray et al. 1998; Żaba and Gregson 1998; Todd et al. 1997; Wachter, Knodel, and Van Landingham 2002; Hunter et al. 2003; Terceira et al. 2003; Lewis et al. 2004; Timaeus and Jasseh 2004; Ford and Hosegood 2005; Garenne et al. 2007; Gregson et al. 2007; Kahn et al. 2007; Nyirenda et al. 2007; Żaba et al. 2007; Clark et al. 2008), it is not possible to understand the dynamics of a population affected by HIV without specifically taking into account these effects. Further complicating this situation, data describing underlying fertility and mortality unaffected by HIV are scarce and often of poor quality, especially for most populations with high HIV prevalence.

These challenges are addressed in an interesting and useful way in a model developed by Heuveline (2003). Heuveline created a multi-state version of the standard CCMPP model (Day 1996; United Nations 2004) that further classifies the population by time since infection with HIV, and uses a set of age-specific incidence parameters to 'infect' HIV negative people and transition them to the first (shortest duration) HIV positive group. This version of the model also includes additional parameters to govern the links between HIV status and fertility. Heuveline used data describing HIV status and survival (mortality) from East African countries to estimate these model parameters, using maximum likelihood techniques.

The purpose of this article is to build on the work of Heuveline (2003) and the multi-

state, HIV-enabled CCMPP in the following ways. Since the original publication of this model in 2003, more data have become available on HIV prevalence and the survival of those infected. Perhaps the most notable source of more recent data is the Demographic and Health Surveys (DHS) program (and the associated AIDS Indicator Surveys – AIS) which have collected information on age-specific HIV prevalence from nationally-representative population-based surveys. We augment the data compiled by Heuveline (2003), taken from sources in East Africa, with over 100,000 observations on HIV prevalence provided by the DHS data in the same region.[3] We estimate the model parameters using the augmented data and compare these new estimates to those obtained by using only the data compiled by Heuveline, as well as to the estimates obtained by using only the more recent data compiled by us. Adding more recent data to the analysis will provide more information about the later stages of the HIV epidemic, and the comparisons across the data compilations will show how sensitive the estimates are to the data being used (particularly with respect to the date when the data were collected).

Relative to the original data used, there is more recent information on the survival of those infected with HIV (e.g. Żaba et al. 2007; Crampin et al. 2002; Urassa et al. 2001; Sewankambo et al. 2000) which motivates an extension of the CCMPP model. Previous work with the HIV-enabled CCMPP uses a fixed survival schedule for the infected population to estimate the model parameters (see the next section for more details). While it is possible to specify (a priori) a reasonable survival schedule for HIV-positive individuals, it is preferable to let data inform the model, so that the estimates and projections are not influenced by problems with the assumed mortality experience. In this paper we extend the CCMPP to include additional parameters that govern the survival of the infected population, and estimate these parameters using new data gathered from the literature. It is worth noting that the estimated age pattern of HIV incidence may be sensitive to the life table used for the infected population, because the incidence parameters are estimated using data on HIV prevalence, which is a function of both incidence and survival. In other words, an observed level of HIV prevalence can arise from different combinations of incidence and survival rates. Given this point, we discuss differences in the estimated age patterns of HIV incidence between the original specification of the model and the version in which survival for the infected population is estimated. A final point that motivates our extension of the model is that the expansion of antiretroviral treatment programs will affect the survival prospects of the infected population. The modeling techniques carried out here are easily adaptable to a population with significant access to antiretroviral treatment, which makes CCMPP an additional tool to study the future effects of these treatment programs.

---

[3] The list of countries from which DHS data are used consists of Ethiopia, Kenya, Malawi, Rwanda, Tanzania, Uganda, Zambia, and Zimbabwe.

The final contribution of our work is to shift the geographic focus from eastern Africa to South Africa, where the HIV epidemic has reached even higher levels (UNAIDS 2008). Data on age-specific HIV prevalence from a population living in a rural area of the KwaZulu-Natal province of South Africa (Welz et al. 2007) is used to estimate the CCMPP parameters, and these estimates are compared to those obtained from the data collected in eastern Africa. Different parameterizations of the model are also explored along with the long term implications for the age structure.

The subsequent sections of this paper are organized as follows. A detailed introduction of Heuveline's multi-state, HIV-enabled CCMPP is provided in the next section, with the Leslie matrix representation of the model. This is followed by a description of the data and a brief discussion of the maximum likelihood estimation used in the analysis. The focus then shifts to the results obtained from using the new data and the modeling of the survival of the infected population. Finally, we apply the CCMPP to a rural population in South Africa and conclude with a discussion section.

2. CCMPP

Heuveline (2003) extends the standard CCMPP to accommodate a population categorized by duration of infection with HIV using five 'HIV duration' groups. There are four HIV+ duration groups (0-4 years, 5-9 years, 10-14 years, and 15+ years) as well as an HIV– group. In this section we present Heuveline's multi-state CCMPP for a population with 17 age groups (0-4, 5-9, ..., 80+) in each of the five HIV duration groups. The model is introduced with a series of equations representing the transition from one group/time period to the next. While the model can be applied to both men and women, the description presented here only includes the details for women.

Begin by dividing the population into age groups where $a = 1, 2, \ldots, 17$ correspond to age groups 0-4, 5-9, ..., 80+. Denote membership in the HIV duration groups by d, with $d = 1, 2, \ldots, 5$ corresponding to HIV–, HIV+ for 0-4 years, ..., HIV+ for more than 15 years. Time is indexed by t noting that the duration between t and $t + 1$ is equal to the width of a standard age interval, i.e. 5 years. Let $n_{a,d,t}$ be the number of women in age group a and duration group d at time t. For $1 < a < 17$, we have:

$$n_{a+1,1,t+1} = n_{a,1,t}\, s_{a,1,t}\, (1 - i_{a,t}) \tag{1}$$

$$n_{a+1,2,t+1} = n_{a,1,t}\, s_{a,1,t}\, i_{a,t}\, s_{a,2} \tag{2}$$

$$n_{a+1,d,t+1} = n_{a,d-1,t}\, s_{a,1,t}\, s_{a,d} \quad \text{for } d > 2 \tag{3}$$

where $s_{a,d}$ is the survivorship ratio for age group a and duration group d. Note that for $2 < d < 5$ this survivorship ratio determines the transition from one age group to the next, as well as from one duration group to the next. Each HIV+ group is exposed

to the same underlying base survivorship ratio as the HIV– group, in addition to this extra survivorship ratio that accounts for the increased mortality associated with different durations of infection. The parameter $i_{a,t}$ is the fraction of women in age group a who become infected with HIV over the projection interval. To allow for the heterogeneity of HIV epidemics across populations, this parameter is decomposed as:

$$i_{a,k,t} = 1 - \exp\{-\Gamma_{t,t_0} \cdot H \cdot j_{a,k}\} \tag{4}$$

where k denotes sex and Γ_{t,t_0} is a parametric curve used to model the time trend in the HIV epidemic from the start time t_0. The actual values for Γ_{t,t_0} are presented in Table 1 (see the next section for more details). The parameter H is a population-specific scale parameter that captures the overall magnitude of the epidemic. The parameter $j_{a,k}$ is an age- and sex-specific scaling factor for incidence that represents the multiplicative difference in HIV incidence between age group a and a reference age group, which is held constant at a value of 1.0 in order to make the model identifiable. Following Heuveline, we set the reference age group females aged 25-29, i.e. $j_{5,female} = 1.0$.

For a given age profile of incidence (a specific set of values for j_a), Figure 1 demonstrates how the different values for H simply scale the incidence profile. Each panel in this figure corresponds to a different time in the epidemic, with incidence, whose overall scale is determined by the values of Γ. Within each panel, each line corresponds to a value of the population-specific scale parameter H ranging from 0.1 to 1.0.

The projection equations are slightly different for the youngest and oldest age groups. The oldest (open-ended) age group is incremented by two sources, those 75-79 and 80+ in the previous time period. Thus for $a = 17$ we have:

$$
\begin{aligned}
n_{17,1,t+1} &= n_{16,1,t}\, s_{16,1,t}\, (1 - i_{16,t}) \\
&\quad + n_{17,1,t}\, s_{17,1,t}\, (1 - i_{17,t}) \tag{5} \\
n_{17,2,t+1} &= n_{16,1,t}\, s_{16,1,t}\, i_{16,t}\, s_{16,2} \\
&\quad + n_{17,1,t}\, s_{17,1,t}\, i_{17,t}\, s_{17,2} \tag{6} \\
n_{17,d,t+1} &= n_{16,d-1,t}\, s_{16,1,t}\, s_{16,d} \\
&\quad + n_{17,d-1,t}\, s_{17,1,t}\, s_{17,d} \quad \text{for } 2 < d < 5 \tag{7} \\
n_{17,5,t+1} &= n_{16,4,t}\, s_{16,1,t}\, s_{16,5} + n_{17,4,t}\, s_{17,1,t}\, s_{17,5} \\
&\quad + n_{16,5,t}\, s_{16,1,t}\, s_{16,5} + n_{17,5,t}\, s_{17,1,t}\, s_{17,5}. \tag{8}
\end{aligned}
$$

Figure 1: **Age-specific HIV incidence rates, $i_{a,t}$, for different values of the population-specific scale parameter, H, over time**
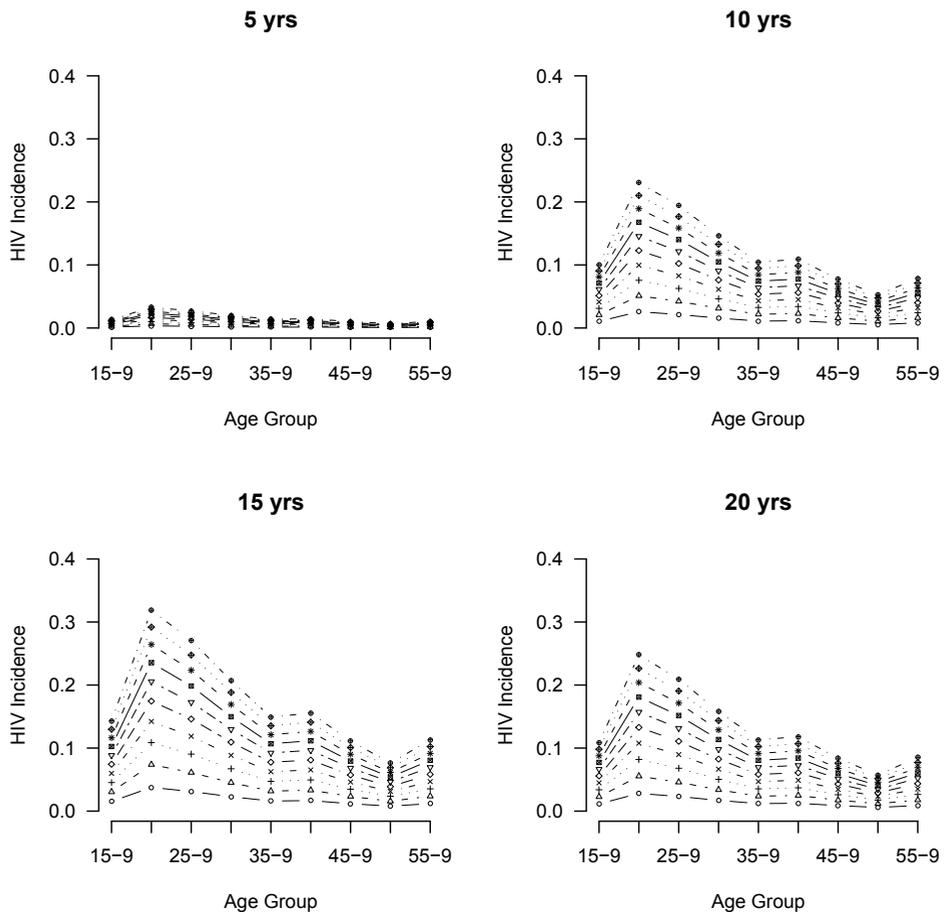
As with the single-state CCMPP, the number of children in the first age group at the end of the projection interval is determined by surviving forward the births that occur during the projection interval. The number of births that occur is calculated by applying age-specific fertility rates to the average number of women in each age group during the

projection interval, taking into account the fact that HIV+ women who have been infected for different durations will, to varying degrees, be less likely to have children. To capture the relationship between fertility and HIV status, Heuveline defined three additional parameters. First, consider the number of HIV– births:

$$
\begin{aligned}
n_{1,1,t+1} \;=\; s_{0,1,t}\,\frac{1}{1+SRB}\times\Big(\,&\sum_{a=\alpha}^{\beta} f_{a,1,t}\,\frac{n_{a,1,t}+p^{-}_{a-1,1,t}n_{a-1,1,t}}{2}\\
+\;&\sum_{d=2}^{5}\sum_{a=\alpha}^{\beta} f^{-}_{a,d,t}\,\frac{n_{a,d,t}+p_{a-1,d-1,t}n_{a-1,d-1,t}}{2}\Big).
\end{aligned}\tag{9}
$$

In Equation 9 above, SRB is the sex ratio at birth, the $f_{a,1,t}$'s are simply the age-specific fertility rates for HIV– women, and the lower and upper bounds of the childbearing age range are α and β. Fertility among HIV+ women introduces the following parameters

$$
f^{-}_{a,d,t} \;=\; f_{a,1,t}\,e_a\,g_d\,(1-v)\tag{10}
$$

for $1 < d$. The superscript in $f^{-}_{a,d,t}$ designates HIV– births (i.e. $d=1$) to women who are HIV+. The parameter v_d is the probability that an HIV+ woman in duration group d will give birth to an HIV+ child, the *vertical transmission* rate. The parameter e_a captures the higher level of sexual activity and resulting fertility among HIV+ women age 15-19 who have been infected for 0-4 years ($d=2$). In other words we expect $e_{a=4} > 1$ while $e_{a\neq4}$ are constrained to be 1.0. The parameter g_d represents the *fertility impairment* experienced by women in duration group d, a number that becomes smaller as the time since infection increases, reflecting increasing fertility impairment with time since infection. The corresponding equations for HIV+ births are:

$$
n_{1,2,t+1} \;=\; s_{0,1,t}\,\frac{1}{1+SRB}\sum_{d=2}^{5}\sum_{a=\alpha}^{\beta} f^{+}_{a,d,t}\,\frac{n_{a,d,t}+p_{a-1,d-1,t}\,n_{a-1,d-1,t}}{2}\tag{11}
$$

$$
f^{+}_{a,d,t} \;=\; f_{a,1,t}\,e_a\,g_d\,v.\tag{12}
$$

Finally, we define the factors used to approximate the average number of women at the beginning and end of the period, $p^{-}_{a,1,t}$ and $p_{a,d,t}$:

$$
p^{-}_{a,1,t} \;=\; s_{a,1,t}\,(1-i_{a,t})\tag{13}
$$

$$
p_{a,1,t} \;=\; s_{a,1,t}\,i_{a,t}\,s_{a,2}\tag{14}
$$

$$
p_{a,d,t} \;=\; s_{a,1,t}\,s_{a,d}\quad\text{for } d>1.\tag{15}
$$

2.1 The HIV incidence trend

Recall that age-specific incidence in CCMPP is modeled as follows:

$$i_{a,k,t} = 1 - \exp\{-\Gamma_{t,t_0} \, H \, j_{a,k}\}$$

where $i_{a,k,t}$ is the fraction of individuals age a who will become infected over the projection interval. Γ_{t,t_0} represents the shape of the incidence trend from the start of the epidemic in year t_0 to the projection period t. This incidence trend is shifted up or down by H, an overall scale parameter for the epidemic. Finally, $j_{a,k}$ is the incidence ratio comparing those of age a and sex k to women age 25-29. The details of the incidence trend Γ_{t,t_0} are described in this section, along with several other possible specifications.

The incidence trend Γ_{t,t_0} used by Heuveline (2003) is borrowed from `EpiModel`, a computer program developed by the World Health Organization to make short-term projections of adult AIDS cases (Chin and Lwanga 1991) (the precursor of EPP, UNAIDS' estimation and projection package software used to estimate the prevalence of HIV (Ghys et al. 2004)), and is based on the gamma family of distributions:

$$g(t) = \frac{t^{\alpha-1}e^{-x/\beta}}{(\alpha-1)!\beta^\alpha}, \quad \text{for } t \geq 0, \ \alpha > 0 \ \beta > 0. \tag{16}$$

The α parameter is typically referred to as the shape parameter since it affects how peaked or flat the density is; as α increases the density appears more flat or uniform. The scale parameter β is associated with how diffuse or spread out the density is; as β increases the density spreads out.

Before discussing the calculations made by Heuveline (2003) it is helpful to discuss a particular property of the gamma distribution. The mode, or the the value for which the the function reaches its maximum, is equal to $(\alpha - 1)\beta$. This quantity has a nice interpretation in that it is the number of years after the start of the epidemic t_0 when the epidemic peaks. For example, if the shape parameters is 5 and the scale parameter is 3, then the epidemic peaks 12 years after it began.[4]

---

[4] Chin and Lwanga (1991) report that setting $\alpha = 5$ results in "the best empirical 'fit' to the reported AIDS-case curves in countries with reliable case-reporting systems." They set $\beta = 1$.

This is precisely the density used by Heuveline (2003) to calculate the five-year incidence rates (i.e. $\alpha = 5, \beta = 3$). The actual rates are calculated by integrating the gamma density over the appropriate five-year span, i.e.:

$$\Gamma_{x,t_0} = \int_{(x-5)}^{x} \frac{t^{5-1}e^{-x/3}}{(5-1)!3^5} dt \quad \text{for } 5 \geq x \leq 20, \tag{17}$$

$$\Gamma_{20+,t_0} = 5 \times \int_{20}^{21} \frac{t^{5-1}e^{-x/3}}{(5-1)!3^5} dt. \tag{18}$$

The five-year incidence rate for twenty years after the start date is different because the decline in the gamma density for values greater than 20 is too rapid to represent an actual decline in incidence. The values used to estimate the CCMPP parameters are presented in Table 1. We are aware that this is relatively crude, and it would be preferable to model this trend using time series techniques with attention given to the uncertainty around the trend, especially at time points far into the future. That is among the improvements that currently occupy us.

Table 1: **Five-year incidence rates calculated from the gamma density and an exponential curve**

Time Period	Γ_{t,t_0}	E_{t,t_0}
0 - 5 years	0.028	0.063
6 - 10 years	0.216	0.191
11 - 15 years	0.316	0.323
16 - 20 years	0.235	0.457
20+ years[a]	0.163	0.540

Notes: [a] For the gamma model it is assumed that the HIV incidence rate will level off at the rate equal to the integral of the gamma density from 20 and 21 multiplied by five. For the exponential model it is assumed that the HIV incidence rate will level off at a rate equal to $5 * (h(t = 21) - h(t = 20))$. See the text for the definition of $h(t)$.

