A model of the impact of HIV/AIDS and antiretroviral treatment in the Masiphumelele community

Centre for Infectious Disease Epidemiology and Research working paper
June 2011

Leigh F. Johnson[1]
Katharina Kranzer[2,3]
Keren Middelkoop[2,4]
Robin Wood[2]

1. Centre for Infectious Disease Epidemiology and Research, University of Cape Town
2. The Desmond Tutu HIV Centre, Institute of Infectious Disease and Molecular Medicine, University of Cape Town
3. Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine
4. Department of Medicine, University of Cape Town

Correspondence to:
Leigh Johnson
Centre for Infectious Disease Epidemiology and Research
University of Cape Town
Faculty of Health Sciences
Anzio Road
Observatory
7925

Tel: +27 21 406 6981
Fax: +27 21 406 6764
Email: Leigh.Johnson@uct.ac.za

Abstract

Background: Guidelines for the initiation of antiretroviral treatment (ART) in developing countries have been revised to recommend ART initiation in all adults with CD4 counts of <350/μl, higher than the threshold of 200/μl that was previously recommended. However, there is uncertainty regarding the impact that this change in guideline will have on the numbers of adult patients starting ART. We aim to quantify the impact of this change using data from Masiphumelele, an informal settlement in Cape Town.
Method: A mathematical model was developed to simulate demographic changes in the Masiphumelele community over time, age- and sex-specific HIV incidence rates, and CD4 decline in untreated HIV-positive adults. Rates of ART initiation were estimated from numbers of patients starting ART in Masiphumelele, stratified by year of ART initiation, sex, baseline CD4 count and ART experience (patients restarting ART or transferring into the Masiphumelele ART programme from elsewhere were considered separately).
Results: During the period 2005-6, when ART was available to all adults in Masiphumelele with CD4 counts below 350/μl, rates of ART initiation in males were substantially higher in those with CD4 <200 (74.3 per 100 ART-eligible person years (AEPY)) than in those with CD4 counts of 200-349 (8.8/100 AEPY). Over the same period, rates of ART initiation were substantially higher in women with CD4 counts <200 (103.6/100 AEPY) than in women with CD4 counts of 200-349 (22.5/100 AEPY). In the 2007-9 period, when ART initiation criteria reverted to those used nationally (CD4 <200 or WHO stage IV), the rate of ART initiation in adults with CD4 <200 increased, but the rate of ART initiation in adults with CD4 200-349 declined.
Conclusions: Although recent changes to ART initiation criteria imply a substantial increase in the numbers of adults who are ART-eligible, these results suggest that the actual increase in the number of adults starting ART is likely to be relatively modest. Improvements in access to HIV testing, as well as follow-up of HIV-diagnosed patients, will be needed in order to ensure greater ART initiation in the CD4 200-349 category.

Background

Many countries in sub-Saharan Africa continue to experience major challenges in rolling out antiretroviral treatment (ART) to the population that is eligible for treatment. Recent changes to ART guidelines have increased substantially the proportion of the HIV-positive population that is eligible to receive ART, with new guidelines recommending that ART be initiated in all adults who have CD4 counts below 350 cells/μl (World Health Organization 2009). At the same time, however, donor commitment to funding ART in developing countries has begun to level off. Realistic projections of the likely future growth in treatment numbers are needed in order for countries to plan for the financial and human resource requirements associated with the ART programme, particularly in the light of the new treatment guidelines. However, projection of the future uptake of ART is hampered by uncertainty regarding the likely rates of ART initiation in individuals who were previously considered ineligible, i.e. asymptomatically infected individuals with CD4 counts between 200 and 350. These individuals are less likely to know their HIV status than symptomatic individuals with CD4 counts below 200, and they are therefore less likely to initiate ART. Projections of the future ART uptake are also hampered by uncertainty regarding rates at which individuals re-initiate ART after stopping therapy. This could potentially be very significant, considering the high rates of loss to follow-up that have been observed in many ART cohorts in developing countries (Fox and Rosen 2010).

Data collected from the Masiphumelele community in South Africa provide a unique opportunity to address these uncertainties. During 2005 and 2006, this community was the focus of an intervention to start ART at higher CD4 counts than were recommended by the South African Department of Health at the time, at CD4 counts less than 350 (Sanne *et al*, 2010). Following 2006, however, the intervention was discontinued, and the criteria for starting ART reverted to those used nationally, i.e. CD4 counts less than 200 or WHO clinical stage IV. The data from the ART programme in this community therefore provide valuable information on the relative numbers of patients starting ART at different CD4 counts, under the old and the new treatment guidelines. However, the interpretation of these relative numbers is complex because of the uncertainty regarding the population distribution of CD4 counts in untreated individuals, which is likely to change substantially over time, both as a result of the natural evolution of the HIV epidemic, and as a result of the ART programme. In order to estimate the relative *rates* of ART initiation in the different untreated patient categories, it is necessary to use mathematical modelling to estimate the change over time in the numbers of untreated HIV-positive individuals in different categories.

The aim of this paper is therefore to estimate the relative rates of ART initiation in different CD4 categories, as well as the relative rates of ART initiation in ART-naïve and previously treated individuals, using a mathematical model of the Masiphumelele community. To ensure that the denominators are estimated with a reasonable degree of accuracy, the model is calibrated to demographic data collected from the Masiphumelele community, as well as HIV prevalence data and data on CD4 distributions in untreated individuals.

Model description

Demographic assumptions

Regular household censuses have been conducted in the Masiphumelele community since 1996, and this has provided information on the numbers of males and females in each five-year age group in 1996, 2002, 2004, 2006, 2008 and 2010. Each of these censuses is assumed to provide information on the population size at the middle of the survey year. In the years in which censuses were not conducted, the numbers of males and females in each five-year age group are linearly interpolated. Prior to 1996, the population is linearly interpolated, from a level of zero in 1985 (at which time the Masiphumelele community did not exist) up to the level observed in the 1996 census. Although this assumption of linear growth is unrealistic, it has little effect on the HIV profile of the population in recent years, which is what we are primarily interested in.

Having obtained estimates of numbers of males and females in each five year age group, we then estimated numbers of males and females at each individual age using Beer's "ordinary" formula (Judson and Popoff 2004).

Of individuals aged x at the start of a particular year, who are of sex g and not receiving ART, a proportion $m_g(x)$ is assumed to migrate out of the Masiphumelele community during the year. These age-specific out-migration rates are based on estimates of the rate of migration from the Western Cape to the Eastern Cape among Africans, over the 2001-2007 period (Dorrington and Moultrie 2009). Most African migration in and out of the Western Cape is believed to be between the Eastern Cape and Western Cape provinces (Dorrington and Moultrie 2009), but to the extent that other provinces are excluded, and to the extent that intra-provincial out-migration is excluded, these estimates may understate the true rate of out-migration in Masiphumelele. It might also be expected that rates of out-migration in urban informal settlements may be higher than in more settled rural communities. We therefore adjust the age-specific rates of out-migration estimated by Dorrington and Moultrie by a constant multiple of 12 in male and 18 in females, in order to obtain the $m_g(x)$ values. These multiples were chosen in order to produce estimates of in-migration consistent with the reported rates of in-migration in the Masiphumelele community in a 2010 household survey (discussed in more detail below). Due to the lack of migration data from other periods, it was assumed that rates of age-specific out-migration remained constant over the 1985-2010 period. Rates of out-migration in individuals receiving ART are calculated from the numbers of ART patients who are known to have transferred to other ART services in each year.

Similarly, numbers of patients known to have transferred into the Masiphumelele ART programme in each year are used to determine the model assumptions about numbers of in-migrants who are known to be on ART at the time of entry into the population. To calculate the number on in-migrants at age x, who are not on ART, it is necessary to make use of the demographic balancing equation. Suppose that at time t, the number of individuals aged x in the population, of sex g and not on ART, is $N_g(x, t)$. Suppose that in this group, the probability of either dying or starting ART in the next 12 months is $d_g(x, t)$. Then the number of individuals who are aged $x + 1$ a year later, $N_g(x + 1, t + 1)$, can be computed using the formula

$$N_g(x+1,t+1) = N_g(x,t)\big(1-d_g(x,t)\big)\big(1-m_g(x)\big)+I_g(x,t), \qquad (1)$$

where $I_g(x, t)$ is the number of people migrating into the population, aged x at time t, who remained in the population until time $t + 1$. Since the values $N_g(x, t)$ and $N_g(x + 1, t + 1)$ are known (from the Beer's formula estimate of the total population, subtracting the model estimates of numbers of patients on ART), the terms of this equation can be rearranged to calculate the number of in-migrants. (The values of $d_g(x, t)$ are also calculated from the model, as described in subsequent sections.) From this it is apparent that the number of in-migrants is positively related to the out-migration rate, and the assumed out-migration rates have been set in such a way that the resulting numbers of in-migrants are consistent with the in-migration rates in a 2010 household survey (Figure 1). Overall, the model estimate of the proportion of 15-49 year old females who have moved into Masiphumelele in the last three years (23.9%) is consistent with the 2010 survey estimate (26.1%, 95% CI: 22.5-30.0%), and similar consistency is observed in males aged 15-49 (27.9% in the model, compared with 29.9% (95% CI: 26.1-34.0%) in the survey). However, in individuals aged 50 and older, the model estimates a significantly lower immigration rate than that measured in the survey, both in males and females. This suggests that the assumed rates of out-migration at the older ages may be too low, or that the rate of mortality in this older population may be under-estimated. In males aged 15-19, the model appears to significantly over-estimate the level of immigration observed in the survey. This implies that the assumed rates of out-migration at the younger ages may be too high in males.

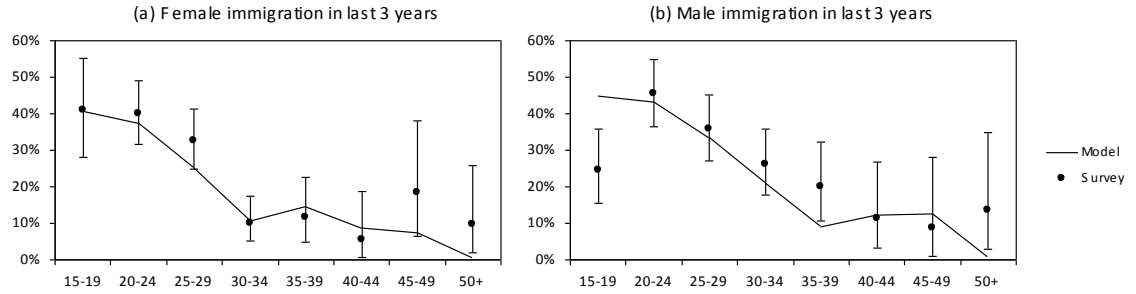

Figure 1: Proportion of the 2010 Masiphumelele population that has migrated into Masiphumelele in the last three years, by age and sex
Error bars surrounding survey estimates represent 95% confidence intervals.

Non-AIDS mortality rates are assumed to be the same as those estimated for Africans in the Western Cape, as obtained from the ASSA2008 AIDS and Demographic model (Actuarial Society of South Africa 2011). These rates differ by sex and individual age, and they are also assumed to reduce slightly from one calendar year to the next.

HIV incidence assumptions

HIV incidence rates, by age and by sex, were obtained from a previously described model of sexual behaviour patterns and HIV transmission patterns in South Africa (Johnson *et al*, 2009). These estimates are representative of the trends that would be expected at a national level. However, HIV incidence rates in Africans in the Western Cape appear to lag those in the country as a whole by approximately one year (Actuarial Society of South Africa 2011). We therefore applied a one-year offset to the national HIV incidence rates, when calculating the HIV incidence rates in Masiphumelele. The age- and sex-specific HIV incidence rates

were also adjusted by a constant factor of 1.35 in females, in order to bring the model estimates of HIV prevalence in Masiphumelele in line with the prevalence levels observed in surveys (Figure 2). These household surveys were conducted in 2005 (Wood *et al*, 2007) and in 2008, and are considered to provide a reasonably unbiased estimate of the actual prevalence in the population. No adjustment to male HIV incidence rates was necessary in order to achieve consistency between the model estimates of HIV prevalence in males and the survey estimates.

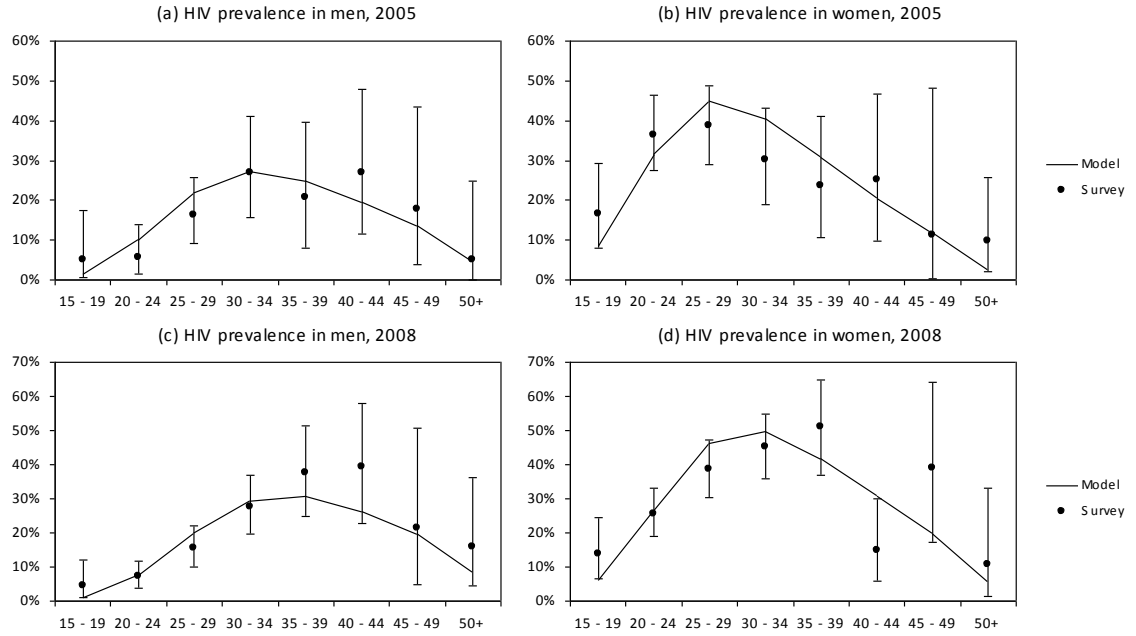

Figure 2: Model calibration to age-specific HIV prevalence data

In order to generate the HIV profile of immigrants into Masiphumelele, it is also necessary to make assumptions about HIV incidence rates in the population of potential immigrants into Masiphumelele. In reality, most immigrants into Masiphumelele are Africans from the Eastern Cape, and HIV incidence rates in this population group appear to peak at around the same time as HIV incidence rates in the general population (Actuarial Society of South Africa 2011), so that no time lag needs to be applied to the national HIV incidence rates. HIV prevalence in the population of potential immigrants is calculated for each age, sex and year, by applying the age- and year-specific HIV incidence rates to each single-year age cohort and assuming that infected individuals progress through a model of CD4 decline (described below) before dying from AIDS. This simple model does not allow for ART initiation, as the modelling of ART patients migrating into Masiphumelele is handled separately from the modelling of in-migrants who are not on ART. The HIV prevalence of female migrants into Masiphumelele is multiplied by a factor of 1.35, the same factor that is used to adjust HIV incidence rates after entry into the Masiphumelele community.