We will now discuss two other possible specifications for the trend in HIV incidence. The first is an exponential curve that models a continual increase in HIV over time. While this may not be realistic in the long run, it does provide an upper bound for the trend. A reasonable lower bound is a constant rate of new infections (i.e. no change) over time. Since the second specification is simply a constant[5] we will focus our attention on the exponential model.

---

[5] Heuveline (2003) uses 0.2 as the five-year incidence rate.

The exponential curve used to model the trend in HIV incidence takes the following functional form:

$$h(t) = \frac{e^{\beta t}}{\beta} - t \quad \text{for } t = 1, 2, 3, \ldots; \text{ and } \beta > 0. \tag{19}$$

The five-year HIV incidence rates are calculated by differencing $h(t)$ (at lag one) and summing over the five year period of interest:

$$E_{t, t_0} = \sum_{j=1}^{5} h(t) - h(t-1) \tag{20}$$

Heuveline (2003) chooses a value of $\beta = 0.005$ to obtain the five-year HIV incidence rates based on the exponential model, shown in Table 1. Finally, similar to the gamma model described above, the incidence rate after twenty years takes a different form. For the exponential curve, $E_{20+, t_0} = 5 \times (h(t=21) - h(t=20))$.

2.2 Additional HIV-related force of mortality

In the HIV-enabled CCMPP individuals infected with HIV ($d > 1$) experience an additional force of mortality that is not experienced by those in the HIV– state ($d = 1$). This mortality differential can be seen in the following projection equations

$$n_{a+1,d=2,t+1} = n_{a,d=1,t} \, s_{a,d=1,t} \, i_{a,t} \, s_{a,d=2} \tag{21}$$

$$n_{a+1,d>2,t+1} = n_{a,d-1,t} \, s_{a,d=1,t} \, s_{a,d>2} \tag{22}$$

where $s_{a,d>1} < 1$; survival in the HIV+ states is reduced compared to the HIV– state. Recall that the vital rates in CCMPP are treated as fixed parameters and need to be set by the user.

Heuveline turns to the epidemiological literature for guidance on choosing values for the survival rates of individuals infected with HIV. One of the more important findings (concerning the HIV-enabled CCMPP) is that the progression from HIV to death is faster for those infected at older ages. Morgan et al. (2002) report this finding in a cohort study in rural Uganda for whom the time of HIV infection is reasonably well known, and the median follow-up time is 5.6 years (interquartile range: 3.6-7.3 years). Among those aged 15-24, 25-39, and 40+ years at seroconversion, the probabilities of surviving seven years (after seroconversion) are 79% (95%CI: 63-88%), 72% (95%CI: 56-83%), and 20% (95%CI: 6-40%), respectively. These estimates are based on 10, 18, and 19 deaths among 65, 68, and 35 participants, going from the youngest to the oldest age group. Morgan et al. (2002) also report a faster progression from seroconversion to AIDS for the oldest age

group (40+ years).[6] Similar results are reported in an updated analysis of these Ugandan data by Van der Paal et al. (2007), and in other cohort studies carried out in sub-Saharan Africa (Lutalo et al. 2007; Isingo et al. 2007; Murray et al. 2007).

Given the evidence from the epidemiological literature, Heuveline (2003) specifies the survival rates for HIV+ individuals as a function of age at infection. This dependence comes through in the choice of a particular survival schedule, defined by the median number of years lived after infection. The schedules include median survival times of 3, 8, 10, and 12 years, which is consistent with the empirical evidence from sub-Saharan Africa (Newell et al. 2004; Marston et al. 2005; Morgan et al. 2002; Van der Paal et al. 2007; Lutalo et al. 2007; Peters et al. 2007; Isingo et al. 2007). Children who are infected perinatally follow the 3 year schedule, while the oldest age groups follow the 8-year schedule. Before describing the age dependence in greater detail, it is helpful to take a slight digression and define some more notation.

Heuveline (2003) defines these survival schedules with reference to the projection interval (i.e. five years). Let $y_{d,m}$ be the expected number of years lived by an individual in duration group d following survival schedule m, where $m = 3, 8, 11, 12$. For example, the average number of years lived by a person infected 5-9 years ago who is following the survival schedule with a median survival time of 11 years is $y_{d=3,m=11} = 3.375$. The values for $y_{d,m}$ are listed in Table 2.

Table 2: **Expected number of person-years lived over a five-year interval by survival schedule and duration group**

Duration Group		Survival Schedule			
		3	8	11	12
0 to 0-4	(d=2)	2.7750	4.7100	4.8000	4.8310
0-4 to 5-9	(d=3)	0.4250	2.4300	3.3750	3.6000
5-9 to 10-4	(d=4)	0.0000	0.8600	2.0000	2.4125
10-4 to 15-9	(d=5)	0.0000	0.3150	1.0000	1.5375

Now we are in a position to define $s_{a,d>1}$. Let us begin with those who have been infected for 0-4 years ($d = 2$). Children born and infected (perinatally) during the projection interval are exposed to:

$$s_{a=1,d=2} = \frac{y_{d=2,m=3}}{5}.$$ (23)

---

[6] In this same study, the median time from seroconversion to AIDS is 9.4 years, IQR: 5.5 – 10.1 years. The median time from seroconversion to death is 9.8 years, IQR: 6.1 – > 10.3 years (Morgan et al. 2002).

The HIV-related survival ratios for the next two age groups are defined to be 1.0 because persons between the of ages 5 and 9 are not able to be infected given our current assumptions about incidence. Recall that the age-specific incidence rates are zero for the first three age groups.[7] For those who are ages 15-19, 25-34, or above age 45, the additional force of mortality caused by HIV takes a form similar to the equation just above:

$$s_{a=4,d=2} = \frac{y_{d=2,m=12}}{5} = 0.9662 \tag{24}$$

$$s_{8 \geq a \geq 6, d=2} = \frac{y_{d=2,m=11}}{5} = 0.9600 \tag{25}$$

$$s_{a \geq 10, d=2} = \frac{y_{d=2,m=8}}{5} = 0.9420. \tag{26}$$

Note how survival declines as age increases before leveling off at age 45 and above. For the age groups not already mentioned, the survival parameters are calculated by taking the average over two adjacent survival schedules:

$$s_{a=5,d=2} = \frac{y_{d=2,m=11}+y_{d=2,m=12}}{2} = 0.9631 \tag{27}$$

$$s_{a=9,d=2} = \frac{y_{d=2,m=8}+y_{d=2,m=11}}{2} = 0.9510. \tag{28}$$

We now turn our attention to the third duration group, individuals who have been infected for 5-9 years. For this group, we start with those ages 5-9:

$$s_{a=2,d=3} = \frac{y_{d=3,m=3}}{y_{d=2,m=3}}. \tag{29}$$

The corresponding parameters for the older age groups take a similar form. The expected number of years lived for the third duration group is divided by the expected number of years lived by the second duration group (for a given survival schedule). The dependence on age for $d=3$ takes the same form as for the previous duration group, only that the age groups are incremented by one. The actual equations are:

$$s_{a=5,d=3} = \frac{y_{d=3,m=3}}{y_{d=2,m=3}} \tag{30}$$

$$s_{a=6,d=3} = \frac{y_{d=3,m=12} + y_{d=3,m=11}}{y_{d=2,m=12} + y_{d=2,m=11}} \tag{31}$$

$$s_{9 \geq a \geq 7, d=3} = \frac{y_{d=3,m=11}}{y_{d=2,m=11}} \tag{32}$$

$$s_{a=10,d=3} = \frac{y_{d=3,m=11} + y_{d=3,m=8}}{y_{d=2,m=11} + y_{d=2,m=8}} \tag{33}$$

$$s_{a \geq 11, d=3} = \frac{y_{d=3,m=8}}{y_{d=2,m=8}}. \tag{34}$$

---

[7] This might not be the best assumption since infants are still able to become infected via breastfeeding.

The pattern continues for the fourth and fifth duration groups. All of the parameters for the additional force of mortality due to HIV are listed in Table 3.

Table 3: **Survival probabilities applied to HIV+ ($s_{a,d>1}$)**

| Age Group | HIV Duration Group | | | |
	0-4 yrs (d=2)	5-9 yrs (d=3)	10-4 yrs (d=4)	15+ yrs (d=5)
0-4	0.5550	–	–	–
5-9	–	0.1532	–	–
10-4	–	–	0.0000	–
15-9	0.9662	–	–	0.0000
20-4	0.9631	0.7452	–	–
25-9	0.9600	0.7242	0.6701	–
30-4	0.9600	0.7031	0.6326	0.6373
35-9	0.9600	0.7031	0.5926	0.5751
40-4	0.9510	0.7031	0.5926	0.5000
45-9	0.9420	0.6104	0.5926	0.5000
50-4	0.9420	0.5159	0.4927	0.5000
55-9	0.9420	0.5159	0.3539	0.4598
60-4	0.9420	0.5159	0.3539	0.3663
65-9	0.9420	0.5159	0.3539	0.3663
70-5	0.9420	0.5159	0.3539	0.3663
75-9	0.9420	0.5159	0.3539	0.3663
80+	0.9420	0.5159	0.3539	0.3663

Life tables can also be constructed using the survival rates presented in Table 3, where cohorts defined by age at infection are exposed to the survival rates aligned along the diagonal cells of the table. For example children infected by their mothers will never reach age 15 years because the survival probability is zero for those infected at birth and between the ages of 10 and 14. The mortality experienced by the cohort of women infected at age 15 is summarized in the life table presented in Table 4.

The size of the cohort exposed to the mortality risks in Table 4 is reduced to half after fifteen years. After thirty years there is just under ten percent of the cohort still living. While the survival experience of this cohort may seem plausible, it is preferable to have the their survival rates informed by data. The current specification for the survival experience results in a stable population with unappealing characteristics. This point is discussed later on in the paper.

Table 4: **Life table for HIV+ women infected at age 15**

Age group	$_np_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	T_x	$\overset{o}{e}$
15	0.9566	0.0434	100,000	4,345	489,138	1,612,020	16.1202
20	0.7347	0.2653	95,655	25,381	414,822	1,122,882	11.7389
25	0.6587	0.3413	70,274	23,983	291,412	708,060	10.0757
30	0.6255	0.3745	46,291	17,338	188,110	416,648	9.0006
35	0.5634	0.4366	28,953	12,642	113,160	228,538	7.8934
40	0.4883	0.5117	16,311	8,346	60,690	115,378	7.0736
45	0.4863	0.5137	7,965	4,092	29,595	546,880	6.8660
50	0.4823	0.5177	3,873	2,005	14,352	25,093	6.4789
55	0.4370	0.5630	1,868	1,052	6,710	10,741	5.7498
60	0.3388	0.6612	816	540	2,730	4,031	4.9395
65	0.3220	0.6780	276	187	912	1,301	4.7125
70	0.2940	0.7060	89	63	288	389	4.3669
75	0.2500	0.7500	26	20	80	101	3.8713
80	0.1586	0.8414	6	6	21	21	3.4425

2.3 Matrix notation for HIV-enabled CCMPP

These equations for the multi-state, HIV-enabled CCMPP can be conveniently expressed in matrix notation. For a population with 17 age groups and five HIV duration groups, the population at time t is represented by an 85×1 column vector

$$\mathbf{n}_t = \begin{bmatrix} n_{1,1,t} \\ n_{2,1,t} \\ \vdots \\ n_{17,1,t} \\ \hline \vdots \\ \hline n_{1,4,t} \\ n_{2,4,t} \\ \vdots \\ n_{17,4,t} \end{bmatrix}. \tag{35}$$

The corresponding *Leslie* matrix is:

$$
\mathbf{A}_t = \begin{bmatrix}
\mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \mathbf{B}_{1,3} & \mathbf{B}_{1,4} & \mathbf{B}_{1,5} \\
\mathbf{B}_{2,1} & \mathbf{B}_{2,2} & \mathbf{B}_{2,3} & \mathbf{B}_{2,4} & \mathbf{B}_{2,5} \\
\mathbf{0} & \mathbf{B}_{3,2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{B}_{4,3} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{5,4} & \mathbf{B}_{5,5}
\end{bmatrix}
\tag{36}
$$

where $\mathbf{B}_{i,j}$ is a 17×17 sub-matrix that models how group j at time t contributes to group i at time $t+1$. Note that $\mathbf{B}_{3,1}$ is a zero matrix since women who are HIV– at time t cannot give birth to children who have been HIV positive for ten years by $t+1$ (i.e. five years into the future). Similar reasoning applies for the other zero matrices.

The calculations involving $\mathbf{B}_{1,j}$ produce the projection for the number of HIV– births (i.e. $n_{1,1,t+1}$) contributed by duration group j. Similarly, $\mathbf{B}_{2,j}$ projects the number of HIV positive births contributed by duration group $j > 1$. $\mathbf{B}_{1,1}$ and $\mathbf{B}_{2,1}$ are a little different in that they project each age group to the next oldest age group *and* from one HIV duration group to the next. Let us first consider $\mathbf{B}_{1,1}$:

$$
\mathbf{B}_{1,1} = \begin{bmatrix}
b^-_{1,1,t} & b^-_{2,1,t} & \cdots & & & b^-_{17,1,t} \\
p^-_{1,1,t} & 0 & \cdots & & & 0 \\
0 & p^-_{2,1,t} & \ddots & & & \vdots \\
0 & 0 & \ddots & & & 0 \\
\vdots & \vdots & \ddots & \ddots & 0 & 0 \\
0 & 0 & \cdots & 0 & p^-_{16,1,t} & p^-_{17,1,t}
\end{bmatrix}.
\tag{37}
$$

Recall that the number in the first age group at time $t+1$ is equal to the number of births summed across the fecund age groups. Let $b^-_{a,d,t}$ be the factor needed to calculate the number of HIV– births to mothers in age group a at time t and in duration group d:

$$
b^-_{a,1,t} = s_{0,1,t} \, \frac{1}{1+SRB} \, f^-_{a,1,t} \, \frac{1 + p^-_{a-1,1,t} \frac{n_{a-1,1,t}}{n_{a,1,t}}}{2}.
\tag{38}
$$

In our application of the multi-state, HIV-enabled CCMPP, fertility only occurs among women aged 15-49 (i.e. $\alpha = 4$, $\beta = 10$). Consequently $b^-_{a<4,1,t} = b^-_{a>10,1,t} = 0$. In the equation above the factor $\frac{n_{a-1,1,t}}{n_{a,1,t}}$ is used to approximate the number of women at risk

of giving birth. If the count in the denominator $n_{a,1,t}$ is ever zero the entire ratio is simply replaced by zero. This issue arises when dealing with the fertility of the HIV+ groups. The same procedure is used in the analogous HIV+ equations if they involve dividing by zero.

$\mathbf{B}_{1,d}$ for $d > 1$ projects forward HIV– births contributed by duration group d and can be written as:

$$
\mathbf{B}_{1,d} =
\begin{bmatrix}
b^-_{1,d,t} & b^-_{2,d,t} & b^-_{3,d,t} & \cdots & b^-_{17,d,t} \\
0 & \cdots & & & 0 \\
\vdots & \ddots & & & \\
& & & & \\
0 & & & & 0
\end{bmatrix}
\tag{39}
$$

where

$$
b^-_{a,d,t} = s_{0,1,t}\, \frac{1}{1 + SRB}\, f^-_{a,d,t}\, \frac{1 + p_{a-1,d-1,t}\left(\frac{n_{a-1,d-1,t}}{n_{a,d,t}}\right)}{2}
\tag{40}
$$

for $d > 1$. The $\mathbf{B}_{2,d}$'s determine the number of people infected with HIV for less than five years at time $t + 1$, contributed by those in duration group d at time t. For the first duration group we have:

$$
\mathbf{B}_{2,1} =
\begin{bmatrix}
b^+_{1,1,t} & b^+_{2,1,t} & \cdots & & & & b^+_{17,1,t} \\
p_{1,1,t} & 0 & \cdots & & & & 0 \\
0 & p_{2,1,t} & \ddots & & & & \vdots \\
0 & 0 & \ddots & & & & 0 \\
\vdots & \vdots & \ddots & \ddots & 0 & 0 \\
0 & 0 & \cdots & & 0 & p_{16,1,t} & p_{17,1,t}
\end{bmatrix}.
\tag{41}
$$

$\mathbf{B}_{2,d}$ for $d > 1$ projects forward the number of HIV+ births contributed by duration group d. It can be written as:

$$\mathbf{B}_{2,d} = \begin{bmatrix} b^+_{1,d,t} & b^+_{2,d,t} & b^+_{3,d,t} & \cdots & b^+_{17,d,t} \\ 0 & \cdots & & & 0 \\ \vdots & & \ddots & & \\ 0 & & & & 0 \end{bmatrix} \tag{42}$$

where

$$b^+_{a,d,t} = s_{0,1,t}\, s_{0,1,t}\, \frac{1}{1+SRB}\, f^+_{a,d,t}\, \frac{1 + p_{a-1,d-1,t}\left(\frac{n_{a-1,d-1,t}}{n_{a,d,t}}\right)}{2}. \tag{43}$$

The remaining non-zero sub-matrices $-\mathbf{B}_{3,2}, \mathbf{B}_{4,3}, \mathbf{B}_{5,4}$ and $\mathbf{B}_{5,5}$ $-$ project people forward in both age and time, and consequently the only non-zero elements occur along the sub-diagonal:

$$\mathbf{B}_{i,j} = \begin{bmatrix} 0 & 0 & \cdots & & & 0 \\ p_{1,d=j,t} & 0 & \cdots & & & \vdots \\ 0 & p_{2,d=j,t} & \ddots & & & \\ 0 & 0 & \ddots & & & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & p_{16,d=j,t} & p_{17,d=j,t} \end{bmatrix}. \tag{44}$$

The Leslie matrix representation of the model greatly facilitates the implementation and use of the multi-state, HIV-enabled CCMPP using the R programming language (R Development Core Team 2011). Repeated matrix multiplication produces the projected population at five-year intervals, making it possible to explore the long term behavior of the population and the epidemic. See the results section for an application to a population living in the KwaZulu-Natal province of South Africa. Finally, if the Leslie matrix is irreducible and primitive, then we can explore the stable age distribution of the population (Keyfitz and Caswell 2005).

3. Parameter estimation

The CCMPP projections are used to estimate thirty-three of the model parameters. As mentioned earlier, the vital rates, the initial population counts and the HIV survival schedules are all fixed. The 33 parameters we estimate are:

- v: vertical transmission parameter that is constrained to be between 0 and 1; although the model is described as having a vertical transmission rate for each duration group, there are not enough data to estimate separate parameters. (1 parameter)
- e_a: fertility selection parameter that is constrained to be equal to 1 for all groups except women aged 15-19 in the first HIV duration group, for whom we expect the value to be greater than 1. (1 parameter)
- g_d: fertility impairment parameter for women in duration group d, for $d = 2, 3, 4$; the fertility impairment parameter for $d = 5$ is constrained such that $g_{d=5} = g_{d=4}$, and the values for all duration groups are constrained to be between 0 and 1. (3 parameters)
- $j_{a,k}$: relative incidence ratio parameter that is constrained to be equal to 1 for women age 25-29 and non-negative for all other groups; values are estimated for women ($k = 1$) age $15 - 19, 20 - 24, 30 - 34, 35 - 39, \ldots, 55 - 59$ and for men ($k = 2$) in the age groups between 15-59 (i.e. 8 and 9 age-specific parameters for women and men, respectively). (17 parameters)
- H_h: scale parameters for the trend in HIV incidence for population h. (11 parameters)

For a given set of parameter values, we obtain a set of not necessarily unique age- and sex-specific counts. These model outputs are used to calculate predicted values for the observed data. For example the ratio of the projected number of HIV+ women age 20-25 over the total number of women projected in that age group, is used to predict the observed HIV prevalence for women in that age group. Several types of observed data, such as HIV prevalence, are used to estimate the values of the parameters that are most likely.

In this section we discuss this topic in greater detail. The first focus is on the types of data used in the analysis. We then shift to the likelihoods specific to each type of data. Finally we turn to the techniques used to estimate the parameters, namely maximum likelihood (ML) estimation.

3.1 Data types

Heuveline (2003) uses data published in the literature to estimate the model parameters. These data consist of observations from eleven different East African populations col-

lected from antenatal clinics (ANCs), demographic surveillance sites, hospitals and general surveys. Both rural and urban areas are included, and the years of data collection range from 1989 to 1998. The data are classified into the following five categories (see Table 1, Heuveline 2003):

1. HIV test results in a general-population sample (10 data sets)
2. HIV test results in an ANC-patient sample (3 data sets)
3. HIV test results in all or a sample of births from HIV+ mothers (3 data sets)
4. HIV test results during a follow-up of an HIV– sample (3 data sets)
5. Survival during a follow-up of HIV+ individuals (3 data sets)

In addition to the data compiled by Heuveline, several new sources are also included in the analysis carried out here. All of the data sources are listed in Table 5, with the sources unique to the current analysis appearing in the shaded rows.

Although we have updated the information used to estimate the CCMPP parameters, we are unable to exploit many of the published results in the literature because those data are not broken down by age. For example, vertical transmission and the subsequent survival of infected infants has been the focus of a great deal of research (e.g. Simonon et al. 1994; Miotti et al. 1999; Spira et al. 1999; Nicoll et al. 2000; De Cock et al. 2000; Nduati et al. 2000; Mandelbrot et al. 2002; Petra Study Team 2002; Jackson et al. 2003; Read and Breastfeeding and HIV International Transmission Study Group 2004; Newell et al. 2004; Zijenah et al. 2004; Marston et al. 2005; Brahmbhatt et al. 2006; Kagaayi et al. 2008). While much has been learned from these studies, the published information is not stratified by the age of the mother, which would allow us to estimate the rate of vertical transmission by HIV duration groups. Without this level of detail, the CCMPP estimate of the vertical transmission rate is simply an average across all the estimates weighted by sample size. If future studies in this area tabulate the data by age of the mother it would greatly benefit modeling efforts.[8] Similarly, many of the results concerning the survival of infected individual (e.g. Van der Paal et al. 2007; Lutalo et al. 2007; Smith et al. 2007; Peters et al. 2007; Isingo et al. 2007; Murray et al. 2007), as well as HIV incidence and prevalence rates (e.g. Urassa et al. 2006; Żaba et al. 2000; Wambura et al. 2007) are not reported by age.[9] Small sample sizes may be a limiting factor, but we encourage future studies *to report age-specific results* whenever possible so that others can utilize the patterns observed along this critical dimension.

---

[8] The median survival time of children infected with HIV is around two years (Newell et al. 2004). Thus, the CCMPP model, which classifies the population into *five-year* age groups, can contribute very little to this question.

[9] Age-specific results are available for other geographic regions (e.g. Nyirenda et al. 2007), but we restrict our focus to countries in East Africa.

Table 5: **Descriptions for each source that contributes data used in the estimation of the CCMPP parameters**

Country	City/Town	Data Type	Citation	Year of Data Collection	Start Year of Epidemic	Urban/ Rural
Burundi	Bujumbura	1	Saidel et al. (1996)	1990	1973	Urban
	Bujumbura	2	Sokal et al. (1993)	1991-2	1973	Urban
	Bujumbura	4	Saidel et al. (1996)	1991	1973	Urban
Ethiopia	national	1	DHS/AIS[a]	2005	1980	National
Kenya	Mombasa	1	Hawken et al. (2002)[a]	2000	1974	Urban
	national	1	DHS/AIS[a]	2003	1980	National
Malawi	Mangochi	2	Slutsker et al. (1994)	1990	1975	Urban
	Karonga	1	McGrath et al. (2007)[a]	2005	1975	Rural
	Karonga	2	Crampin et al. (2008)[a]	1994-2001	1975	Rural
	Karonga	5	Crampin et al. (2002)[a]	1998-2000	1975	Rural
	Blantyre	2	Taha et al. (1998)[a]	1990, 1995	1975	Urban
	national	1	DHS/AIS[a]	2004	1980	National
Rwanda	national	1	DHS/AIS[a]	2005	1976	National
Tanzania	Mara	1	Shao et al. (1994)	1990	1975	Urban
	Mwanza	1	Grosskurth et al. (1995)	1992	1975	Rural
	Mwanza	1	Mwaluko et al. (2003)[a]	1994, 1999	1975	Rural
	Mwanza	4	Boerma et al. (1999)[a]	1994-7	1975	Rural
	Mwanza	5	Urassa et al. (2001)[a]	1994-8	1975	Rural
	Mwanza	5	Todd et al. (1997)	1994	1975	Rural
	Kagera	1	Kwesigabo et al. (2005)[a]	1987, 1996	1975	Urban
	Kagera	1	Kwesigabo et al. (2005)[a]	1996	1975	Rural
	national	1	DHS/AIS[a]	2004, 2007	1980	National
Uganda	Fort Portal	1	Kilian et al. (1999)	1995	1975	Urban
	Gulu	2	Fabiani et al. (2001)	1993, 1997	1975	Rural
	Masaka	1	Nunn et al. (1994)	1989	1975	Rural
	Masaka	2	Carpenter et al. (1997)	1990	1975	Rural
	Masaka	4	Kengeya-Kayondo et al. (1996)	1990-4	1975	Rural
	Masaka	4	Shafer et al. (2008)[a]	1995-2005	1975	Rural
	Masaka	5	Nunn et al. (1997)	1990	1975	Rural
	Rakai	1	Wawer et al. (1991)	1989-91	1980	Rural
	Rakai	4	Wawer et al. (1994)	1990	1980	Rural
	Rakai	1	Serwadda et al. (1992)	1989-91	1980	Rural
	Rakai	2	Gray et al. (1998)	1995	1980	Rural
	Rakai	5	Sewankambo et al. (1994)	1980	1990	Rural
	Rakai	5	Sewankambo et al. (2000)[a]	1994-8	1980	Rural
	Nsambya & Jinja	2	Asiimwe-Okiror et al. (1997)[a]	1990	1975	Urban

Table 5: **(Continued)**

Country	City/Town	Data Type	Citation	Year of Data Collection	Start Year of Epidemic	Urban/ Rural
Uganda	Nsambya, Jinja, & Rubaga	2	Asiimwe-Okiror et al. (1997)[a]	1995	1975	Urban
	national	1	DHS/AIS[a]	2004	1980	National
Zambia	Chelston	1	Fylkesnes et al. (1998)	1995	1975	Urban
	Kapiri Mposhi	1	Fylkesnes et al. (1998)	1996	1980	Rural
	Lusaka	3	Hira et al. (1989)	1987	1975	Urban
	national	1	DHS/AIS[a]	2002,2007	1976	National
Zimbabwe	Mutasa	1	Gregson and Garnett (2000)	1998	1977	Rural
	Manicaland	1	Gregson et al. (2002)[a]	2000	1976	Rural
	national	1	DHS/AIS[a]	2006	1980	National

Data types: (1) HIV test results in a general-population sample;
(2) HIV test results in an ANC-patient sample;
(3) HIV test results in all or a sample of births from HIV+ mothers;
(4) HIV test results during a follow-up of an HIV– sample;
(5) Survival during a follow-up of HIV+ individuals
Notes: The data sources that are unique to the current analysis appear in the shaded rows.
[a] New sources of data not included in the compilation used by Heuveline (2003). DHS/AIS refers to the Demographic and Health Survey program and the AIDS Indicator Surveys.

It should be noted that the CCMPP also requires vital rates for the uninfected population and an initial age distribution. These model inputs are taken from the United Nations global population estimates (United Nations 1999) and model life tables (United Nations 1982).

3.2 Likelihoods

Each category of data provides the pieces needed for a proportion which leads to the use of the binomial distribution in the likelihood specification. The binomial likelihood can be written as:

$$\mathcal{L} = \prod \binom{N}{x} \pi^x (1 - \pi)^{N-x} \tag{45}$$

where N is the total number of events, x is the number of "successes", π is the probability of success, and the product is taken over age, sex, the data types, and the various populations. The first two quantities N and x are taken from the data and we use CCMPP to calculate π given values for the parameters.

Before discussing the finer details of how the CCMPP outputs are used in the likelihood, it is important to cover two points. First, we need to temporally match up the CCMPP projections with the observed data. The start year for the model is the year when widespread transmission of HIV began, t_0. The population is then projected forward to the year when the data were collected. For example if widespread transmission in a country began in 1980 and the observed data are from 2000, then we can take the projected counts 20 years from the start time and compare these to the observed data. Given that the projections are in five year increments, it is sometimes necessary to take the average across two projection periods to match the year of data collection. Estimates of when widespread HIV transmission began are taken from a report by the United Nations (1998, Table 1).

Second, the data come from populations at twenty-six different locations.[10] At a given location there can also be several different types of data. For example data from the population living in Mwanza, Tanzania, include both HIV prevalence from a general population and survival information for those who are HIV+. As a result there is a separate likelihood for the 51 combinations of location, year, and data type retrieved from the literature. These are indicated using h for the population (and location) and c for the type or category of data. Finally, the data consist of sex- and age-specific information so the likelihoods may also be indexed by these characteristics as well.

3.2.1 HIV test results in a general-population sample

Various studies have collected data on sex- and age-specific HIV prevalence in a sample from the general population. The age groups range from 15-19 to 55-59. This type of data, labeled '1' ($c = 1$), usually includes the number of people tested and the percent who tested positive for HIV by sex and age. The likelihood, however, requires a count of individuals who are HIV+, so we calculate this quantity from the data and round it to the nearest integer. Let $N_{a,k,t,h,c=1}$ denote the total number of individuals in age group a of sex k at time t at location h and let $x_{a,k,t,h,c=1}$ be the number in this group who tested positive.

---

[10] See Table 5 for the locations.

$n_{a,d,t,h}$ the projected counts from CCMPP are used to predict sex- and age-specific prevalence for a given location as follows:

$$\pi_{a,k,t,h,c=1} = \frac{\sum_{d=2}^{5} n_{a,d,t,h}}{\sum_{d=1}^{5} n_{a,d,t,h}},$$ (46)

where the sum is taken across HIV duration groups. Having chosen the projection period that matches up with the year the data were observed, we can use $\pi_{a,k,t,h,c=1}$ in the binomial likelihood. A final note is that the observed data may be reported by age groups that do not match those of our projections. In this case weighted sums of the projected counts can be used to estimate HIV prevalence. For example if observed prevalence is reported for individuals age 17-25, then predicted prevalence can be calculated as:

$$\pi_{age=(17-25),k,t,h,c=1} = \frac{\sum_{d=2}^{5}(n_{a=4,d,t,h} \times 0.6 + n_{a=5,d,t,h})}{\sum_{d=1}^{5}(n_{a=4,d,t,h} \times 0.6 + n_{a=5,d,t,h})}.$$ (47)

This issue arises with the other data types as well.

3.2.2 HIV test results in an ANC-patient sample

Seven of the data sets used in this analysis provide age-specific information on HIV prevalence for female attendees of ANCs, typically from age 15 to 49. This type of data, indexed by $c = 2$, takes a form similar to the observed prevalence from a general population, except that they only refer to women. Both the total number of women tested $N_{a,k=1,t,h,c=2}$ and the age-specific prevalence are reported. The data are included in the binomial likelihood as counts, so we calculate the number of women who tested positive $x_{a,k=t,h}$ rounded to the nearest integer.

The predicted prevalence for the ANC attendees is calculated differently than for the general population. Recall that there are two primary assumptions of CCMPP concerning the fertility of HIV+ women. The first is that HIV+ women age 15-19 have higher fertility which is captured by the fertility selection parameter $e_{a=4}$. Second, fertility is expected to decline as time since infection increases, modelled by the fertility impairment parameters $g_{d>1}$. Since the HIV+ women observed in the data are pregnant, these parameters are included in the calculation of predicted prevalence. The formula is:

$$\pi_{a,k=1,t,h,c=2} = \frac{\sum_{d=2}^{5} n_{a,d,k=1,t,h} \times e_a \times g_d}{n_{a,d,k=1,t,h} + \sum_{d=2}^{5} n_{a,d,k=1,t,h} \times e_a \times g_d},$$ (48)

where the sum is taken over the duration groups. Having chosen the projection period that matches up with the year the data were observed, we can use $\pi_{a,k=1,t,h,c=2}$ in the binomial likelihood.

3.2.3 HIV test results in all or a sample of births from HIV-positive mothers

Heuveline (2003) found three data sets consisting of information on the fertility of HIV+ women. However one of these sources, Hira et al. (1989), differs from the others in that it provides information on whether or not an HIV+ mother infected her child. Data from the other two sources, Carpenter et al. (1997) and Gray et al. (1998), consist of the number of children born to both HIV+ and HIV– women, by age group. The likelihoods for the latter two sources are nearly identical to those in the data category $c = 2$, HIV test results in an ANC sample. The only difference is that the observed counts (i.e. the data) refer to the total number of children born to female ANC attendees in a specific age group, and the number of children born to HIV+ attendees. The probability that a child is born to an infected mother is calculated in exactly the same way as $\pi_{a,k=1,t,h,c=2}$. Given this similarity the data reported by Carpenter et al. (1997) and Gray et al. (1998) are classified here as $c = 2$.

Data that take the form of Hira et al. (1989) will also be indicated by $c = 3$. The corresponding counts used in the likelihood refer to the total number of children born to infected mothers in age group o, $N_{o,t,h,c=3}$, and the number of these children who are infected by their mothers $x_{o,d=2,t,h,c=3}$. The predicted rate of vertical transmission using the model outputs is calculated as:

$$\pi_{o,t,h,c=3} = \frac{\sum_{d=2}^{5} n_{a,k=1,d,t,h} \times f_{a,d=1,t,h} \times e_a \times g_d \times v_d}{\sum_{d=2}^{5} n_{a,k=1,d,t,h} \times f_{a,d=1,t,h} \times e_a \times g_d}, \tag{49}$$

where the sum is taken over the duration groups.

Unfortunately, this leaves only one data source to inform the estimate of the vertical transmission parameter. This issue is especially problematic when the parameter is constrained to be equal across duration groups (i.e. $v_d = v$, for all d). Note that all the like terms in the numerator and denominator cancel out, so the projections have no influence on the likelihood. Thus including more data of this type would be beneficial for future analysis using this model. As mentioned earlier, a substantial amount of the published data on vertical transmission does not stratify the information by age of the mother, which prevents us from utilizing these data to estimate the CCMPP parameters. Without this level of detail, the CCMPP estimate of the vertical transmission rate is simply an average across all the estimates weighted by sample size.

3.2.4 HIV test results during a follow-up of an HIV-negative sample

Data on sex- and age-specific HIV incidence are also used to estimate CCMPP parameters. These data indexed by $c = 4$ are typically reported in terms of the number of people who become infected and the total number of person-years lived while uninfected. For the binomial likelihood however, we need the counts of the initial population observed $N_{a,t,d=1,t,h,c=4}$ and the number who become infected $X_{a,t,d=2,t,h,c=4}$. Thus the initial population size needs to be calculated from the observed data, and this calculation can be done as follows:

$$\text{Initial Population} = \frac{\text{\# Converted}}{1 - \exp\left\{-T \times \frac{\text{\# Converted}}{\text{Person-Years}}\right\}} \tag{50}$$

where T is the total number of years observed.[11]

The model outputs are then used to calculate the probability of becoming infected for men or women in a certain age group at a given location and time. That is:

$$\pi_{a,k,d=1,t,h,c=4} = \frac{n_{a,k,d=1,t,h} \times s_{a,k,d=1,t,h} - n_{a+1,k,d=1,t+1,h}}{n_{a,k,d=1,t,h} \times s_{a,k,d=1,t,h}} \tag{51}$$

where $\pi_{a,k,d=1,t,h,c=4}$ is the proportion of HIV– women/men who become infected after five years. As discussed earlier, the period of observation for the data may not be equal to the projection interval of five years. If the observation period is only four years, then the quantity of interest is calculated as:

$$1 - \exp\left\{\frac{4}{5} \times \log\left(1 - \frac{\text{\# Converted}}{\text{Initial Population}}\right)\right\}. \tag{52}$$

3.2.5 Survival during a follow-up of HIV+ individuals

The final category of data describes the survival (mortality) of HIV+ individuals. This category indexed by $c = 5$ is similar to the previous one in that the data reported include the number of deaths observed among a cohort of HIV+ individuals of a particular age and sex $X_{a,k,t,h,c=5}$ and the number of person-years observed for each group. As before, the likelihoods require the initial population size for each group $N_{a,k,t,h,c=5}$ (see Equation 50). These two inputs $X_{a,k,t,h,c=5}$ and $N_{a,k,h,c=5}$ are the counts needed for

---

[11] In deriving this equation it is helpful to note that the number of person-years lived by a population of initial size N_0 and of size N_T T years later is equal to $\frac{(N_T - N_0) \times T}{\log(N_T/N_0)}$.

the binomial likelihood with the corresponding proportion referring to the probability of death over a given period of time.

The procedure for calculating the probability of death from the model outputs is best described in two steps. First, we calculate the probabilities by age, sex, and duration group. This can be written as:

$$q_{a,k,d,t,h,c=5} = 1 - \left(\frac{n_{a+1,k,d+1,t+1,h}}{n_{a,k,d,t,h}} \right)^{\frac{T}{5}} \quad \text{for} \quad d \geq 2 \tag{53}$$

where T is the number of years over which the data were observed. Since the observed data do not contain information on duration group, we must calculate the weighted average where the weights are the counts in each duration group. This second step is performed as follows:

$$\pi_{a,k,t,h,c=5} = \frac{\sum_{d=2}^{5} q_{a,k,d,t,h,c=5} \times n_{a,k,d,t,h}}{n_{a,k,d,t,h}}. \tag{54}$$

This is the probability used in the binomial likelihood.