Survival of untreated adults

The survival of HIV-infected adults, in the absence of ART, is simulated using a multi-state model of HIV survival, represented in Figure 3. The parameters in this model are estimated from South African surveys of CD4 distributions in HIV-positive individuals, as described in the appendix. This diagram is a simplification of the model that is actually used, as

movements out of the population due to non-AIDS mortality are not shown. In addition, the age dependency of the transition rates is not shown. The parameters that are estimated in the appendix and shown in Figure 3 are assumed to apply to HIV-positive individuals who are aged 35, since this corresponds to the average age of individuals who become infected in their late twenties. For an individual aged *x*, the rate of transition out of state *s* is assumed to be

$$\lambda_s(x) = \lambda_s \times 1.015^{x-35},$$

where 1.015 is the factor by which the rate of progression increases for each one-year increase in the individual age. The factor of 1.015 is the factor by which the average HIV survival time (in the absence of ART) increases for each one-year *decrease* in the age at which HIV acquisition occurs, and is estimated by fitting Weibull models to age-specific survival data from the CASCADE Collaboration (Collaborative Group on AIDS Incubation and HIV Survival 2000). Since the rate of mortality is inversely proportional to the average survival time, the factor by which the rates of progression increase, for each one-year *increase* in age, can be crudely approximated as being of a similar magnitude (1.015).

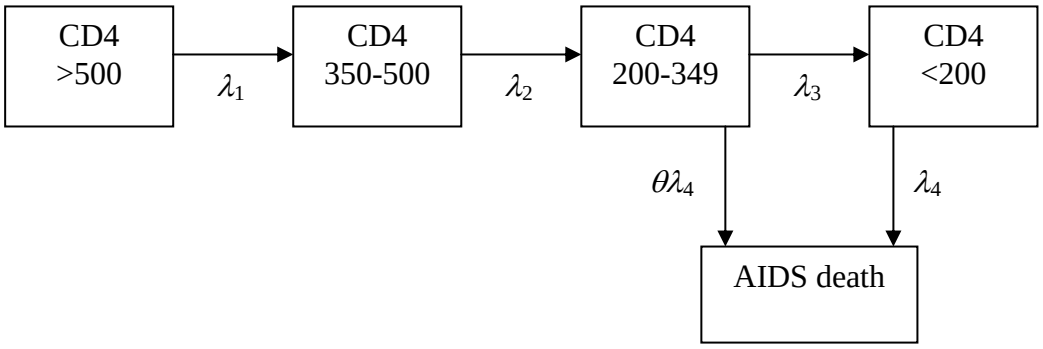

Figure 3: Multi-state model of decline in CD4 count in HIV-infected adults

Rates of ART initiation

Programme statistics, summarized in Table 1, provide information on the numbers of ART-naive adults starting ART, by sex, by year and by CD4 category. In addition, information is available on numbers of patients re-initiating ART after having temporarily stopped ART (also included in Table 1). The high proportions of ART-naive patients starting ART with CD4 >200, in 2005 and 2006, are a reflection of the change in ART initiation criteria over this period. However, even after 2006, when treatment initiation criteria reverted to those used nationally, the proportion of new ART patients with CD4 >200 remained substantial.

Table 1: Profile of patients initiating ART in Masiphumelele

	2004	2005	2006	2007	2008	2009
Males starting ART						
ART naive	35	64	88	77	81	83
Re-initiating ART	0	0	4	6	10	14
% of ART-naive with CD4 >200	7%	27%	23%	16%	26%	20%
Females starting ART						
ART naive	81	153	197	105	167	155
Re-initiating ART	0	1	4	10	24	11
% of ART-naive with CD4 >200	11%	38%	44%	32%	24%	31%

In order to use these numbers as model inputs it is necessary to express them as rates of ART initiation (numbers of patients starting ART per 100 patient years of untreated ART eligibility). This rate is calculated for each year, for males and females separately, and for each of three eligible groups: ART-naive adults with CD4 <200, ART-naive adults with CD4 200-350, and adults who have discontinued ART. However, the calculation of the denominator is complex because the number of patient years of untreated ART eligibility is affected by the numbers of patients starting ART (the numerator). Suppose that $r_{g,c}(t)$ is the rate of ART initiation in adults of sex *g*, in eligibility group *c* (the three eligibility groups are defined above) in year *t*. Further suppose that $S_{g,c}(t)$ is the corresponding number of adults starting ART (calculated from the information in Table 1). To calculate $r_{g,c}(t)$ in terms of $S_{g,c}(t)$, we need to know the model estimates of the numbers of individuals in eligibility category *c* at the start of year *t*, at each age *x*, $E_{g,c}(x,t)$, as well as the numbers progressing into eligibility category *c* over the course of year *t*, $P_{g,c}(x,t)$, and the rate of mortality that would be expected in the absence of ART in eligibility category *c*, $q_{g,c}(x)$. Then if it is assumed that the individuals who progress to eligibility category *c* do so uniformly over the course of the year, it can be shown that

$$S_{g,c}(t) = \sum_x E_{g,c}(x,t)\frac{r_{g,c}(t)}{r_{g,c}(t)+q_{g,c}(x)}\left[1-\exp\left(-r_{g,c}(t)-q_{g,c}(x)\right)\right]$$
$$+\sum_x P_{g,c}(x,t)\frac{r_{g,c}(t)}{r_{g,c}(t)+q_{g,c}(x)}\left\{1-\frac{1}{r_{g,c}(t)+q_{g,c}(x)}\left[1-\exp\left(-r_{g,c}(t)-q_{g,c}(x)\right)\right]\right\}$$

The rate of ART initiation, $r_{g,c}(t)$, is then calculated using Newton's method.

Survival after ART initiation

Figure 4 shows how the model is extended to incorporate survival after ART initiation. Individuals on ART are categorized according to the number of years since they started ART and according to their CD4 count at the time they initiated ART. Individuals on ART are assumed to stop ART either as a result of death or discontinuation due to other reasons. After discontinuation of ART, individuals can restart treatment, but for the sake of simplicity this is assumed to occur only in individuals who have CD4 <200.