3.3 Parameter estimation

A maximum likelihood (ML) approach is used to estimate the most likely parameter values (given the data and the model) and the uncertainty around those point estimates. Given the data from an individual site, the likelihood of a specific set of CCMPP parameter values can be calculated using the binomial expressions described above. There are twenty-two likelihoods in the original data compilation and an additional twenty-nine likelihoods included in the analysis below. We follow Heuveline (2003) and combine these likelihoods by taking the product across all locations and data types, assuming independence. The set of parameter values that maximizes the combined likelihood is the ML point estimate.

The ML estimation as well as the implementation of CCMPP is performed using the R programming language (R Development Core Team 2011). This languages provides an optimization routine `optim` that is used to find the parameter values that maximize the combined likelihood described above. In addition `optim` calculates the Hessian matrix of the likelihood function at the maximum. Standard errors are obtained by inverting the Hessian matrix (after multiplying it by negative one) and taking the square root of the diagonal elements.

4.   Results

Five sets of ML estimates for the CCMPP parameters are presented in Table 6, with the 95% confidence intervals shown in parentheses. The first column of estimates (moving from left to right) contains the results published by Heuveline (2003) and the second column lists our replication of his work using the same data. The point estimates for these two sets of results are generally consistent, as is seen with the vertical transmission rate. In both of the analyses it is estimated that 38.5% of children born to HIV+ mothers will be infected by their mothers, with the confidence intervals reaching from a low of around 30% to a high of nearly 48%. This result is well in line with previous estimates of vertical transmission in sub-Saharan Africa that range from 25% to 45% (Nicoll et al. 2000; Marston et al. 2005).[12] Both sets of results also suggest that infected women 15 to 19 years old have fertility that is considerably higher than HIV– women in the same age group. The multiplicative factor by which fertility is higher among infected women is equal to the early-selection fertility coefficient times the fertility impairment coefficient for women infected less than five years. These factors are 1.42 for Heuveline's estimates and 1.39 for our replication of his analysis using the same data. The fertility of women infected for five or more years is estimated by Heuveline to be lower than that of uninfected women or those infected for less than five years. Our results are consistent in that the fertility of women who have been infected for 5 to 9 years is likely to be at least 40% lower than uninfected women (i.e. the upper bound of the 95% confidence interval is 0.60). However, the large amount of uncertainty around the fertility impairment coefficient for women who have been infected for 10 or more years makes it difficult to draw conclusions about the fertility experience of these women. Finally, our estimated age patterns of HIV incidence, for both women and men, are similar to those reported by Heuveline (2003). Among women the risk of infection is highest for those 20 to 24 years old, then the relative incidence ratios decline with age. There is a similar pattern for men, but the estimated risk of infection peaks at an older age group, those 25 to 29 years old. Although our estimate of the relative incidence ratio for men 15 to 19 years old is higher than that of Heuveline and our 95% confidence intervals tend to be wider,[13] we feel that our results are similar enough to Heuveline's to justify using our implementation of the model to analyze the new data.[14]

---

[12] These estimates are higher than the 20% reported by Newell et al. (2004) from an analysis of data pooled across sub-Saharan Africa, but the infection status of 17% of the children was undetermined.

[13] It is worth noting that some of our 95% confidence intervals include negative values, which is troubling given that zero is the natural lower bound for the parameters. The same is true for the upper bound, namely 1.00, of the duration-specific fertility impairment coefficient for women infected for 0 to 4 years (5 to 9 years) when using gamma trend with the new data (all of the data) to estimate the parameters.

[14] For a given set of model inputs, our population projections match exactly those made by Heuveline. This suggests that differences between the two types of software (Microsoft Excel and R) used to carry out the

Table 6: **Maximum likelihood estimates of the CCMPP parameters for the original data sources compiled by Heuveline (2003), the new data sources only, and all of the data sources combined**

Parameters	Heuveline Original Data	Thomas & Clark Original Data	New Data Gamma	New Data Constant	All Data
Vertical Transmission Rate (%)	38.5	38.5	–	–	38.5
	(29.7, 47.8)	(29.3, 47.7)	–	–	(29.4, 47.6)
Early-Selection Fertility Coefficient	1.672	1.538	–	–	1.954
	(1.492, 1.865)	(1.028, 2.048)	–	–	(1.578, 2.330)
Duration-Specific Fertility Impairment Coefficients					
0-4 years	0.848	0.905	–	–	0.749
	(0.798, 0.909)	(0.664, 1.146)	–	–	(0.576, 0.922)
5-9 years	0.357	0.245	–	–	1.00
	(0.276, 0.450)	(-0.1, 0.59)	–	–	(0.749, 1.251)
10 years and above	0.293	0.461	–	–	0.433
	(0.078, 0.607)	(-0.027, 0.949)	–	–	(0.138, 0.728)
Female Age-Specific Relative Incidence Ratio					
15-9	0.594	0.585	0.762	0.564	0.849
	(0.545, 0.650)	(0.471, 0.698)	(0.672, 0.852)	(0.504, 0.623)	(0.738, 0.960)
20-4	1.325	1.297	1.496	1.183	1.255
	(1.239, 1.412)	(1.04, 1.553)	(1.279, 1.712)	(1.038, 1.328)	(1.499, 1.744)
25-9 (referent)	1.000	1.000	1.000	1.000	1.000
	–	–	–	–	–
30-4	0.752	0.724	0.922	0.877	0.975
	(0.647, 0.886)	(0.52, 0.927)	(0.718, 1.126)	(0.713, 1.041)	(0.750, 1.200)
35-9	0.635	0.518	0.790	0.702	0.586
	(0.482, 0.762)	(0.32, 0.716)	(0.554, 1.027)	(0.553, 0.851)	(0.403, 0.770)
40-4	0.551	0.577	0.686	0.551	0.807
	(0.409, 0.795)	(0.324, 0.83)	(0.375, 0.996)	(0.376, 0.725)	(0.559, 1.055)
45-9	0.356	0.339	0.177	0.323	0.300
	(0.159, 0.544)	(0.085, 0.594)	(-0.095, 0.450)	(0.170, 0.476)	(0.091, 0.509)
50-4	0.295	0.304	–	–	0.112
	(0.095, 0.679)	(-0.021, 0.63)	–	–	(-0.083, 0.307)
55-9	0.246	0.395	–	–	0.176
	(0.087, 0.627)	(0.027, 0.764)	–	–	(0.013, 0.338)

---

analysis are responsible for the discrepancies between the two sets of results. For example, it is possible to evaluate the binomial distribution for non-integer counts using the binomial function in Microsoft Excel, but this is not possible when using the binomial function in R. Therefore, we feel it is inappropriate to compare the log likelihood obtained via our estimation procedure to that of Heuveline. When we use our software and code to evaluate the log likelihood at Heuveline's ML estimates, the value is much lower than the log likelihood at our ML estimates (-654.02 vs -579.58).

Table 6: **(Continued)**

Parameters	Heuveline Original Data	Original Data	Thomas & Clark New Data Gamma	Constant	All Data
Male Age-Specific Relative Incidence Ratio					
15-9	0.059	0.11	0.262	0.167	0.274
	(0.024, 0.109)	(0.065, 0.155)	(0.214, 0.310)	(0.138, 0.196)	(0.226, 0.323)
20-4	0.583	0.508	0.427	0.329	0.466
	(0.483, 0.684)	(0.387, 0.63)	(0.333, 0.521)	(0.272, 0.385)	(0.376, 0.556)
25-9	1.149	1.024	1.171	0.883	1.216
	(0.986, 1.285)	(0.807, 1.242)	(0.980, 1.362)	(0.764, 1.002)	(1.033, 1.400)
30-4	0.963	0.998	1.387	1.201	1.383
	(0.773, 1.130)	(0.743, 1.252)	(1.128, 1.647)	(1.032, 1.370)	(1.150, 1.617)
35-9	0.759	0.723	0.849	0.894	0.708
	(0.573, 0.944)	(0.452, 0.993)	(0.533, 1.165)	(0.696, 1.091)	(0.485, 0.931)
40-4	0.769	0.759	0.818	0.835	1.088
	(0.554, 1.007)	(0.442, 1.077)	(0.413, 1.223)	(0.609, 1.062)	(0.790, 1.385)
45-9	0.622	0.628	0.444	0.335	0.379
	(0.409, 0.879)	(0.297, 0.959)	(0.145, 0.743)	(0.137, 0.534)	(0.130, 0.628)
50-4	0.417	0.288	0.519	0.458	0.483
	(0.120, 0.773)	(-0.093, 0.668)	(0.145, 0.743)	(0.249, 0.668)	(0.194, 0.772)
55-9	0.168	0.219	0.526	0.466	0.189
	(0.001, 0.445)	(-0.114, 0.552)	(0.081, 0.971)	(0.189, 0.743)	(-0.084, 0.462)
Log Likelihood	-567	-580	-1799	-2121	-3444

Note: There are two sets of estimates based on the new data sources: one using the gamma trend for HIV incidence, and another using a constant trend. The 95% confidence intervals for each parameter are shown in parentheses.

There are two sets of results, presented in Table 6, from our analysis of the new data sources. This new data compilation does not contain enough information to calculate the CCMPP parameters related to fertility or the relative incidence ratios for women in the 50-4 and 55-9 year age groups. The set of estimates shown under the column heading "Gamma" are based on the same specification of the HIV-enabled CCMPP used to analyze the original data. More specifically, the trend in HIV incidence is derived from the gamma distribution (see Equation 16 and Table 1) with a peak level of incidence occurring 12 years into the epidemic and a decline thereafter. The second set of results estimated with the new data sources is based on a constant rate in HIV incidence over time. These two sets of estimates are very similar to each other, and the estimated age pattern of HIV infection is consistent with the results obtained using the original data. The age profile for men, however, is different when the new data are used, regardless of the assumed underlying trend in HIV incidence. Although the uncertainty around the point estimates

makes it difficult to precisely identify the age at which the risk of infection peaks, the results suggest that men aged 30-34 experience the highest risk of infection. This is sligthly older than the peak age of infection for men estimated from the original data, the 25-29 year age group. This finding is particularly interesting given that the new data compilation consists of observations that are closer to the present, which may suggest that the age pattern of HIV incidence changes as the epidemic matures. Shafer et al. (2008) report a similar finding from a cohort study in Uganda. The results from the new data also suggest that the model based on the trend in HIV incidence derived from the gamma distribution fits the data better than a model with a constant trend in HIV incidence, as seen by comparing the log likelihoods for each model (gamma: -1790, vs. constant: -2121). Although the gamma trend fits these new data relatively well, the CCMPP estimates for more complicated models (presented in the next section) are much more unstable than those obtained from the specification of a constant incidence trend. Thus, a constant incidence trend is assumed for extensions of the CCMPP, and the corresponding results for the original model are presented here for the purpose of comparison.

The final set of estimates presented in Table 6 are the result of using all of the data to estimate the parameters. The estimate of the vertical transmission rate is virtually identical to the previous estimates, but the results concerning the other parameters related to fertility are noticeably different. When all the data are used in the analysis the point estimate of the early-selection fertility coefficient is higher, and the fertility impairment parameters are 0.749 (0-4 years), 1 (5-9), and 0.433 (10+) which is different from the monotonic decline we expect to see. Turning now to the relative incidence ratios, we see a different pattern of infection by age than what is estimated using solely the original data, or the new data only. When all the data are used, the risk of infection among women rises until peaking at ages 20-24 years, then declines before increasing again among those aged 40-44 years, and finally reaching much lower levels at the oldest ages. The age pattern for men is similar in that the risk of infection rises after the peak years of infection at ages 30-34. Similar patterns of HIV incidence have been observed in other African populations (Żaba et al. 2008), and a much more muted version is estimated using the original data. Perhaps the subsequent rise in the risk of infection after the peak is simply due to sampling variability or heterogeneity across the data compilations, but there may be behavioral patterns associated with union instability that can generate these results (Żaba et al. 2008).

4.1 Modeling survival for the infected population

One modeling assumption in the HIV-enabled CCMPP is that the lower survival rates associated with HIV/AIDS (captured by the parameters $s_{a,d}$ for $d \geq 2$) are known. While it is possible to specify a realistic survival schedule for the infected population, a more appealing extension of the model is to estimate the additional force of mortality associated with HIV/AIDS. This approach also seems reasonable since there are more data available on the survival experiences of HIV+ cohorts that are not included in the original data compilation used by Heuveline (2003). Recall that the projected counts in the CCMPP are calculated as

$$
\begin{aligned}
n_{a+1,1,t+1} &= n_{a,1,t} \, s_{a,1,t} \, (1 - i_{a,t}) \\
n_{a+1,2,t+1} &= n_{a,1,t} \, s_{a,1,t} \, i_{a,t} \, s_{a,2} \\
n_{a+1,d,t+1} &= n_{a,d-1,t} \, s_{a,1,t} \, s_{a,d} \quad \text{for } d > 2,
\end{aligned}
$$

with $s_{a,d \geq 2}$ capturing the reduced survival prospects associated with HIV/AIDS for each duration group. Instead of treating these model inputs as fixed, we can estimate the $s_{a,d \geq 2}$ in various ways. A parsimonious approach is to ignore the dependence on age at infection[15] and assume that survival is only a function of the duration of infection, which can be expressed as s_d for $d \geq 2$ (dropping the subscript for age group). These four new model inputs (one for each HIV+ duration group) can be estimated along with the other CCMPP parameters estimated in the previous section. The results for this model are presented in Table 7 along with estimates from the model with fixed survival, for the purpose of comparison. As with the previous results, estimates are shown for results obtained from using the original data compiled by Heuveline, the new data with the constant trend in HIV incidence, and all of the data sources combined.

---

[15] Heuveline (2003) specified the model so that expected survival time declines as the age at infection increases (among adults).

Table 7: **Maximum likelihood estimates for the modified CCMPP that includes duration-specific survivorship coefficients for the infected population**

Parameters	Original Data Fixed Surv.	Original Data	New Data	All Data
Female Age-Specific Relative Incidence Ratio				
15-9	0.585	0.529	0.621	0.837
	(0.471, 0.698)	(0.423, 0.635)	(0.537, 0.705)	(0.724, 0.950)
20-4	1.297	1.28	1.277	1.553
	(1.04, 1.553)	(1.039, 1.521)	(1.092, 1.461)	(1.316, 1.791)
25-9 (referent)	1.000	1.000	1.000	1.000
	–	–	–	–
30-4	0.724	0.703	0.930	0.891
	(0.52, 0.927)	(0.513, 0.893)	(0.736, 1.123)	(0.688, 1.093)
35-9	0.518	0.498	0.836	0.434
	(0.32, 0.716)	(0.308, 0.688)	(0.654, 1.019)	(0.266, 0.602)
40-4	0.577	0.543	0.534	0.822
	(0.324, 0.83)	(0.319, 0.766)	(0.343, 0.725)	(0.595, 1.048)
45-9	0.339	0.304	0.259	0.191
	(0.085, 0.594)	(0.076, 0.531)	(0.061, 0.457)	(-0.009, 0.392)
50-4	0.304	0.246	–	0.169
	(-0.021, 0.63)	(-0.052, 0.544)	–	(-0.058, 0.395)
55-9	0.395	0.331	–	0.171
	(0.027, 0.764)	(0.011, 0.65)	–	(-0.005, 0.346)
Male Age-Specific Relative Incidence Ratio				
15-9	0.11	0.1	0.183	0.263
	(0.065, 0.155)	(0.059, 0.141)	(0.147, 0.219)	(0.216, 0.310)
20-4	0.508	0.479	0.354	0.493
	(0.387, 0.63)	(0.367, 0.59)	(0.285, 0.422)	(0.403, 0.583)
25-9	1.024	0.982	0.942	1.219
	(0.807, 1.242)	(0.782, 1.182)	(0.797, 1.086)	(1.046, 1.393)
30-4	0.998	0.974	1.277	1.283
	(0.743, 1.252)	(0.739, 1.209)	(1.075, 1.478)	(1.062, 1.504)
35-9	0.723	0.71	0.866	0.721
	(0.452, 0.993)	(0.461, 0.959)	(0.642, 1.090)	(0.510, 0.932)
40-4	0.759	0.705	0.842	0.726
	(0.442, 1.077)	(0.425, 0.986)	(0.575, 1.110)	(0.481, 0.971)
45-9	0.628	0.564	0.442	0.477
	(0.297, 0.959)	(0.27, 0.858)	(0.189, 0.695)	(0.238, 0.717)
50-4	0.288	0.197	0.360	0.204
	(-0.093, 0.668)	(-0.152, 0.545)	(0.075, 0.646)	(-0.064, 0.472)
55-9	0.219	0.183	0.331	0.108
	(-0.114, 0.552)	(-0.105, 0.471)	(-0.013, 0.676)	(-0.085, 0.301)

Table 7: **(Continued)**

Parameters	Original Data Fixed Surv.	Original Data	New Data	All Data
Duration-Specific HIV Survivorship Coefficients				
0-4 years	–	1	1.000	1.000
	–	(0.843, 1.157)	(0.800, 1.200)	(0.922, 1.078)
5-9 years	–	0.602	0.821	0.658
	–	(0.557, 0.647)	(0.754, 0.889)	(0.602, 0.714)
10-4 years	–	0.609	0.726	0.611
	–	(0.29, 0.929)	(0.611, 0.840)	(0.451, 0.771)
15 years and above	–	1	0.147	0.968
	–	(-2.887, 4.887)	(0.043, 0.252)	(0.612, 1.325)
Log Likelihood	-580	-569	-2030	-3214

Notes: Results are presented for analyses obtained by using the original data compiled by Heuveline (2003), new data, and all of the data sources combined. The 95% confidence intervals for each parameter are shown in parentheses. Survival for the infected populations is equal to the product of these duration-specific coefficients and the appropriate survivorship ratio applied to the uninfected population; see text for more details. (CCMPP parameters related to fertility are not shown.)

First, focus on the results obtained using the original data compiled by Heuveline (2003). Estimates of the CCMPP parameters related to fertility are very similar to the corresponding results from the model with fixed survival (results not shown), and the same is true for the age pattern of HIV incidence for both men and women. Among the first HIV duration group, there is no significant difference between the survival prospects of the infected and uninfected populations, but survival declines by a factor of 0.6 (relative to the uninfected population in the same age group) for the next two HIV duration groups, 5 to 9 and 10 to 15 years. The survivorship coefficient for the longest HIV duration group is estimated to be 1, which suggests no difference in the survival prospects between those infected for 15 or more years and the uninfected population (in the corresponding age group). This result is obviously nonsense, but the large amount of uncertainty around the point estimate drastically reduces the amount of weight that should be placed on this finding. A final note is that the ML estimates for the CCMPP modified to estimate the survival of the infected population have a larger log likelihood than the result for the original model. This result is expected, since the former model includes four additional parameters, but it is encouraging to see the extension of the model gaining some traction.