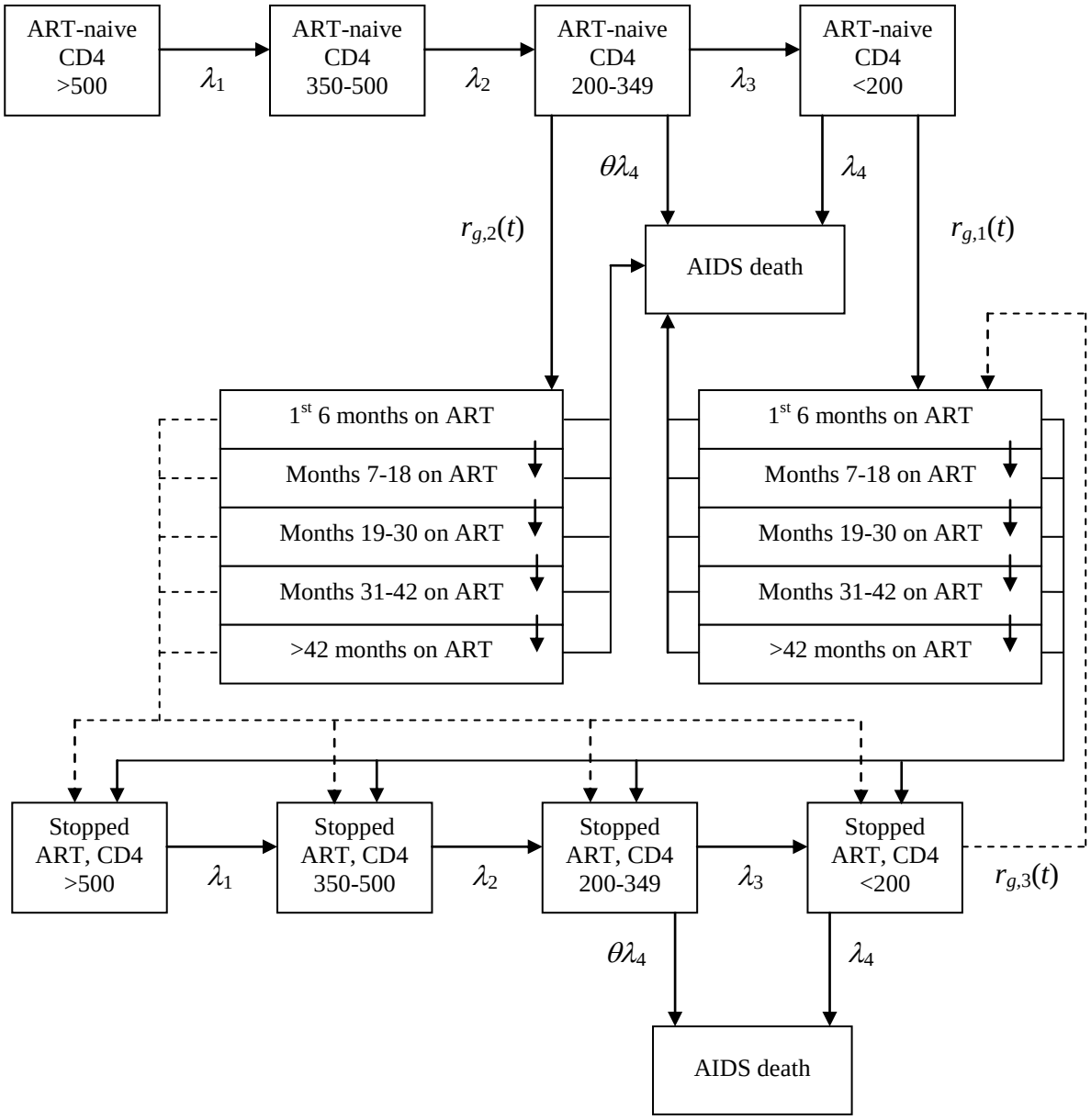

Figure 4: Multi-state model of survival after ART initiation
For the sake of simplicity, transitions due to non-HIV mortality and migration into and out of the population (including ART patients transferring in and out) are not shown.

Based on recorded numbers of deaths in the Masiphumelele ART programme, it is assumed that the annual AIDS mortality rate in patients starting ART with CD4 <200 is 6.6 per 100 person years during the first 6 months after ART initiation, 2 per 100 person years during months 7-18, and 1 per 100 person years thereafter. Mortality rates are assumed to be lower in those patients who start ART with CD4 counts of 200-349: 2 per 100 person years during the first 6 months, 0.8 per 100 person years during months 7-18, and 0.6 per 100 person years thereafter. This means that during the first 6 months after starting ART, the rate of mortality in patients with baseline CD4 ≥200 is assumed to be 0.3 times the mortality rate in patients with CD4 <200. This ratio is consistent with the ratio of 0.34 observed in a pooled analysis of

data from ART programmes across South Africa, comparing the same CD4 categories during the first 4 months after starting ART (Cornell *et al,* 2010). The above mortality rates are assumed to apply at age 35, the average age at which adults start ART in the South African context (Cornell *et al,* 2010). There are few studies in developing countries that have distinguished HIV mortality and non-HIV mortality after ART initiation, but studies that have done so in industrialized settings have shown a strong association between older age and higher HIV mortality (van Sighem *et al,* 2003; Braithwaite *et al,* 2005). Based on data from the ATHENA cohort, the HIV-related mortality rate is assumed to increase by a factor of 1.029 for each year of increase in age (van Sighem *et al,* 2003). For example, in adults who started ART with CD4 <200 and who have survived for more than 18 months on ART, the annual HIV mortality rate at age x is assumed to be $0.01 \times 1.029^{x-35}$.

Assumed rates of ART discontinuation (for reasons other than death) are also based on Masiphumelele estimates of numbers of patients lost to follow-up. Consistent with data from other South African cohorts (Nglazi *et al,* 2011; Fatti *et al,* 2010; Boulle *et al,* 2010; Cornell *et al,* 2010), loss to follow-up in Masiphumelele has become increasingly common in recent years, particularly during the first year after ART is initiated. During the first 6 months after starting ART, the annual rate of stopping ART is assumed to increase from 10.3 per 100 person years in 2004, to 13.7 per 100 person years in 2005-6, to 29.2 per 100 person years in 2007-8, to 40.3 per 100 person years thereafter. The annual rate of stopping ART after the first 6 months on ART is assumed to be 6 per 100 per person years, and this is assumed to remain constant over time. As with the mortality rates, these rates of stopping ART are assumed to apply at age 35, and age adjustment factors are applied to allow for the fact that ART discontinuation tends to be much more common at younger ages than at older ages (Cornell *et al,* 2010; Fatti *et al,* 2010). Based on South African data sources (Cornell *et al,* 2010; Fatti *et al,* 2010), the rate of stopping ART is assumed to be reduced by a factor of 0.989 for each one-year increase in the age of the patient on ART. South African studies have also shown that even after adjusting for age, rates of loss to follow-up are significantly higher in men than in women (Nglazi *et al,* 2011; Fatti *et al,* 2010). It is therefore assumed that rates of stopping ART are multiplied by a factor of 1.15 in males and 0.92 in females, so that the male:female ratio is consistent with the adjusted hazard ratio of 1.26 (95% CI: 1.05-1.51) estimated by Nglazi *et al* (2011). For example, in a woman of age x who has been on ART for at least 6 months, the rate of ART discontinuation is calculated as $0.06 \times 0.92 \times 0.989^{x-35}$.

For the purpose of calibrating the model to CD4 data, and for the purpose of determining the CD4 groups into which individuals move after stopping ART, it is necessary to make assumptions about changes in CD4 counts after ART initiation. These assumptions about changes in CD4 counts are based on published estimates of CD4 trajectories in different settings (Boulle *et al,* 2010; Lok *et al,* 2010; Nash *et al,* 2008). Boulle *et al* (2010) have reported medians and interquartile ranges of CD4 counts in patients on ART for different durations, in a South African ART programme in which almost all patients started ART with CD4 <200. Gamma distributions have been fitted to these data to determine (approximately) the mean CD4 and coefficient of variation in CD4 counts at each duration after ART initiation, in patients who start ART with CD4 <200. In patients who start ART at higher CD4 counts, the change in mean CD4 count is approximated from a collaborative analysis of CD4 changes in resource-limited settings (Nash *et al,* 2008). However, the latter analysis suggests a less substantial long-term increase in CD4 counts than that observed by Boulle *et al,* when the comparison is restricted to patients with baseline CD4 <200. This may be

because virological monitoring (which is not routine in most resource-limited settings) is having a pronounced effect on CD4 recovery in South African settings. The mean CD4 trajectories estimated by Nash *et al*, in patients starting ART with CD4 >200, have therefore been adjusted upwards to produce a trend more consistent with CD4 patterns observed in South Africa. The coefficients of variation in CD4 counts in patients starting ART at CD4 counts >200 have been estimated from fitting gamma distributions to CD4 data from US patients starting ART at higher CD4 counts (Lok *et al*, 2010). Based on these mean CD4 counts and assumed coefficients of variation, it is possible to calculate the proportions of patients in different CD4 categories, by assuming that CD4 counts are gamma-distributed. The resulting CD4 distributions are shown in Table 2. CD4 distributions are assumed to remain stable after 54 months on ART.

Table 2: Proportions of ART patients in different CD4 categories

	Months since ART initiation (with baseline CD4 <200)					Months since ART initiation (with baseline CD4 200-349)				
	6	18	30	42	54+	6	18	30	42	54+
CD4 <200	0.568	0.184	0.114	0.079	0.059	0.002	0.001	0.002	0.002	0.001
CD4 200-349	0.367	0.402	0.327	0.268	0.225	0.447	0.100	0.079	0.054	0.028
CD4 350-500	0.060	0.266	0.295	0.292	0.279	0.526	0.450	0.292	0.200	0.132
CD4 >500	0.006	0.148	0.264	0.360	0.437	0.026	0.449	0.627	0.744	0.839

Studies have shown that after treatment interruptions, there are typically very rapid declines in CD4 count during the first three months after treatment is discontinued (in excess of the rate of CD4 decline that would normally be expected in untreated patients), followed by slower rates of CD4 decline thereafter (El-Sadr *et al*, 2006; Sungkanuparph *et al*, 2007; Touloumi *et al*, 2006). To model this, it is assumed that there is an instantaneous change in CD4 at the point of ART discontinuation, after which patients experience the same rate of CD4 decline as untreated ART-naive adults of the same age and CD4 count. The percentage reduction in CD4 count after ART interruption has been shown to be strongly negatively associated with both the baseline CD4 count at the time of ART initiation and the CD4 count at the time of ART interruption (Touloumi *et al*, 2006). Suppose that $F_{c,s}(l)$ represents the probability that an individual who was in baseline CD4 category c, with a CD4 count in category s just prior to stopping ART, experiences a proportionate drop in CD4 count of $< l$, after stopping ART (this implies that $F_{c,s}(0) = 0$ and $F_{c,s}(1) = 1$, assuming that the CD4 count cannot increase after ART interruption). Further suppose that individuals in CD4 category s have CD4 counts uniformly distributed between lower limit a and upper limit b prior to ART interruption. Then the probability that their CD4 count will drop below limit l is

$$\int_a^b \frac{1}{b-a}\left(1-F_{c,s}\left(1-\frac{l}{u}\right)\right)du$$
$$\approx 0.25\left(1-F_{c,s}\left(1-\frac{l}{a}\right)\right)+0.25\left(1-F_{c,s}\left(1-\frac{l}{b}\right)\right)+0.5\left(1-F_{c,s}\left(1-\frac{2l}{a+b}\right)\right)$$

This formula can be easily evaluated if it is assumed that F is a cumulative beta distribution. The means of the corresponding beta probability density functions, $M_{c,s}$, are estimated from the average percentage reductions in CD4 counts estimated by Touloumi *et al* (2006). The standard deviations of the beta probability densities, $\sigma_{c,s}$, are calculated according to the formula

$$\sigma_{c,s} = \begin{cases} 0.2 \times M_{c,s} & \text{if } M_{c,s} < 0.5 \\ 0.2 \times (1 - M_{c,s}) & \text{if } M_{c,s} \geq 0.5 \end{cases}$$

where the multiple of 0.2 has been chosen in order to produce a range of CD4 reductions consistent with that reported by Sungkanuparph *et al* (2007). The assumed beta parameters are then used to determine the proportions of individuals falling into different CD4 categories after stopping ART, and these proportions are shown in Table 3.

Table 3: Changes in CD4 counts after ART interruptions

	Started ART with CD4 <200				Started ART with CD4 200-349			
	CD4 just prior to stopping ART				CD4 just prior to stopping ART			
	>500	350-500	200-349	<200	>500	350-500	200-349	<200
Mean proportionate CD4 reduction ($M_{c,s}$)	0.6	0.6	0.65	0.75	0.45	0.45	0.55	0.65
Proportion in CD4 group after stopping ART								
CD4 <200	0.20	0.78	1.00	1.00	0.01	0.26	0.98	1.00
CD4 200-349	0.70	0.22	0.00	0.00	0.47	0.73	0.02	0.00
CD4 350-500	0.10	0.00	0.00	0.00	0.46	0.01	0.00	0.00
CD4 >500	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00

The proportion of individuals on ART who move into a particular CD4 category after stopping ART is thus calculated by multiplying the appropriate column vector in Table 2 by the appropriate row vector in Table 3. For example, if an individual who started ART with a CD4 count of <200 stops ART after 30 months, the probability that their CD4 count is <200 after stopping ART is calculated as

$$0.114 \times 1.00 + 0.327 \times 1.00 + 0.295 \times 0.78 + 0.264 \times 0.20 = 0.72.$$

Results

Model calibration and validation

Figure 5 compares the modelled CD4 distribution in the Masiphumelele population, in the middle of 2010, with the CD4 distribution that was observed in a survey of CD4 counts in HIV-positive individuals in the Masiphumelele community. There is reasonable consistency between the model and the survey, although the model appears to estimate slightly too low a proportion of untreated HIV-positive adults in the CD4 <200 category.