Turning now to the results obtained using the new data, we see results that are qualitatively similar to those obtained from the model with a fixed survival schedule for the infected population. Among women, the risk of infection peaks among those aged 20-24, and declines monotonically into the older ages. Again, men in the 30-34 year age group are estimated to experience the highest risk of infection, which appears to be slightly older than the results obtained using the original data, but the uncertainty around the point esti-

mates make it difficult to draw any strong conclusions. The estimated survival parameters show a decline in the prospects of survival as the duration of infection increases, with a fairly large drop for the longest duration group. These additional parameters increase the log likelihood from -2,121 to -2,030, suggesting an improvement in the model fit to the data.

The final set of results shown in Table 7 pertains to estimating the survivorship coefficients of the CCMPP model with all of the data sources combined. The age patterns of HIV incidence are somewhat similar to the estimates from the model with a fixed survival schedule for the infected population. Among women, the risk of infection rises and then peaks at the 20-4 age group, with a subsequent increase among those aged 40-44. The age pattern for men is characterized by a peak in the risk of infection at ages 30-4, as seen before, but when survival is estimated, the subsequent rise in HIV incidence after the peak is smoothed out. The survival estimates suggest no difference in the survival prospects between the HIV+ population infected for less than five years and the uninfected population in the same age group. Survival decreases by a factor of around 0.66 for those infected for 5 to 9 years, but the point estimates do not show a continued decline for the longer duration groups and there is too much uncertainty around the estimates to detect any differences for these groups. Finally, we find that adding the survivorship coefficients to the estimation procedure increases the log likelihood at the ML estimates from -3,444 to -3,214, again suggesting that the model fit is improved when the survival schedules of the infected population are estimated.

Given the large number of parameters already included in the HIV-enabled CCMPP, the parsimonious model extension explored above is a reasonable first step. Previous research, however, suggests that the age at infection is a critical determinant of survival post infection (UNAIDS Reference Group on Estimates, Modelling and Projections 2002). One possible strategy is to take a parametric approach, which is useful here because it allows for more complicated patterns of mortality to be specified with only a few additional parameters. We follow the lead of Żaba et al. (2007) and use a Weibull distribution to model the change in the survival prospects of the infected population, so that it depends on age at infection. Let the mortality rate for those infected in age group a after t years of HIV infection be

$$m(a, t) = \alpha_a t^{\alpha_a - 1} e^{-\beta}$$

where α_a and β are the Weibull parameters to be estimated. The mortality rates from this model increase with higher values of the α_a parameter, whereas mortality decreases with higher values of the β parameter. To convert the mortality rate into the survivorship ratio needed for the CCMPP, exponentiate the negative of the cumulative hazard function over the appropriate range as follows

$$s_{a,d=2} = e^{-\int_0^5 m(a,t)\partial t}$$

$$s_{a,d=3} = e^{-\int_0^{10} m(a,t)\partial t}$$

$$s_{a,d=4} = e^{-\int_0^{15} m(a,t)\partial t}$$

$$s_{a,d=5} = e^{-\int_0^{20} m(a,t)\partial t}.$$

In the interest of parsimony, we combine several age groups and estimate two α_a parameters for those aged 15-29 and 30-59 years, as well as a single β parameter that is shared across all groups. Estimates of the CCMPP parameters using this specification of survival for the infected population are shown in Table 8. Results are only shown for both the original data and the new data, with the latter obtained from assuming a constant level of HIV incidence over time. The corresponding results obtained by combining all of the data sources are not shown because of the lack of stability in the parameter estimates.

Table 8: **Maximum likelihood estimates for the modified CCMPP with a Weibull model used to specify the survival for the infected population (see text for more details)**

Parameters	Original Data Fixed Surv.	Original Data	New Data
Female Age-Specific Relative Incidence Ratio			
15-9	0.585	0.544	0.653
	(0.471, 0.698)	(0.435, 0.654)	(0.573, 0.733)
20-4	1.297	1.278	1.261
	(1.04, 1.553)	(1.031, 1.525)	(1.080, 1.443)
25-9 (referent)	1.000	1.000	1.000
	–	–	–
30-4	0.724	0.742	0.913
	(0.52, 0.927)	(0.542, 0.943)	(0.707, 1.119)
35-9	0.518	0.561	0.751
	(0.32, 0.716)	(0.358, 0.764)	(0.569, 0.933)
40-4	0.577	0.618	0.878
	(0.324, 0.83)	(0.370, 0.866)	(0.628, 1.128)
45-9	0.339	0.320	0.288
	(0.085, 0.594)	(0.074, 0.566)	(0.062, 0.514)
50-4	0.304	0.293	–
	(-0.021, 0.63)	(-0.033, 0.620)	–
55-9	0.395	0.363	–
	(0.027, 0.764)	(0.016, 0.710)	–
Male Age-Specific Relative Incidence Ratio			
15-9	0.11	0.104	0.195
	(0.065, 0.155)	(0.060, 0.147)	(0.159, 0.232)
20-4	0.508	0.486	0.346
	(0.387, 0.63)	(0.371, 0.602)	(0.277, 0.416)
25-9	1.024	0.982	0.959
	(0.807, 1.242)	(0.777, 1.186)	(0.812, 1.106)
30-4	0.998	1.009	1.257

Table 8:   **(Continued)**

Parameters	Original Data Fixed Surv.	Original Data	New Data
Male Age-Specific Relative Incidence Ratio			
	(0.743, 1.252)	(0.758, 1.260)	(1.044, 1.469)
35-9	0.723	0.763	0.779
	(0.452, 0.993)	(0.489, 1.037)	(0.554, 1.005)
40-4	0.759	0.806	1.334
	(0.442, 1.077)	(0.495, 1.117)	(0.991, 1.677)
45-9	0.628	0.607	0.343
	(0.297, 0.959)	(0.282, 0.931)	(0.051, 0.635)
50-4	0.288	0.262	0.557
	(-0.093, 0.668)	(-0.121, 0.645)	(0.238, 0.867)
55-9	0.219	0.197	0.410
	(-0.114, 0.552)	(-0.113, 0.507)	(0.048, 0.771)
Weibull Parameter $\alpha_{\mathbf{a}}$			
15-29 years	–	2.007	4.612
	–	(1.214, 2.800)	(3.457, 5.767)
30-59 years	–	2.118	4.951
	–	(1.332, 2.904)	(3.766, 6.135)
Weibull Parameter β	–	0.557	1.340
	–	(0.372, 0.742)	(1.019, 1.661)
Log Likelihood	-580	-577	-2101

Notes: Results are presented for analyses obtained by using the original data compiled by Heuveline (2003) and the new data sources. The 95% confidence intervals for each parameter are shown in parentheses. (CCMPP parameters related to fertility are not shown.

Focusing first on the results obtained using the original data, we see that the estimates from the CCMPP with survival for the infected population based on the Weibull are very similar to the estimates based on the CCMPP with a fixed survival schedule for the infected population. The estimates of the Weibull parameters suggest, as expected, that the survival prospects are poorer for those infected in the 30-59 age category relative to the 15-29 category. The high level of uncertainty around these point estimates does not allow us to identify differences between the α_a parameters, and thus an even simpler model that ignores age at infection may suffice. In the corresponding model presented in Table 7, survival does not depend on age at infection and the log likelihood is -569, which compares favorably against the Weibull model shown in Table 8, with a log likelihood of -579.

Estimating the survival of the infected population via our Weibull specification with the new data results in age patterns of HIV incidence that are different from those presented earlier, particularly among men. Recall that in the previous results estimated from the new data risk of infection for men increased until peaking at ages 30-34, and then declined with age. For our Weibull specification, there appears to be a post-peak increase in HIV incidence among men aged 40-44. The width of the confidence intervals casts uncertainty over this finding, but it does seem clear that the estimated age pattern of HIV

incidence is sensitive to the survival schedule used for the infected population. A similar pattern is observed for women, with a peak in the risk of incidence at ages 20-24 and a subsequent increase among those aged 40-44, but the magnitudes of these differences are smaller than those observed among men. The survival estimates suggest that the risk of mortality is greater when HIV infection occurs after age 29, but there is too little power in these data to determine if this is a real difference. A final point is that when the new data are analyzed, the Weibull specification has a smaller log likelihood than the specification which ignores age at infection (-2101 vs. -2030).

To assist with the interpretation of our survival results, we used the estimated sampling distribution of the model parameters to generate 300 projections of the age-specific mortality rates for Uganda, as shown by the grey lines in the plots in Figure 2. The mortality rates shown in this figure correspond to the results estimated using the new data for time periods that are 5, 15, and 30 years into the HIV epidemic. Projections from the model based on directly estimating the survivorship coefficients are shown in the first row, and the projections from the Weibull model are presented in the second row. For comparitive purposes, fitted values from the model estimated by Żaba et al. (2007) are also included in each plot. Żaba et al. (2007) note that the mortality rates of the infected population will change as the epidemic matures and the average duration of infection increases. Their theoretical and empirical investigation points to two characteristics that help identify the appropriate age pattern of mortality (among the infected population) for a given stage of the epidemic. The first characteristic is a decline in HIV prevalence, indicating higher mortality, and the second is the ratio of HIV prevalence among the dead to HIV prevalence among the living, which mirrors the infected mortality rate at a lag of roughly three years. Żaba et al. (2007) include these indicators as covariates in a Weibull model to estimate the mortality rates of the infected population using data from various cohort studies in sub-Saharan Africa. Their model produces four age patterns of mortality for infected individuals: HIV prevalence is *not declining* and the ratio of dead to living HIV prevalence is *less than* 4 (denoted as Żaba 1); HIV prevalence is *declining* and the ratio of dead to living HIV prevalence is *less than* 4 (Żaba 2); HIV prevalence is *not declining* and the ratio of dead to living HIV prevalence is *greater than* 4 (Żaba 3); and HIV prevalence is *declining* and the ratio of dead to living HIV prevalence is *greater than* 4 (Żaba 4). These four patterns are included in the plots in Figure 2.

Figure 2: **Estimated mortality rates for Uganda based on CCMPP parameter estimates for survivorship model (method 1), the Weibull model (method 2), and model age-specific mortality patterns estimated from Żaba et al. (2007)**
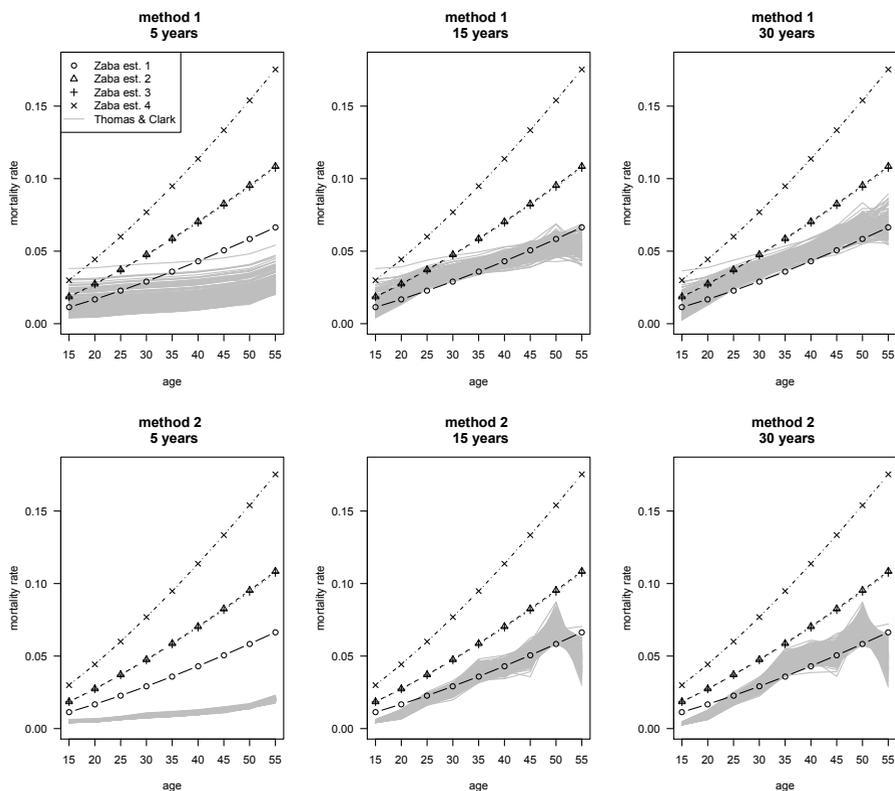
Note: See text for more details. Żaba1 – HIV prevalence is not declining, ratio of HIV prevalence among the dead to prevalence among the living < 4; Żaba2 – HIV prevalence is declining, ratio < 4; Żaba3 – HIV prevalence is not declining, ratio > 4; Żaba4 – HIV prevalence is declining, ratio > 4.

Both of the CCMPP specifications produce the same general mortalty patterns by age, as seen in Figure 2. As the epidemic matures and the average duration of infection increases the mortality rates at the older ages increase. There is very little change over time at the younger ages because these groups primarily consist of recent infections (due

to the close proximity between the current age and the age at which risk exposure, via sexual intercourse, begins). Our estimates suggest that the mortality rates among the infected population continue to increase over the first three decades of the epidemic. For the Weibull specification of the CCMP, the dual peaks in the estimated age patterns of HIV incidence, particularly among men, are reflected in the age-specific mortality rates among those aged 35-39 and 50-54. The levels of mortality associated with our estimates are generally lower than the age patterns estimated by Żaba et al. (2007). This descrepancy is most likely attributable to the differences in the sources of data used in each analysis. The vast majority of the data used to generate our results consists of observations on HIV prevelance, whereas Żaba et al. (2007) conduct a survival analysis which primarily uses observations on deaths and person-years at risk. Given that the approach of Żaba et al. (2007) is more direct, this comparison suggests that we may be underestimating the force of mortality experienced by infected individuals.[16] Some studies have reported mortality rates among the infected population that are fairly similar to our estimates. For example, a study of HIV-positive South African miners reports a crude mortality rate (from natural causes) of .0271 during the 1990s (Murray et al. 2007), and point estimates of age-specific mortality rates from a cohort study in Zimbabwe are less than .07 (Smith et al. 2007). Data from Kwazulu-Natal, South Africa, collected during the period from 2004 to 2006, suggest that mortality is slightly higher among the infected population, with average annual mortality rates of .025, .060, .076, and .0100 among those 15-24, 25-34, 35-44, and 45-54 years of age (Nyirenda et al. 2007). Estimates of the median survival time from HIV infection to death include 8.7 years in Rakai, Uganda (Lutalo et al. 2007), 9.0 years in Masaka, Uganda (Van der Paal et al. 2007), and 11.9 years in Kigali Rwanda (Peters et al. 2007), which also suggest that our mortality estimates, and the associated median survival times of roughly 14 (method 1) and 16 (method 2) years, are understating the probability of dying among the infected population.

As a further check on the validity of our parameter estimates for the CCMPP with estimated survival for the infected population, we compare our estimated age patterns of HIV incidence to those of Żaba et al. (2008). Again, we use the estimated sampling distribution of the model parameters to generate 300 projections of the annual age-specific HIV incidence rates for women and men in Rakai, Uganda; Kisesa, Tanzania; and Manicaland, Zimbabwe. Figure 3 shows our estimated age patterns (grey lines) along with those esimated by Żaba et al. (2008), who fit a cubic splines model to survival data on HIV incidence for these same populations. Our estimates correspond to the CCMPP specification in which we directly estimate the survivorship coefficients (see Table 7).

---

[16] Another possible reason that our results are lower is that the rates we are using for background mortality, i.e. the force of mortality in the absence of HIV/AIDS, are too high, since the lowering background mortality would require an increase in mortality among the infected to produce the same population projections.

Figure 3:   **A comparison of age patterns of annual HIV incidence rates in East Africa between CCMPP estimates (grey lines) and those estimated by Żaba et al. (2008) (black lines)**
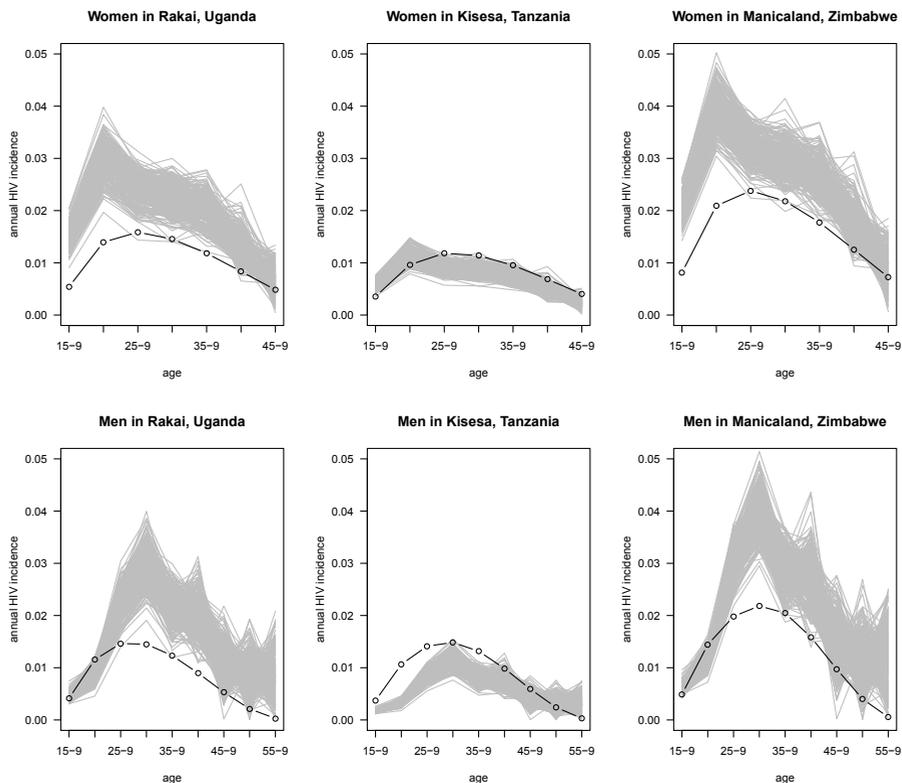
There are two types of differences between the age patterns shown in Figure 3. The first refers to the overall level of the age pattern, and the second consists of differences in the shape of the pattern. Our estimates for both women and men tend to be higher for Rakai and Manicaland, but lower in Kisesa. Relative to our results, the estimates of Żaba et al. (2008) peak at an older age among women, while the shape of the age patterns among men is fairly similar, with the possible exception that our results are lower at the youngest ages. Before addressing the possible explanations for these differences, it is important to note that none of the uncertainty around the estimates of Żaba et al. (2008) is presented

in these plots, which may exagerate the magnitude of the differences. One possible cause of the differences is that very little data are used to estimate the size of the epidemic in Rakai and Manicaland. For each of these sites, only one data set is used to estimate the CCMPP parameter (see Equation 4), whereas three data sets contribute information to the size of the epidemic in Kisesa and the resulting level differences between the two models are smaller.

The discrepancies observed in Figure 3 may also be due to misspecifications in the CCMPP. For example, our results are based on the assumption of a flat trend in HIV incidence, which may be unrealistic and erroneously influence the estimated age pattern. Heuveline (2003) found only minor differences in the age patterns of HIV incidence based on the CCMPP with different trends in incidence. Similarly, we find similar patterns based on different trends (see Table 7). While these findings suggest that the age profile is fairly insensitive to the assumed trend in HIV incidence for the CCMPP, the question is still open, given that only a small number of different trends have been explored with this model. Another problem may be that we are assuming that the shape of the age pattern of HIV incidence stays the same over the course of the epidemic. Basic demographic changes, as well as differential behavioral responses by age could generate these dynamics in the risk of infection (Żaba et al. 2008). A final point is that our age patterns are estimated primarily using data on HIV prevalence, which depends on both incidence and survival. Our earlier findings, shown in Figure 2, suggest that we may be underestimating the mortality rates of the infected population. If the CCMPP parameters governing the survival of the infected population are decreased to reflect higher mortality, then the incidence parameters would also need to be increased to generate the same levels of HIV prevalence. Since our findings suggest that the bias in our mortality estimates varies systematically by age, with larger differences at older ages, correcting this bias would most likely change the estimated age pattern of HIV incidence. As stated earlier, future studies on the survival of the infected population should publish their results by age so that this information can be utilized by age-specific models such as CCMPP.

4.2 An application to a rural South African population

The estimates presented above are obtained using data from countries located in East Africa. This gives rise to the question of regional variation within sub-Saharan Africa. We explore this question with an application of CCMPP to a population living in a rural area of the KwaZulu-Natal province of South Africa. Data from this population published by Welz et al. (2007) include age-specific HIV prevalence for the following age groups: 15-19, ..., 45-49 for women; and 15-19, ..., 50-54 for men.[17] The data are further

---

[17] We also attempted to use the mortality rates of the infected population published by Nyirenda et al. (2007), but we were unable to obtain maximum likelihood estimates for which we could estimate a covariance matrix.

classified by resident status, with non-residents being individuals whose main residence is elsewhere but who maintain regular contact with the surveyed household through periodic visits. Prevalence is generally much higher among non-residents, which is consistent with the finding that migrants are at a higher risk of becoming infected. We model this difference by estimating CCMPP with separate parameters for the level of the epidemic for residents and non-residents. The results from this model are presented in Table 9 under the column labeled 'full model'. We also report estimates for a reduced model, for which incidence is constrained to be equal across certain age groups: for women 30-49 and for men 25-34 and 35-54.

Relative to males, incidence among females is much higher in the youngest age group. Incidence peaks among women between the ages of 20 and 29 while the peak for men occurs during the late twenties and early thirties. There is a relatively large drop in relative incidence after the peak for women. Recall that women between the ages of 25 and 29 serve as the reference group, with their relative incidence fixed at a value of 1. Relative incidence for women in the next oldest age group has a 95% confidence interval of (-0.11, 0.51). There is a similar finding for men, but it is much less pronounced. With respect to the two resident groups, residents and non-residents, the estimates do not indicate a difference in the level of the epidemic.

The amount of uncertainty around the estimates in the full model motivated the estimation of a more parsimonious model. Note how the confidence intervals become increasingly wider with age.[18] With relatively few data, we are unable to identify differences between the point estimates for relative incidence after the peak. The 'reduced model' column in Table 9 displays estimates for a simpler model that collapses all post-peak age groups for women and includes only two parameters for relative incidence among men older than 34. The results show an age pattern similar to that seen for the full model, and there is more precision around the estimates for this reduced model. We again find little support for the hypothesis that the HIV epidemic is at a higher level among non-residents, but in both cases the estimates of epidemic scale are much higher than those from the sites in East Africa.

---

[18] It should also be pointed out that two of the confidence intervals include negative values that are problematic, since these parameters are naturally bounded at zero.

Table 9: **CCMPP parameter estimates and 95% confidence intervals using data from KwaZulu-Natal, South Africa**

Parameters	Full Model	Reduced Model
Female Relative Incidence Ratio		
15-9	0.26	0.27
	(0.2, 0.33)	(0.22, 0.32)
20-4	0.63	0.64
	(0.43, 0.82)	(0.49, 0.79)
25-9	1	1
	–	–
30-4	0.2	0.29^a
	(-0.11, 0.51)	(0.24, 0.34)
35-39	0.4	
	(0.15, 0.64)	
40-44	0.25	
	(0.02, 0.49)	
45-49	0.21	
	(0.04, 0.37)	
Male Relative Incidence Ratio		
15-19	0.02	0.02
	(0.01, 0.03)	(0.01, 0.03)
20-24	0.2	0.21
	(0.14, 0.27)	(0.16, 0.27)
25-29	0.69	0.65^b
	(0.47, 0.91)	(0.54, 0.77)
30-34	0.58	
	(0.28, 0.88)	
35-39	0.22	0.28^c
	(-0.12, 0.56)	(0.23, 0.34)
40-44	0.4	
	(0.05, 0.74)	
45-49	0.18	
	(-0.09, 0.45)	
50-54	0.32	
	(0.09, 0.55)	
Scale		
Residents	1.77	1.73
	(1.41, 2.13)	(1.5, 1.95)
Non-Residents	2.45	2.39
	(1.86, 3.03)	(1.95, 2.83)

Notes: [a] This is the parameter estimate for women between the ages of 30-49.
[b] This is the parameter estimate for men between the ages of 25-34.
[c] This is the parameter estimate for men between the ages of 35-54.

An interesting comparison between the full and reduced models is made by projecting prevalence and combining the age groups to obtain an estimate of adult prevalence which can then be compared to observed prevalence.[19] To get a sense of the uncertainty around these estimates, we appeal to asymptotics and sampled CCMPP inputs from a multivariate normal distribution, with the mean and covariance matrix taken from the ML estimates. For each set of inputs in the sample we run CCMPP and the corresponding outputs are used as a predictive distribution. The 0.025 and 0.975 quantiles of this predictive distribution are reported together as the 95% predictive interval (95% PI).

Welz et al. (2007) report that 21.5% of all adult residents aged 15-49 are HIV+ (prevalence for non-residents is not given). Using the ML estimates as the model inputs gives a predicted prevalence of 25.9% (95% PI: 24.8 – 26.9%) which is much too narrow. Figures 4 & 5 show the projected prevalence for women and men, respectively. There is a pair of boxplots for each five-year increment over time up to 2003, the year of data collection, and ten years into the future. The boxplot on the left (right) refers to the full (reduced) model. Among women the projections match the observed prevalence of adults much better for non-residents than for residents. Predicted prevalence is too high for adult female residents. As seen in Panel (a) of Figure 4 the predictive interval for the full model barely includes the observed prevalence in 2003, while the corresponding interval for the reduced model is too narrow. Among non-resident women (Panel (b)), for whom prevalence is much higher, the predictive intervals for both models cover the observed prevalence in 2003. Finally the predictive intervals tend to be larger for non-residents in both the full and reduced models.

The results for men are shown in Figure 5, with residents presented in Panel (a) and non-residents in Panel (b). The findings are similar to those of women, in that the projections tend to be too high for residents and too low for non-residents. The fit appears to be slightly worse among men in that the predictive intervals appear to be too narrow. As found with women, the reduced model provides very similar projections to the more complicated model.

---

[19] This partially serves as a check for the age distribution of the initial population and the choice of vital rates.

Figure 4: **Projected prevalence over time for women aged 15-50**
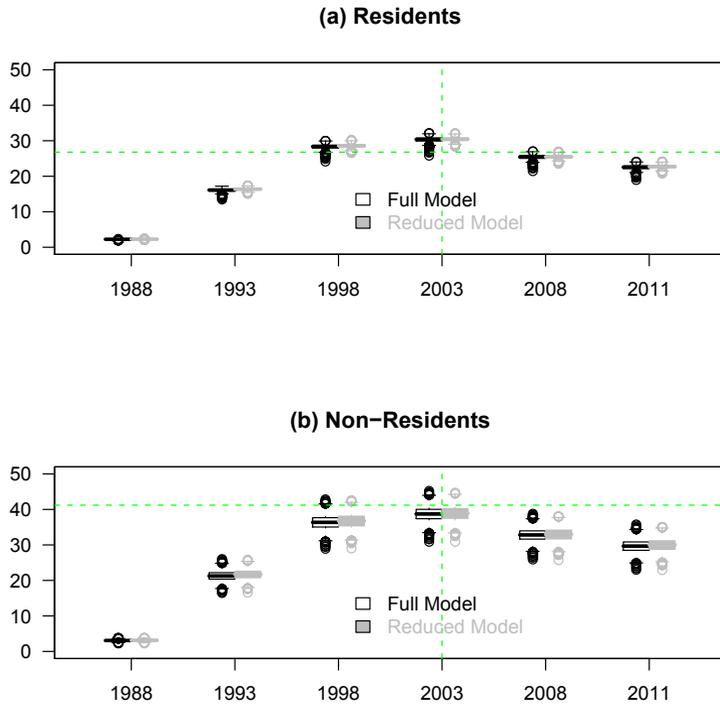
(a) Residents

(b) Non−Residents

Note: The horizontal and vertical dashed lines indicate the prevalence and year, respectively, from the observed data.

**Figure 5:**   **Projected prevalence over time for men aged 15-54**
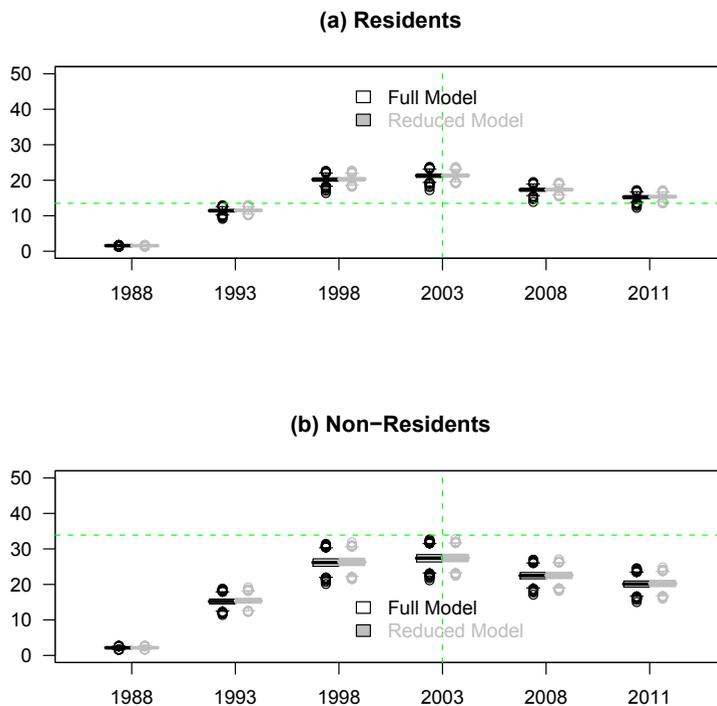
(a) Residents

(b) Non−Residents

Note: The horizontal and vertical dashed lines indicate the prevalence and year, respectively, from the observed data.

Given the high level of observed prevalence in KwaZulu-Natal it is interesting to explore the long term implications of the estimated CCMPP parameters for the population. The eigenvector corresponding to the largest real eigenvalue of the Leslie matrix is the stable age distribution (Keyfitz and Caswell 2005). The population pyramid for the stable equivalent population is presented in Figure 6 with men on the left and women on the right. The white horizontal bars represents the proportion of the total population in an age group and the black section of each bar indicates the proportion of the population in that age group that is HIV+. The dots at the end of each bar indicate the 95% confidence intervals around the proportion represented by the bar – they are very narrow.

In the stable population HIV prevalence resembles the age profile for incidence. There is an earlier peak for women occurring among women between the ages of 25 and 29. Among men the peak is at the next older age group, and HIV prevalence is generally lower relative to women. Prevalence is so high among women (25%) that the female stable population is actually shrinking. In contrast the male stable population (prevalence 17%) is growing slowly – which hints at possible pressure on the sex ratio of this population.

Figure 6:   **Projected population pyramid for the stable equivalent population for KwaZulu-Natal South Africa**

Stable Age Distribution
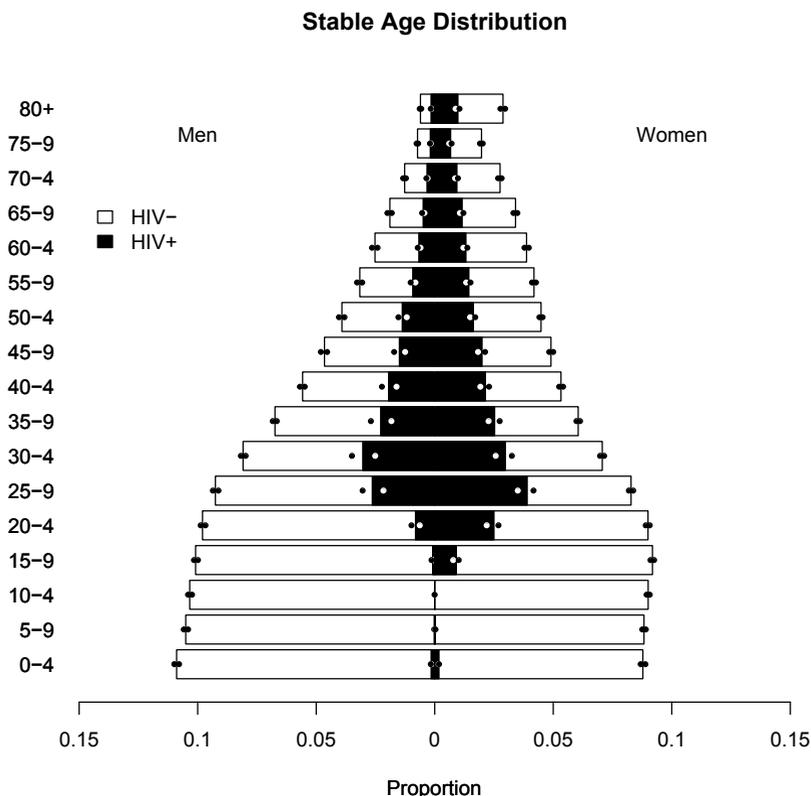
Note: The white bars indicate the proportion of the population in each age group (HIV– and HIV+ combined), while the black bars indicate the proportion of the total population who are HIV+ in each age group. 95% confidence intervals are depicted by the dots.

5.   Discussion

In this article we extend the work of Heuveline (2003) by developing a Leslie matrix representation of his HIV-enabled CCMPP, updating the data sources used to estimate the parameters, estimating the survival of the infected population, and by providing new estimates for a rural population living in the KwaZulu-Natal province of South Africa. Our findings based on the original data are broadly similar to Heuveline's, in that for women incidence peaks between the ages of 20 and 24 and for men between the ages of 25 and 34. However, we also find that the estimated age patterns are sensitive to the data used to fit the CCMPP. When "newer" data, collected more recently than those compiled by Heuveline (2003), are analyzed the peak age of HIV incidence among men is more narrowly peaked in the 30 to 34 year age range. Among men, the age pattern estimated from the South African data has the same shape, but the risk of HIV incidence peaks among women in the 25-29 age group.

One of the benefits of a modeling approach to studying HIV/AIDS epidemics is that data on prevalence, incidence, and survival can be jointly utilized to estimate consistent age patterns for each outcome (Ghys et al. 2006). Our results, based on generalizing the HIV-enabled CCMPP to estimate mortality associated with HIV, suggest that a simple model of survival that is independent of the age at infection can perform quite well relative to the CCMPP based on a fixed survival schedule. It is possible to specify more complicated models for the survival of the HIV-positive groups, but our attempt at this has added very little to the fit of the model (as indicated by changes in the log likelihood evaluated at the ML estimates). A comparison of our estimates to those from a different study, however, suggests that we may be underestimating the mortality rates for the infected population. This potential bias is troubling since estimates of HIV incidence depend on the pattern of mortality among the infected population, as noted in previous work on devising methods for estimating incidence by age (e.g. Hallett et al. 2008).

The shortcomings in our analysis have implications for modeling efforts and population projections more generally. Perhaps the most likely cause of the discrepancies between our estimates and those of other studies is that we are not using enough information to estimate the parameters, particularly with respect to the survival experiences of the infected population. There are numerous high quality HIV cohort studies (e.g. Van der Paal et al. 2007; Lutalo et al. 2007; Smith et al. 2007; Peters et al. 2007; Isingo et al. 2007; Murray et al. 2007), and we encourage future work with these data to stratify their results by age (with reports of the number of events and person-years at risk when appropriate) so that age-specific models such as CCMPP can take advantage of these valuable resources. This will be very helpful in future efforts aimed at modeling the dynamics associated with the expanding availability and coverage of antiretroviral therapies.

Another implication of our work is highlighted by the potential misspecifications of

the CCMPP, with the most problematic being the assumed trend in HIV incidence. The empirical record suggests that both mortality and fertility depend on the duration of HIV infection (Porter and Żaba 2004; Żaba et al. 2007; Lee et al. 2000; Desgrées du Loû et al. 1999; Sedgh et al. 2005; Nguyen et al. 2006; Hunter et al. 2003). As an epidemic matures, the distribution of duration times will change, and thus the impact of HIV at the population level will also change over time. Modeling this dynamic depends heavily on the trend in HIV incidence, yet we are assuming a specific trend and treating it as a fixed input to the CCMPP model. This obvious weakness is introduced for simplicity and tractability, but improvements can be made by furthering our understanding of HIV incidence and how the distribution of duration times change as the epidemic matures. Some steps have been made in this direction by Hallett et al. (2008), who classify the maturation of epidemics into an early stage, a mature and stable stage, and a mature and declining stage, and condition their estimates of HIV incidence on this dimension. More work needs to be done along these lines, with the aim of quantifying the magnitude of differences in the level of incidence between these stages and the duration of each stage. Continued tracking of age-specific trends in HIV prevalence (and accounting for changes due to the relative contributions of incidence and mortality) is a useful tool for moving forward on this front (Ghys et al. 2006; Stover et al. 2006; Hallett et al. 2008). Also, explicitly modeling and estimating the heterogeneity of sexual risk behavior, across age and over time, will help to improve models for population projections in the context of HIV/AIDS (Gregson et al. 2007; Stover et al. 2006). While it is plausible that the age pattern of infection remains stable over time and the overall level changes as the epidemic matures, it is just as reasonable to expect changes in the age pattern of HIV incidence due to age-specific behavioral responses to the risk of infection (Stover et al. 2006; Żaba et al. 2008). Incorporating behavioral data into the specification of the trend in HIV incidence is a very promising way of improving the CCMPP and other models for population projection.[20]

Our final point is a methodological one which is motivated by the fact that our results raise some questions concerning the reliability of the maximum likelihood approach to estimate the CCMPP parameters. We find that several confidence intervals for the CCMPP parameters include negative values, which is troubling since all of the parameters are bounded below by zero. Furthermore, there is some evidence that the ML estimates understate the uncertainty around the parameter estimates. Recall that the predictive intervals for adult prevalence at the KwaZulu-Natal site were too narrow for resident men and women as well as non-resident men, although this may be related to other sources of

---

[20] Accounting for changes in the infectivity of an individual as a function of the duration of infection, as suggested by Leclerc and Garenne (2007), could also enhance the CCMPP.

uncertainty not captured by the model (e.g. uncertainty around the trend in HIV incidence based on the gamma trend).