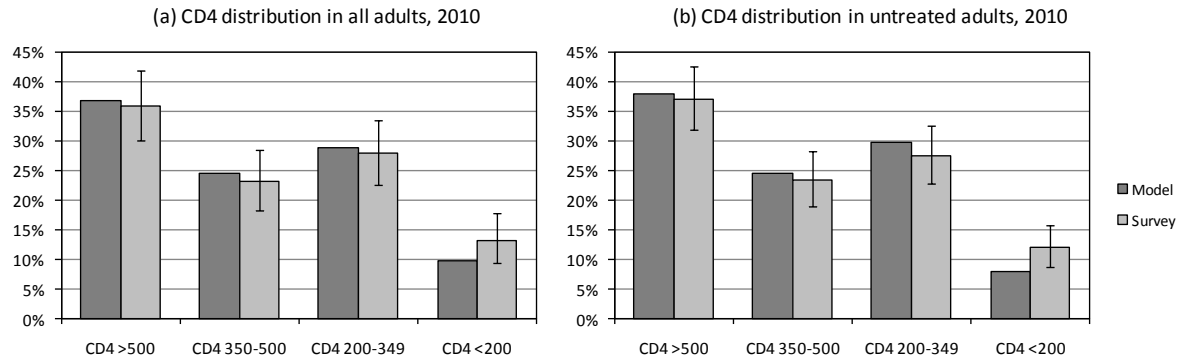

Figure 5: Comparison of observed and modelled CD4 distributions in HIV-positive adults
Error bars around survey estimates represent 95% confidence intervals.

CD4 data collected from the ART programme can also be used to validate the model assumptions about changes in CD4 distributions after ART initiation. For each patient in care at the end of each calendar year, an average CD4 count has been calculated by averaging across all CD4 measurements between the previous April and the following February. This means that the average date to which the CD4 measurements relate is mid-September, which differs from the date of the model estimates of the CD4 distributions (mid-year). Nevertheless, there is a reasonable degree of correspondence between the model estimates of CD4 distributions in treated adults and the actual CD4 distributions, when plotted on comparable time scales (Figure 6). However, the model does slightly over-estimate the proportions of treated adults with CD4 counts of 350-500 (Figure 6c). There is a clear improvement in CD4 distributions over time as the average duration on ART increases.

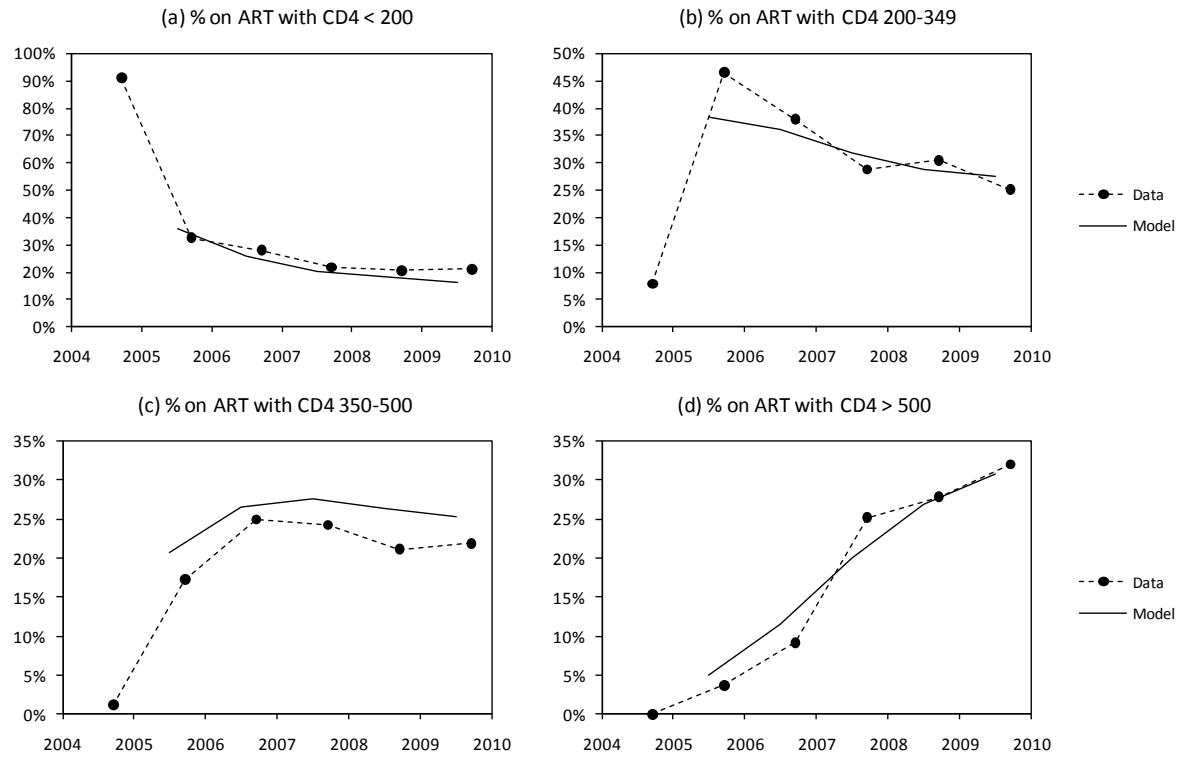

Figure 6: Comparison of observed and modelled CD4 distributions in adults receiving ART

The model estimates of the numbers of patients leaving care due to either transfer or loss to follow-up can also be compared with Masiphumelele programme statistics. However, since the model projects the change in population from mid-year to mid-year, it is necessary to convert the programme statistics (estimated by calendar year) into statistics that run from mid-year to mid-year. This is done by assuming, for example, that the numbers of transfers between mid-2005 and mid-2006 is half of the number in the 2005 calendar year and half of the number in the 2006 calendar year. After making this adjustment to the programme data, the programme data appear to be reasonably consistent with the model estimates, both in males and in females (Figure 7). No comparison with recorded numbers of deaths is shown, as the numbers of deaths in each year is very small, and there is thus substantial random variation.

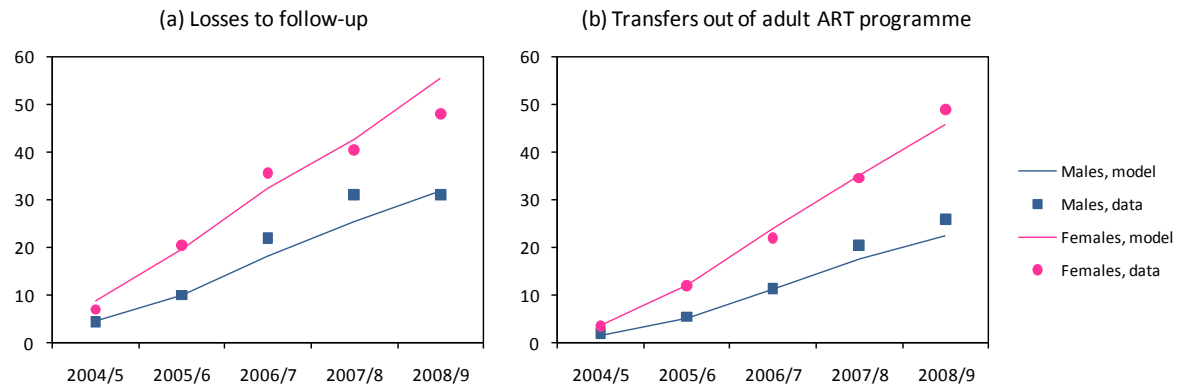

Figure 7: Comparison of modelled and recorded numbers of exits from the ART programme

As a final check on the reasonability of the demographic assumptions, we compare the model estimates of the size of the Masiphumelele population, aged 15 and older, with the data collected in the periodic censuses of the community (Figure 8). Although a high level of consistency would be expected (since the censuses have been used to determine the population growth assumptions and the demographic balancing equation has been used), complete consistency is not achieved because the model does not allow for negative numbers of in-migrants (which can be implied by equation 1 if the assumed rates of out-migration or mortality are too low). At a few ages this has led to population growth rates greater than those observed.