An alternative approach to estimation is the use of Bayesian methods. Poole and Raftery (2000) developed the Bayesian melding technique specifically for use with deterministic models like CCMPP, see also Alkema, Raftery, and Clark (2007). A Bayesian framework is particularly appealing given the potential need for model comparison, such as the full and reduced models fit to the South African data. Recall that for the KwaZulu-Natal model the projections for adult prevalence were generally too low for non-residents and too high for residents. Given that prevalence is higher among women, it may be inappropriate to model the epidemic's level for men and women simultaneously, since non-resident men and resident women have similar levels of prevalence. This suggests another model with level parameters for each combination of sex and resident status, perhaps with additional variation related to the full and reduced specifications shown in Table 9. In a situation like this it would be invaluable to have an estimation approach that objectively compares or even combines model fits, and Bayesian approaches exist for both, see Raftery (1995). Future modeling efforts of HIV/AIDS epidemics should consider these tools for estimation, projection, and quantifying the uncertainty around the events of interest.

6. Acknowledgements

References

Alkema, L., Raftery, A.E., and Clark, S.J. (2007). Probabilistic projections of HIV prevalence using Bayesian melding. *The Annals of Applied Statistics* 1(1): 229–248. doi:10.1214/07-AOAS111.

Asiimwe-Okiror, G., Opio, A.A., Musinguzi, J., Madraa, E., Tembo, G., and Caraël, M. (1997). Change in sexual behaviour and decline in HIV infection among young pregnant women in urban Uganda. *AIDS* 11(14): 1757–1763. doi:10.1097/00002030-199714000-00013.

Boerma, J.T., Urassa, M., Senkoro, K., Klokke, A., and Ng'weshemi, J.Z.L. (1999). Spread of HIV infection in a rural area of Tanzania. *AIDS* 13(10): 1233–1240. doi:10.1097/00002030-199907090-00013.

Bowley, A.L. (1924). Births and population in Great Britain. *Economic Journal* 34(134): 188–192. doi:10.2307/2223159.

Brahmbhatt, H., Kigozi, G., Wabwere-Mangen, F., Serwadda, D., Lutalo, T., Nalugoda, F., Sewankambo, N., Kiduggavu, M., Wawer, M., and Gray, R. (2006). Mortality in HIV-infected and uninfected children of HIV-infected and uninfected mothers in rural Uganda. *Journal of Acquired Immune Deficiency Syndrome* 41(4): 504–508. doi:10.1097/01.qai.0000188122.15493.0a.

Cannan, E. (1895). The probability of a cessation of the growth of population in England and Wales during the next century. *Economic Journal* 5(20): 505–515. doi:10.2307/2956626.

Carpenter, L.M., Nakiyingi, J.S., Ruberantwari, A., Malamba, S.S., Kamali, A., and Whitworth, J.A. (1997). Estimates of the impact of HIV infection on fertility in a rural Ugandan population cohort. *Health Transition Review* 7(Supplement 2): 113–126.

Central Bureau of Statistics (CBS) [Kenya], Ministry of Health (MOH) [Kenya], and ORC Macro (2004). *Kenya Demographic and Health Survey 2003*. Calverton, Maryland: CBS, MOH, ORC Macro (490 pp.).

Central Statistical Agency [Ethiopia] and ORC Macro (2006). *Ethiopia Demographic and Health Survey 2005*. Addis Ababa, Ethiopia and Calverton, Maryland, USA: Central Statistical Agency and ORC Macro (410 pp.).

Central Statistical Office (CSO), Ministry of Health (MOH), Tropical Diseases Research Centre (TDRC), University of Zambia, and Macro International Inc. (2009). *Zambia Demographic and Health Survey 2007*. Calverton, Maryland, USA: CSO and Macro International Inc. (490 pp.).

Central Statistical Office (CSO) [Swaziland], and Macro International Inc. (2008). *Swaziland Demographic and Health Survey 2006-07*. Mbabane, Swaziland: Central Statistical Office and Macro International Inc.

Central Statistical Office (CSO) [Zambia], Central Board of Health [Zambia], and ORC Macro (2003). *Zambia Demographic and Health Survey 2001-2002*. Calverton, Maryland, USA: Central Statistical Office, Central Board of Health, and ORC Macro (388 pp.).

Central Statistical Office (CSO) [Zimbabwe] and Macro International Inc. (2007). *Zimbabwe Demographic and Health Survey 2005-2006*. Calverton, Maryland, USA: CSO and Macro International Inc. (456 pp.).

Chin, J. and Lwanga, S. (1991). Estimation and projection of adult AIDS cases: A simple epidemiological model. *Bulletin of the World Health Organization* 69(4): 399–406.

Clark, S.J., Eaton, J.W., Elmquist, M.M., Ottenweiller, N.R., and Snavely, J.K. (2008). Demographic consequences of HIV epidemics and effects of different male circumcision intervention designs: Suggestive findings from microsimulation. Seattle: Center for Statistics and the Social Sciences.

Crampin, A.C., Floyd, S., Glynn, J.R., Sibande, F., Mulawa, D., Nyondo, A., Broadbent, P., Bliss, L., Ngwira, B., and Fine, P.E.M. (2002). Long-term follow-up of HIV-positive and HIV-negative individuals in rural Malawi. *AIDS* 16(11): 1545–1550. doi:10.1097/00002030-200207260-00012.

Crampin, A.C., Jahn, A., Kondowe, M., Ngwira, B.M., Hemmings, J., Glynn, J.R., Floyd, S., Fine, P.E.M., and Żaba, B. (2008). Use of antenatal clinic surveillance to assess the effect of sexual behavior on HIV prevalence in young women in Karonga District of Malawi. *Journal of Acquired Immune Deficiency Syndrome* 48(2): 196–202. doi:10.1097/QAI.0b013e31817236c4.

Day, J.C. (1996). Population projections of the United States by age, sex, race, and Hispanic origin: 1995 to 2050. U.S. Bureau of the Census. (Current Population Reports; P25-1130).

De Cock, K.M., Fowler, M.G., Mercier, E., de Vincenzi, I., Saba, J., Hoff, E., Alnwick, D.J., Rogers, M., and Shaffer, N. (2000). Prevention of mother-to-child HIV transmission in resource-poor countries. *Journal of the American Medical Association* 283(9): 1175–1182. doi:10.1001/jama.283.9.1175.

Desgrées du Loû, A., Msellati, P., Yao, A., Noba, V., Viho, I., Ramon, R., Welffens-Ekra, C., and Dabis, F. (1999). Impaired fertility in HIV-1-Infected pregnant women: A clinic-based survey in Abidjan, Côte d'Ivoire, 1997. *AIDS* 13(4): 517–521.

doi:10.1097/00002030-199903110-00011.

Dorn, H.F. (1950). Pitfalls in population forecasts and projections. *Journal of the American Statistical Association* 45(251): 311–334. doi:10.2307/2280290.

Fabiani, M., Accorsi, S., Lukwiya, M., Rosolen, T., Ayella, E.O., Onek, P.A., and Declich, S. (2001). Trend in HIV-1 prevalence in an antenatal clinic in north Uganda and adjusted rates for the general population. *AIDS* 15(1): 97–103. doi:10.1097/00002030-200101050-00014.

Ford, K. and Hosegood, V. (2005). AIDS mortality and the mobility of children in KwaZulu Natal, South Africa. *Demography* 42(4): 757–768. doi:10.1353/dem.2005.0029.

Fylkesnes, K., Ndhlovu, Z., Dasumba, D., Musonda, R., and Sichone, M. (1998). Studying the dynamics of the HIV epidemic: Population-based data compared with sentinel surveillance in Zambia. *AIDS* 12(10): 1227–1234. doi:10.1097/00002030-199810000-00015.

Garenne, M.L., Tollman, S.M., Collinson, M.A., and Kahn, K. (2007). Fertility trends and net reproduction in Agincourt, rural South Africa, 1992-2004. *Scandanavian Journal of Public Health* 35(69): 68–76. doi:10.1080/14034950701355650.

Ghys, P.D., Brown, T., Grassly, N.C., Garnett, G., Stanecki, K.A., Stover, J., and Walker, N. (2004). The UNAIDS estimation and projection package: A software package to estimate and project national HIV epidemics. *Sexually Transmitted Infections* 80(Supplement 1): i5–i9. doi:10.1136/sti.2004.010199.

Ghys, P., Kufa, E., George, M., and for the UNAIDS Reference Group on Estimates, Modelling and Projections' ad hoc Working Group on interpreting trends in prevalence and incidence of HIV infection in countries with generalised epidemics (2006). Measuring trends in prevalence and incidence of HIV infection in countries with generalised epidemics. *Sexually Transmitted Infections* 82(Supplement 1): i52–i56. doi:10.1136/sti.2005.016428.

Gray, R.H., Wawer, M.J., Serwadda, D., Sewankambo, N., Li, C., Wabwire-Mangen, F., Paxton, L., Kiwanuka, N., Kigozi, G., Konde-Lule, J., Quinn, T.C., and Gaydos, C.A. (1998). Population-based study of fertility in women with HIV-1 infection in Uganda. *Lancet* 351(9096): 98–103. doi:10.1016/S0140-6736(97)09381-1.

Gregson, S. and Garnett, G.P. (2000). Contrasting gender differentials in HIV-1 prevalence and associated mortality increase in eastern and southern Africa: Artefact of data or natural course of epidemics? *AIDS* 14(Supplement 3): S85–S99.

Gregson, S., Nyamukapa, C., Lopman, B., Mushati, P., Garnett, G.P., Chandiwanat, S.K., and Anderson, R.M. (2007). Critique of early models of the demographic impact of HIV/AIDS in sub-Saharan Africa based on contemporary empirical data from Zimbabwe. *Proceedings of the National Academy of Sciences of the United States of America* 104(37): 14586 – 14591. doi:10.1073/pnas.0611540104.

Gregson, S., Terceira, N., Kakowa, M., Mason, P.R., Anderson, R.M., Chandiwana, S.K., and Caraël, M. (2002). Study of bias in antenatal clinic HIV-1 surveillance data in a high contraceptive prevalence population in sub-Saharan Africa. *AIDS* 16(4): 643–652. doi:10.1097/00002030-200203080-00017.

Grosskurth, H., Mosha, F., Todd, J., Senkoro, K., Newell, J., Lkokke, A., Changalucha, J., West, B., Mayaud, P., Gavyole, A., Gabone, R., Mabey, D., and Hayes, R. (1995). A community trial of the impact of improved sexually transmitted disease treatment on the HIV epidemic in rural Tanzania. *AIDS* 9(8): 927–934. doi:10.1097/00002030-199508000-00015.

Hallett, T.B., Żaba, B., Todd, J., Lopman, B., Mwita, W., Biraro, S., Gregson, S., Boerman, J.T., and on behalf of the ALPHA Network (2008). Estimating incidence from prevalence in generalised HIV epidemics: Methods and validation. *PLoS Medicine* 5(4): 611–622. doi:10.1371/journal.pmed.0050080.

Hawken, M.P., Melis, R.D.J., Ngombo, D.T., Mandaliya, K.N., Ng'ang'a, L.W., Price, J., Dallabetta, G., and Temmerman, M. (2002). Opportunity for prevention of HIV and sexually transmitted infections in Kenyan youth: Results of a population-based survey. *Journal of Acquired Immune Deficiency Syndrome* 31(5): 529–535. doi:10.1097/00126334-200212150-00011.

Heuveline, P. (2003). HIV and population dynamics: A general model and maximum-likelihood standards for east Africa. *Demography* 40(2): 217–245. doi:10.1353/dem.2003.0013.

Hira, S.K., Kamanga, J., Bhat, G.J., Mwale, C., Tembo, G., Luo, N., and Perine, P.L. (1989). Perinatal transmission of HIV-1 in Zambia. *British Medical Journal* 299(6718): 1250–1252. doi:10.1136/bmj.299.6710.1250.

Hunter, S., Isingo, R., Boerma, J.T., Urassa, M., Mwaluko, G.M.P., and Żaba, B. (2003). The association between HIV and fertility in a cohort study in rural Tanzania. *Journal of Biosocial Science* 35(2): 189–199. doi:10.1017/S0021932003001895.

Institut National de la Statistiqu du Rwanda (INSR) and ORC Macro (2006). *Rwanda Demographic and Health Survey 2005*. Calverton, Maryland, USA: INSR and ORC Macro (447 pp.).

Institut National de la Statistique (INS) and ORC Macro (2004). *Enquête Démographique et de Santé du Cameroun 2004*. Calverton, Maryland, USA: INS and ORC Macro.

Isingo, R., Żaba, B., Marston, M., Ndege, M., Mngara, J., Mwita, W., Wringe, A., Beckles, D., Changalucha, J., and Urassa, M. (2007). Survival after HIV infection and the pre-antiretroviral therapy era in a rural Tanzanian cohort. *AIDS* 21(Supplement 6): S5–S13. doi:10.1097/01.aids.0000299405.06658.a8.

Jackson, J.B., Musoke, P., Fleming, T., Guay, L.A., Bagenda, D., Allen, M., Nakabiito, C., Sherman, J., Bakaki, P., Owor, M., Ducar, C., Deseyve, M., Mwatha, A., Emel, L., Duefield, C., Mirochnick, M., Fowler, M.G., Mofenson, L., Miotti, P., Gigliotti, M., Bray, D., and Mmiro, F. (2003). Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of moether-to-child transmission of HIV-1 in Kampala, Uganda: 18-month follow-up of the HIVNET 012 randomised trial. *Lancet* 362(9387): 859–868. doi:10.1016/S0140-6736(03)14341-3.

Kagaayi, J., Gray, R.H., Brahmbhatt, H., Kigozi, G., Galugoda, F., Wabwier-Mangen, F., Serwadda, D., Sewankambo, N., Ddungu, V., Ssebagala, D., Sekasanvu, J., Kigozi, G., Makumbi, F., Kiwanuka, N., Lutalo, T., Reynolds, S.J., and Wawer, M.J. (2008). Survival of infants born to HIV-positive mothers, by feeding modality, in Rakai, Uganda. *PLoS ONE* 3(12): e3877. doi:10.1371/journal.pone.0003877.

Kahn, K., Tollman, S., Collinson, M.A., Clark, S.J., Twine, R., Clark, B.D., Shabangu, M., Gomez-Olive, F.X., Mokoena, O., and Garenne, M. (2007). Research into health, population, and social transitions in rural South Africa: Data and methods of the Agincourt Health and Demographic Surveillance System. *Scandanavian Journal of Public Health* 35(69): 8–20. doi:10.1080/14034950701505031.

Kengeya-Kayondo, J.F., Kamali, A., Nunn, A.J., Ruberantwari, A., Wagner, H.U.H., and Mulder, D.W. (1996). Incidence of HIV-1 infection in adults and socio-demographic characteristics of seroconverters in a rural population in Uganda: 1990-1994. *International Journal of Epidemiology* 25(5): 1077–1082. doi:10.1093/ije/25.5.1077.

Keyfitz, N. and Caswell, H. (2005). *Applied Mathematical Demography*. Springer.

Kilian, A.H.D., Gregson, S., Ndyanabangi, B., Walusaga, K., Kipp, W., Sahlmüller, G., Garnett, G.P., Asiimwe-Okiror, G., Kabagambe, G., Weis, P., and von Sonnenburg, F. (1999). Reductions in risk behaviour provide the most consistent explanation of declining HIV prevalence in Uganda. *AIDS* 13(3): 391–398. doi:10.1097/00002030-199902250-00012.

Kwesigabo, G., Killewo, J., Urassa, W., Lugalla, J., Emmelin, M., Mutembei, A., Mhalu, F., Biberfeld, G., Wall, S., and Sandstrom, A. (2005). HIV-1 infection prevalence

and incidence trends in areas of contrasting levels of infection in the Kagera Region, Tanzania, 1987-2000. *Journal of Acquired Immune Deficiency Syndrome* 40(5): 585–591. doi:10.1097/01.qai.0000168183.22147.c4.

Leclerc, P.M. and Garenne, M. (2007). Inconsistencies in age profiles of HIV prevalence: A dynamic model applied to Zambia. *Demographic Research* 16(5): 121–140. doi:10.4054/DemRes.2007.16.5.

Lee, L.M., Wortley, P.M., Fleming, P.L., Eldred, L.J., and Gray, R.H. (2000). Duration of human immunodeficiency virus infection and likelihood of giving birth in a Medicaid population in Maryland. *American Journal of Epidemiology* 151(10): 1020–1028.

Leslie, P.H. (1945). On the use of matrices in certain population mathematics. *Biometrika* 33(3): 183–212. doi:10.1093/biomet/33.3.183.

Lewis, J.J.C., Ronsmans, C., Ezeh, A., and Gregson, S. (2004). The population impact of HIV on fertility in sub-Saharan Africa. *AIDS* 18: S35–S43. doi:10.1097/00002030-200406002-00005.

Liberia Institute of Statistics and Geo-Information Services (LISGIS) [Liberia], Ministry of Health and Social Welfare [Liberia], National AIDS Control Program [Liberia], and Macro International Inc. (2008). *Liberia Demographic and Health Survey 2007*. Monrovia, Liberi: Liberia Institute of Statistics and Geo-Information Services (LISGIS) and Macro International Inc.

Lutalo, T., Gray, R.H., Wawer, M., Sewankambo, N., Serwadda, D., Laeyendecker, O., Kiwanuka, N., Nalugoda, F., Kigozi, G., Ndyanabo, A., Bwanika, J.B., Reynolds, S.J., Quinn, T.C., and Opendi, P. (2007). Survival of HIV-infected treatment-naive individuals with documented dates of seroconversion in Rakai, Uganda. *AIDS* 21(Supplement 6): S15–S19. doi:10.1097/01.aids.0000299406.44775.de.

Mandelbrot, L., Msellati, P., Meda, N., Leroy, V., Likikouët, R., Van de Perre, P., Dequae-Merchadoux, L., Sylla-Koko, F., Ouangre, A., Ouassa, T., Ramon, R., Gautier-Charpentier, L., Cartoux, M., Dosso, M., Dabis, F., and Welffens-Ekra, C. (2002). 15 month follow up of African children following vaginal cleansing with benzalkonium chloride of their HIV infected mothers during late pregnancy and delivery. *Sexually Transmitted Infections* 78(4): 267–270. doi:10.1136/sti.78.4.267.