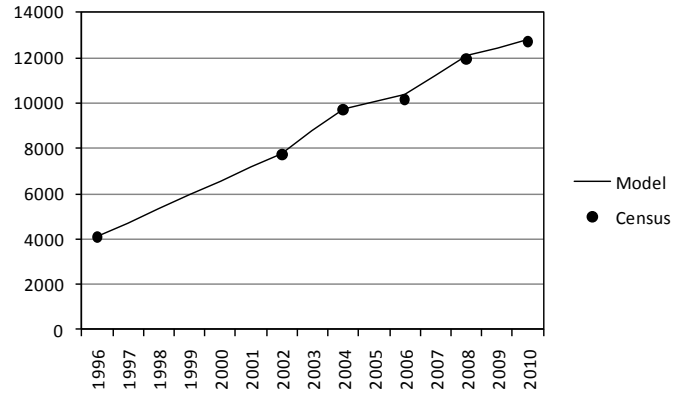

Figure 8: Growth in population aged 15 and older over time

Rates of ART initiation

Figure 9 compares two measures of ART access in the Masiphumelele community, described in detail elsewhere (Johnson and Boulle 2011). ART coverage in the community, defined as the proportion of people on ART or eligible for ART (CD4 <350) who are currently receiving ART, has increased rapidly over the 2004-6 period, but growth has subsequently slowed. By the middle of 2010, ART coverage was 44% in males and 55% in females. If ART access is instead measured in terms of the numbers of patients starting ART for the first time in a particular year, divided by the number of people becoming eligible for ART in the same year, this enrolment ratio reaches its highest level in 2005 and 2006. The ratio has since stabilized at around 1 in females (indicating that universal access in females will be achieved if

enrolment continues at its current rate) and at around 0.75 in men (indicating that if enrolment continues at its current rate, a quarter of HIV-positive men will never receive ART).

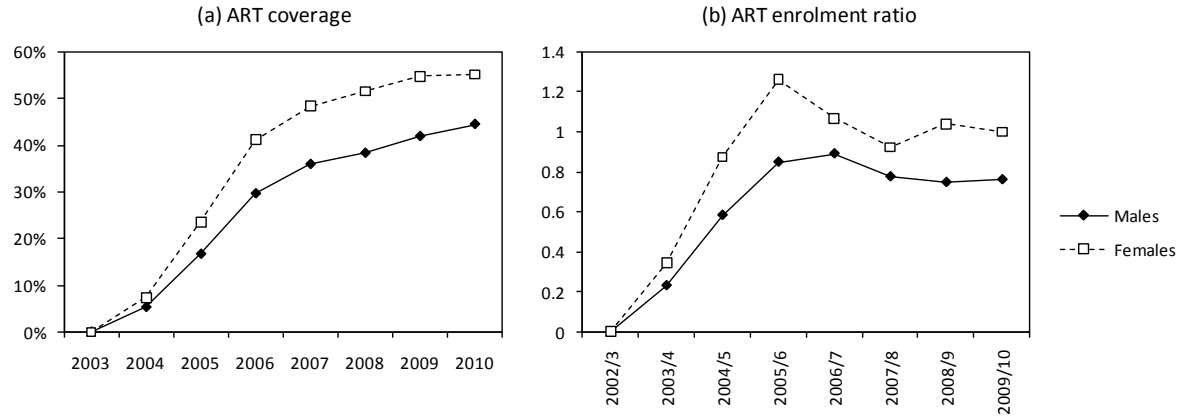

Figure 9: Trends in adult ART access in Masiphumelele
Treatment eligibility is defined in terms of the 2010 WHO ART guidelines.

Another way in which to measure ART access is by calculating rates of ART initiation per 100 person years of untreated ART eligibility. This measure is shown in Table 4. In both males and females who are ART-naive and have CD4 <200, rates of ART initiation have increased steadily over the 2004-2009 period. Male ART initiation rates have stabilized since 2006, while female ART initiation rates have risen to implausibly high levels in 2008 and 2009 – possibly suggesting that the denominator (numbers of untreated women with CD4 <200) may be underestimated by the model. During the period when the old ART initiation criteria were in place (2004 and 2007-2009) rates of ART initiation in the CD4 200-349 category were only about 6% of those in the CD4 <200 category. Most of these individuals would have started ART because of their clinical symptoms. During the period when the CD4 threshold for ART initiation was changed to 350 (2005 and 2006), the ART initiation rate in the CD4 200-349 category was somewhat higher in men (12% of the rate in men with CD4 <200) and substantially higher in women (22% of the rate in women with CD4 <200). Rates of re-initiating ART after interrupting therapy were generally lower than the rates of starting ART for the first time in patients with CD4 <200.

Table 4: Rates of ART initiation (number of patients starting ART per 100 person years of ART eligibility)

	2004	2005	2006	2007	2008	2009	2005-6	2007-9	2004-9
Males									
ART-naive, CD4 <200	36.5	56.1	96.0	102.1	94.5	107.5	74.3	101.3	78.3
ART-naive, CD4 200-349	1.2	8.4	9.3	4.9	7.9	5.8	8.8	6.2	6.4
Discontinued ART	0.0	20.2	36.7	51.7	82.1	27.8			32.2
Females									
ART-naive, CD4 <200	51.9	82.2	133.5	117.1	271.7	303.1	103.6	213.3	121.4
ART-naive, CD4 200-349	2.8	18.0	27.2	10.1	11.4	13.1	22.5	11.6	13.8
Discontinued ART	67.0	20.0	31.3	40.1	46.6	10.8			24.9

As a result of these high levels of ART initiation, substantial reductions in AIDS mortality have occurred. The model estimate of the trend in adult AIDS mortality is shown in Figure

10a, and is compared with the trend that would have been expected if there had been no provision of ART in Masiphumelele. Between mid-2008 and mid-2009, the AIDS mortality rate in the Masiphumelele community was 6.7 per 1000, 53% lower than the rate of 14.0 per 1000 that would have been expected in the absence of ART. Most of the AIDS deaths occurring in Masiphumelele are still occurring in individuals who have never accessed ART, although an increasingly high proportion of AIDS deaths are occurring in patients who have stopped ART (Figure 10b).