Marston, M., Żaba, B., Salomon, J.A., Brahmbhatt, H., and Bagenda, D. (2005). Estimating the net effect of HIV on child mortality in African populations affected by generalized HIV epidemics. *Journal of Acquired Immune Deficiency Syndrome* 38(2): 219–227. doi:10.1097/00126334-200502010-00015.

McGrath, N., Kranzer, K., Saul, J., Ctrampin, A.C., Malema, S., Kachiwanda, L., Żaba,

B., Jahn, A., Fine, P.E.M., and Glynn, J.R. (2007). Estimating the need for antiretroviral treatment and an assessment of a simplified HIV/AIDS case definition in rural Malawi. *AIDS* 21: S105–S113. doi:10.1097/01.aids.0000299417.69432.65.

Ministrère du Plan et Macro International (2008). *Enquête Démographique et de Santé, Rèpublique Démographique du Congo 2007*. Calverton, Maryland, USA: Ministrère du Plan et ORC Macro.

Ministry of Health and Social Welfare (MOHSW) [Lesotho], Bureau of Statistics (BOS)[Lesotho], and ORC Macro. (2005). *Lesotho Demographic and Health Survey 2004*. Calverton, Maryland: MOH, BOS, and ORC Macro.

Ministry of Health (MOH) [Uganda] and ORC Macro (2006). *Uganda HIV/AIDS Sero-behavioural Survey 2004-2005*. Calverton, Maryland, USA: Ministry of Health and ORC Macro (194 pp.).

Miotti, P.G., Taha, T.E.T., Kumwenda, N.I., Braodhead, R., Mtimavalye, L.A.R., der Ho-even, L.V., Chiphangwi, J.D., Limoba, G., and Biggar, R.J. (1999). HIV transmission through breastfeeding: A study in Malawi. *Journal of the American Medical Association* 282(8): 744–749. doi:10.1001/jama.282.8.744.

Morgan, D., Mahe, C., Mayanja, B., Okongo, J.M., Lubega, R., and Whitworth, J.A.G. (2002). HIV-1 infection in rural Africa: Is there a difference in median time to AIDS and survival compared with that in industrialized countries? *AIDS* 16(4): 597–603.

Murray, J., Sonnenberg, P., Nelson, G., Bester, A., Shearer, S., and Glynn, J.R. (2007). Cause of death and presence of respiratory disease at the autopsy in an HIV-1 serconversion cohort of southern African gold miners. *AIDS* 21(Supplement 6): S97–S104. doi:10.1097/01.aids.0000299416.61808.24.

Mwaluko, G., Urassa, M., Isingo, R., Żaba, B., and Boerma, J.T. (2003). Trends in HIV and sexual behaviour in a longitudinal study in a rural population population in Tanzania, 1994-2000. *AIDS* 17(18): 2645–2651. doi:10.1097/00002030-200312050-00012.

National Statistical Office (NSO) [Malawi] and ORC Macro (2005). *Malawi Demographic and Health Survey 2004*. Calverton, Maryland, USA: NSO and ORC Macro (454 pp.).

Nduati, R., John, G., Mbori-Ngacha, D., Richardson, B., Overbaugh, J., Mwatha, A., Ndinya-Achola, J., Bwayo, J., Onyango, F.E., Hughes, J., and Kreiss, J. (2000). Effect of breastfeeding and formula feeding on transmission of HIV-1: A randomized clinical trial. *Journal of the American Medical Association* 283(9): 1167–1174. doi:10.1001/jama.283.9.1167.

Newell, M., Coovadia, H., Cortina-Borja, M., Rollins, N., Gaillard, P., and Dabis, F. (2004). Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: A pooled analysis. *Lancet* 364(9441): 1236–1243. doi:10.1016/S0140-6736(04)17140-7.

Nguyen, R.H.N., Gange, S.J., Wabwire-Mangen, F., Sewankambo, N.K., Serwadda, D., Wawer, M.J., Quinn, T.C., and Gray, R.H. (2006). Reduced fertility among HIV-infected women associated with viral load in Rakai district, Uganda. *International Journal of STD and AIDS* 17(12): 842–846. doi:10.1258/095646206779307586.

Nicoll, A., Newell, M., Peckham, C., Luo, C., and Savage, F. (2000). Infant feeding and HIV-1 infection. *AIDS* 14(Supplement 3): S57–S74.

Nunn, A.J., Mulder, D.W., Kamali, A., Ruberantwari, A., Kengeya-Kayondo, J.F., and Whitworth, J. (1997). Mortality associated with HIV-1 infection over five years in a rural Ugandan population: Cohort study. *British Medical Journal* 315(7111): 767–771.

Nunn, A.J., Kengeya-Kayondo, J.F., Malamba, S.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(1): 81–86. doi:10.1097/00002030-199401000-00012.

Nyirenda, M., Hosegood, V., Bärnighausen, T., and Newell, M. (2007). Mortality levels and trends by HIV serostatus in rural South Africa. *AIDS* 21(Supplement 6): S73–S79. doi:10.1097/01.aids.0000299413.82893.2b.

Pearl, R. and Reed, L.J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Science* 6(6): 275–288. doi:10.1073/pnas.6.6.275.

Peters, P.J., Karita, E., Kayitenkore, K., Meinzen-Derr, J., Kim, D., Tichacek, A., and Allen, S.A. (2007). HIV-infected Rwandan women have a high frequency of long-term survival. *AIDS* 21(Supplement 6): S31–S37. doi:10.1097/01.aids.0000299408.52399.e1.

Petra Study Team (2002). Efficacy of three short-course regimens of zidovudine and lamivudine in preventing early and late transmission of HIV-1 from mother to child in Tanzania, South Africa, and Uganda (Petra Study): A randomised, double-blind, placebo-controlled trial. *Lancet* 359: 1178–1186.

Poole, D. and Raftery, A.E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association* 95(452): 1244–1255. doi:10.2307/2669764.

Porter, K. and Żaba, B. (2004). The empirical evidence for the impact of HIV on adult

mortality in the developing world: Data from serological studies. *AIDS* 18(Supplement 2): S9–S17.

Pritchett, H.S. (1891). A formula for predicting the population of the United States. *Publications of the American Statistical Association* 2(14): 278–286. doi:10.2307/2276575.

R Development Core Team (2011). The r project for statistical computing. [electronic resource]. Vienna, Austria: R Foundation for Statistical Computing. Http://www.r-project.org.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology* 25: 111–163. doi:10.2307/271063.

Read, J.S. and Breastfeeding and HIV International Transmission Study Group (2004). Late postnatal transmission of HIV-1 in breast-fed children: An individual patient data meta-analysis. *Journal of Infectious Diseases* 189(12): 2154–2166. doi:10.1086/420834.

Saidel, T., Sokal, D., Rice, J., Buzingo, T., and Hassig., S. (1996). Validation of a method to estimate age-specific human immunodeficiency virus (HIV) incidence in developing countries using population-based seroprevalence data. *American Journal of Epidemiology* 144(3): 214–223.

Sedgh, G., Larsen, U., Spiegelman, D., Msamanga, G., and Fawzi, W.W. (2005). HIV-1 disease progression and fertility in Dar es Salaam, Tanzania. *Journal of Acquired Immune Deficiency Syndrome* 39(4): 439–445. doi:10.1097/01.qai.0000148529.58963.83.

Serwadda, D., Wawer, M.J., Musgrave., S.D., Sewankambo, N.K., Kaplan, J.E., and Gray, R.H. (1992). HIV risk factors in three geographic strata of rural Rakai District, Uganda. *AIDS* 6(9): 983–989. doi:10.1097/00002030-199209000-00012.

Sewankambo, N.K., Wawer, M.J., Gray, R.H., Serwadda, D., Li, C., Stallings, R.Y., Musgrave, S.D., and Konde-Lule, J. (1994). Demographic impact of HIV infection in rural Rakai District, Uganda: Results of a population-based cohort study. *AIDS* 8(12): 1707–1713.

Sewankambo, N.K., Gray, R.H., Ahmad, S., Serwadda, D., Wabwire-Mangen, F., Nalugoda, F., Kiwanuka, N., Lutalo, T., Kigozi, G., Li, C., Meehan, M.P., Brahmbatt, H., and Wawer, M.J. (2000). Mortality associated with HIV infection in rural Rakai District, Uganda. *AIDS* 14(15): 2391–2400. doi:10.1097/00002030-200010200-00021.

Shafer, L.A., Biraro, S., Nakiyingi-Miiro, J., Kamali, A., Ssematimba, D., Ouma, J., Ojwiya, A., Hughes, P., Van der Paal, L., Whitworth, J.A.G., Opio, A.A., and

Grosskurth, H. (2008). HIV prevalence and incidence are no longer falling in southwest Uganda: Evidence from a rural population cohort 1989-2005. *AIDS* 22(13): 1641–1649. doi:10.1097/QAD.0b013e32830a7502.

Shao, J., Burbaker, G., Levin, A., Kibauri, A., Massesa, E., Siso, Z., Konings, E., Clayton, Y., Kumby, D., Alexander, S., Waters, D., Drummond, J., Biggar, R., Scott, G., Miller, G., Goedert, J.J., and Blattner, W.A. (1994). Population-based study of HIV-1 infection in 4,086 subjects in northwest Tanzania. *Journal of Acquired Immune Deficiency Syndrome* 7(4): 397–402.

Simonon, A., Lepage, P., Karita, E., Hitimana, D., Dabis, F., Msellati, P., Van Goethem, C., Nsengumuremyi, F., Bazubagira, A., and Van de Perre, P. (1994). An assessment of the timing of mother-to-child-transmission of human immunodeficiency virus type 1 by means of polymerase chain reaction. *Journal of Acquired Immune Deficiency Syndrome* 7(9): 952–957.

Slutsker, L., Cabeza, J., Wirima, J.J., and Steketee, R.W. (1994). HIV-1 infection among women of reproductive age in a rural district in Malawi. *AIDS* 8(9): 1337–1340. doi:10.1097/00002030-199409000-00019.

Smith, J., Mushati, P., Kurwa, F., Mason, P.R., Gregson, S., and Lopman, B. (2007). Changing patterns of adult mortality as the HIV epidemic matures in Manicaland, eastern Zimbabwe. *AIDS* 21(Supplement 6): S81–S86. doi:10.1097/01.aids.0000299414.60022.1b.

Smith, S.K. and Sincich, T. (1988). Stability over time in the distribution of population forecast errors. *Demography* 25(3): 461–474. doi:10.2307/2061544.

Sokal, D.C., Buzingo, T., Nitunga, N., Kadende, P., and Standaert, B. (1993). Geographic and temporal stability of HIV seroprevalence among pregnant women in Bujumbura, Burundi. *AIDS* 7(11): 1481–1484. doi:10.1097/00002030-199311000-00012.

Spira, R., Lepage, P., Msellati, P., Van de Perre, P., Leroy, V., Simonon, A., Karita, E., and Dabis, F. (1999). Natural history of human immunodeficiency virus type 1 infection in children: A five-year prospective study in Rwanda. *Pediatrics* 104(5): e56. doi:10.1542/peds.104.5.e56.

Stover, J., Walker, N., Grassly, N.C., and Marston, M. (2006). Projecting the demographic impact of AIDS and the number of people in need of treatment: Updates to the spectrum projection package. *Sexually Transmitted Infections* 82(Supplement 3): iii45–iii50. doi:10.1136/sti.2006.020172.

Taha, T.E., Dallabetta, G.A., Hoover, D.R., Dhiphangwi, J.D., Mtimavalye, L.A.R., Liomba, G.N., Kumwenda, N.I., and Miotti, P.G. (1998). Trends of HIV-1 and sexually

transmitted diseases among pregnant and postpartum women in urban Malawi. *AIDS* 12(2): 197–203. doi:10.1097/00002030-199802000-00010.

Tanzania Commission for AIDS (TACAIDS), National Bureau of Statistics (NBS) and ORC Macro (2005). *Tanzania HIV/AIDS Indicator Survey 2003-2004*. Calverton, Maryland, USA: TACAIDS and ORC Macro (132 pp.).

Tanzania Commission for AIDS (TACAIDS), Zanzibar AIDS Commission (ZAC), National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and Macro International Inc. (2008). *Tanzania HIV/AIDS and Malaria Indicator Survey 2007-2008*. Dar es Salaam, Tanzania: TACAIDS, ZAC, NBS, OCGS, and Macro International Inc. (272 pp.).

Terceira, N., Gregson, S., Żaba, B., and Mason, P.R. (2003). The contribution of HIV to fertility decline in rural Zimbabwe, 1985-2000. *Population Studies* 57(2): 149–164. doi:10.1080/0032472032000097074.

Timaeus, I.M. and Jasseh, M. (2004). Adult mortality in sub-Saharan Africa: Evidence from demographic and health surveys. *Demography* 41(4): 757–772. doi:10.1353/dem.2004.0037.

Todd, J., Balira, R., Grosskurth, H., Mayaud, P., Mosha, F., ka-Gina, G., Klokke, A., Gabbone, R., Gavyole, A., Mabey, D., and Hayes, R. (1997). HIV-associated adult mortality in a rural Tanzanian population. *AIDS* 11(6): 801–807. doi:10.1097/00002030-199706000-00013.

UNAIDS (2008). *Report on the global AIDS epidemic 2008*. Geneva: UNAIDS.

UNAIDS Reference Group on Estimates, Modelling and Projections (2002). Improved methods and assumptions for estimation of the HIV/AIDS epidemic and its impact: Recommendations of the UNAIDS Reference Group on Estimates, Modelling and Projections. *AIDS* 16(9): W1–W14. doi:10.1097/00002030-200206140-00024.

United Nations (1982). *Model Life Tables for Developing Countries*. New York: United Nations.

United Nations (1998). Aids, mortality, and population change – technical meeting on the demographic impact of HIV/AIDS. New York: United Nations. Http://www.un.org/esa/population/pubsarchive/hivmtg/aidsrep.pdf (February 13, 2008).

United Nations (1999). *World Population Prospects: The 1998 Revision*. New York: United Nations.

United Nations (2004). *World Population Prospects: The 2002 Revision*. New York:

United Nations.

Urassa, M., Kumogola, Y., Isingo, R., Mwaluko, G., Makelemo, B., Mugeye, K., Boerma, T., Calleja, T., Slaymaker, E., and Żaba, B. (2006). HIV prevalence and sexual behaviour changes measured in an antenatal clinic setting in northern Tanzania. *Sexually Transmitted Infections* 82(4): 301–306. doi:10.1136/sti.2005.016766.

Urassa, M., Boerma, J.T., Isingo, R., Ngalula, J., Ng'weshemi, J., Mwaluko, G., and Żaba, B. (2001). The impact of HIV/AIDS on mortality and household mobility in rural Tanzania. *AIDS* 15(15): 2017–2023. doi:10.1097/00002030-200110190-00015.

Van der Paal, L., Shafer, L.A., Todd, J., Mayanja, B.N., Whitworth, J.A.G., and Grosskurth, H. (2007). HIV-1 disease progression and mortality before the introduction of highly active antiretroviral therapy in rural Uganda. *AIDS* 21(Supplement 6): S21–S29. doi:10.1097/01.aids.0000299407.52399.05.

Wachter, K.W., Knodel, J.E., and Van Landingham, M. (2002). AIDS and the elderly in Thailand: Projecting familial impacts. *Demography* 39(1): 25–41. doi:10.1353/dem.2002.0012.

Wambura, M., Urassa, M., Isingo, R., Ndege, M., Marston, M., Slaymaker, E., Mngara, J., Changalucha, J., Boerma, T.J., and Żaba, B. (2007). HIV prevalence and incidence in rural Tanzania: Results from 10 years of follow-up in an open-cohort study. *Journal of Acquired Immune Deficiency Syndrome* 46(5): 616–623. doi:10.1097/QAI.0b013e31815a571a.

Wawer, M.J., Serwadda, D., Musgrave, S.D., Konde-Lule, J., Musagara, M., and Sewankambo., N.K. (1991). Dynamics of spread of HIV-1 infection in a rural district of Uganda. *British Medical Journal* 303(6813): 1303–1306. doi:10.1136/bmj.303.6813.1303.

Wawer, M.J., Sewankambo, N.K., Berkley, S., Serwadda, D., Musgrave, S.D., Gray, R.H., Musagara, M., Stallings, R.Y., and Konde-Lule, J.K. (1994). Incidence of HIV-1 infection in a rural region of Uganda. *British Medical Journal* 308(6922): 171–173.

Welz, T., Hosegood, V., Jaffar, S., Batzing-Feigenbaum, J., Herbst, K., and Newell, M. (2007). Continued very high prevalence on HIV infection in rural Kwazulu-Natal, South Africa: A population-based longitudinal study. *AIDS* 21(11): 1467–1472. doi:10.1097/QAD.0b013e3280ef6af2.

Whelpton, P.K. (1936). An empirical method of calculating future population. *Journal of the American Statistical Association* 31(195): 457–473. doi:10.2307/2278370.

Żaba, B., Bärnighausen, T., Biraro, S., Bwanika, J.B., Gregson, S., Hallett, T.B., Hose-

good, V., Isingo, R., Lutalo, T., Marston, M., Mushati, P., Mwita, W., Ndyanabo, A., Shafer, L.A., Todd, J., Nyirenda, M., and Wringe, A. (2008). *Diverse age patterns of HIV incidence rates in Africa*. Paper presented at the XVII International AIDS conference, Mexico City, August 3-8, 2008.

Żaba, B., Carpenter, L.M., Boerma, J.T., Gregson, S., Nakiyingi, J.S., and Urassa, M. (2000). Adjusting ante-natal clinic data for improved estimates of HIV prevalence among women in sub-Saharan Africa. *AIDS* 14(17): 2741–2750. doi:10.1097/00002030-200012010-00014.

Żaba, B. and Gregson, S. (1998). Measuring the impact of HIV on fertility in Africa. *AIDS* 12(1): S41–S50.

Żaba, B., Marston, M., Crampin, A.C., Isingo, R., Biraro, S., Bärnighausen, T., Lopman, B., Lutalo, T., Glynn, J.R., and Todd, J. (2007). Age-specific mortality patterns in HIV-infected individuals: A comparative analysis of African community study data. *AIDS* 21(Supplement 6): S87–S96. doi:10.1097/01.aids.0000299415.67646.26.

Zijenah, L.S., Moulton, L.H., Iliff, P., Nathoo, K., Munjoma, M.W., Mutasa, K., Malaba, L., Zvandasara, P., Ward, B.J., Humphrey, J., and ZVITAMBO Study Group (2004). Timing of mother-to-child transmission of HIV-1 and infant mortality in the first 6 months of life in Harare, Zimbabwe. *AIDS* 18(2): 273–280. doi:10.1097/00002030-200401230-00017.