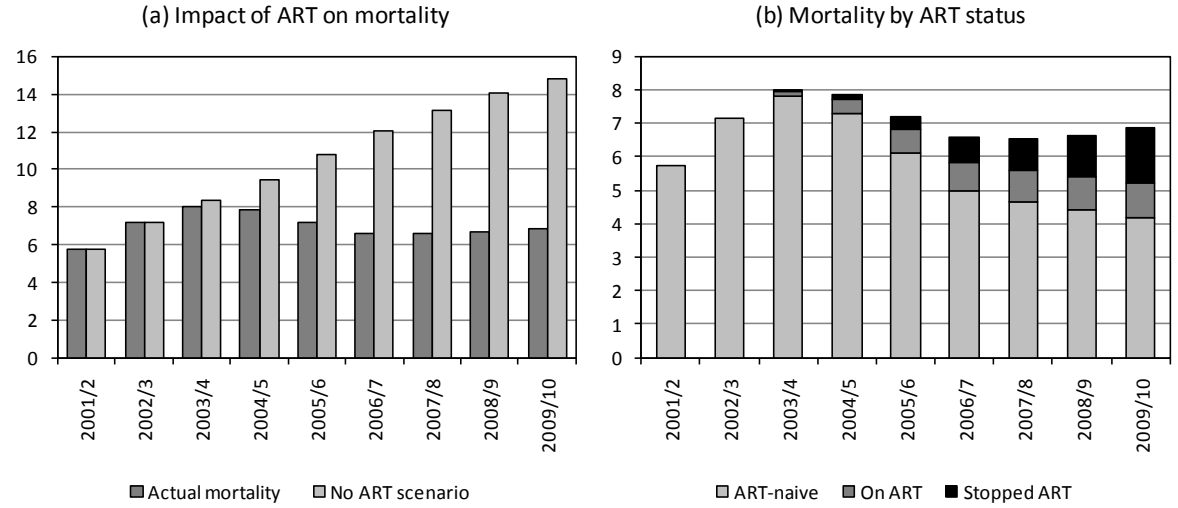

Figure 10: Trends in adult AIDS death rates, per 1000 population aged 15+

Discussion

The ART initiation rates in Table 4 show that even when ART eligibility criteria are expanded to include all individuals with CD4 counts below 350, the rates of ART initiation in the CD4 200-349 category remain substantially lower than the rates of ART initiation in the CD4 <200 category. This difference is particularly substantial in males, possibly because HIV-positive men are less likely to have been tested and to know their HIV status than HIV-positive women, who would typically be diagnosed through antenatal HIV screening. The extent of the difference in ART initiation rates between the 200-349 and <200 CD4 categories is likely to be highly dependent on the proportion of HIV-positive individuals who know their HIV status and who are receiving regular monitoring. Models that are used to project future rates of ART uptake will need to take into account the lower rates of ART initiation at higher CD4 counts, and should ideally allow for rates of ART initiation at higher CD4 counts to depend on proportions of people who have received HIV testing.

Rates of restarting ART after treatment interruption are generally lower than rates of ART initiation observed in patients who are ART-naive with CD4 counts below 200. The overall rate of ART re-initiation in females who have stopped ART is 24.9 per 100 person-years, similar to the rate of 24.1 estimated by Kranzer *et al* (2010), using data from the same community. However, our estimated rate of ART resumption in males, 32.2 per 100 person years, is substantially higher than that estimated by Kranzer *et al* (17.6). This may be because of differences in the method used to calculate the denominator in the rate calculation: our

method excludes follow-up in patients after they have left the Masiphumelele population, due to migration and death (both of which occur at substantially higher rates in men who have discontinued ART than in women who have discontinued ART).

Model limitations

The most significant limitation of the current analysis is that it does not include any consideration of sensitivity of the model to changes in key parameters. A more thorough uncertainty analysis is required in order to assess the ranges of uncertainty around the model outputs. Key sources of uncertainty to consider will include rates of CD4 change in untreated adults, rates of ART interruption in treated individuals, adjustments to HIV incidence rates from national population models, and assumptions about out-migration.

Another problem is that although the model estimates of total numbers of adults on ART are in reasonable agreement with the recorded numbers of adults on ART, the modelled age distribution of patients on ART is quite different from the actual age distribution. In particular, the model over-estimates the numbers of 25-34 year olds on ART and under-estimates the numbers of ART patients aged 45 and older, both in males and in females (results not shown). One possible explanation for this discrepancy is that the model does not allow for age differences in the rate of ART resumption after interruption of therapy. If rates of resumption are significantly higher at older ages than at younger ages, as suggested by data from the Masiphumelele community (Kranzer *et al*, 2010), the numbers on ART in the 25-34 age group would be expected to be lower than those modelled, and the number in the 45+ age group would be expected to be higher than those modelled. Alternatively, there may be other age-specific factors affecting knowledge of HIV status and socioeconomic status, which in turn influence rates of access to ART at different ages. Since the model in its current form does not provide a good fit to age-specific ART data, caution should be applied in the use of age-specific outputs from the model, particularly in the case of outputs relating to ART.

A limitation of the method used to estimate mortality after ART initiation is that this is based only on deaths recorded in patients' clinical records. This is likely to be an under-estimate of the true numbers of deaths that have occurred. Tracing studies conducted in South Africa (Maskew *et al*, 2007; Dalal *et al*, 2008) and studies that have linked patients with known ID numbers to the national population register (Fairall *et al*, 2008; Van Cutsem *et al*, 2011; Fox *et al*, 2010) have demonstrated that a substantial proportion of patients who are classified as "lost to follow-up" are actually dead, with death often occurring at the same time that the patient was lost to the system. Our model may therefore understate the number of HIV deaths that occur while patients are on ART.

The model could also be improved if ART programme statistics were reported over periods from mid-year to mid-year, to be consistent with the way in which the model projects the Masiphumelele population. Although we have approximated the numbers over mid-year to mid-year from the statistics reported in each calendar year, there is some loss of accuracy as a result of the associated assumption that events occur uniformly over each calendar year.

References

Actuarial Society of South Africa (2011) *ASSA2008 AIDS and Demographic Model*. Available: http://aids.actuarialsociety.org.za. Accessed 5 April 2011

Adam M. A. and Johnson L. F. (2009) Estimation of adult antiretroviral treatment coverage in South Africa. *South African Medical Journal.* 99(9): 661-7

Auvert B., Males S., Puren A., *et al.* (2004) Can highly active antiretroviral therapy reduce the spread of HIV? A study in a township of South Africa. *Journal of Acquired Immune Deficiency Syndromes.* 36(1): 613-621

Badri M., Lawn S. D. and Wood R. (2006) Short-term risk of AIDS or death in people infected with HIV-1 before antiretroviral therapy in South Africa: a longitudinal study. *Lancet.* 368(9543): 1254-9

Boulle A., Van Cutsem G., Hilderbrand K., *et al.* (2010) Seven year experience of a primary care antiretroviral treatment programme in Khayelitsha, South Africa. *AIDS.* 24: 563-72

Braithwaite R. S., Justice A. C., Chang C. C., *et al.* (2005) Estimating the proportion of patients infected with HIV who will die of comorbid diseases. *American Journal of Medicine.* 118(8): 890-8

Collaborative Group on AIDS Incubation and HIV Survival (2000) Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *Lancet.* 355: 1131-1137

Connelly D., Veriava Y., Roberts S., *et al.* (2007) Prevalence of HIV infection and median CD4 counts among health care workers in South Africa. *South African Medical Journal.* 97(2): 115-20

Cornell M., Grimsrud A., Fairall L., *et al.* (2010) Temporal changes in programme outcomes among adult patients initiating antiretroviral therapy across South Africa, 2002-2007. *AIDS.* 24(14): 2263-70

Coutsoudis A., England K., Rollins N., *et al.* (2010) Women's morbidity and mortality in the first 2 years after delivery according to HIV status. *AIDS.* 24(18): 2859-66

Dalal R. P., Macphail C., Mqhayi M., *et al.* (2008) Characteristics and outcomes of adult patients lost to follow-up at an antiretroviral treatment clinic in Johannesburg, South Africa. *Journal of Acquired Immune Deficiency Syndromes.* 47(1): 101-7

Dorrington R. and Moultrie T. (2009) *Making use of the consistency of patterns to estimate age-specific rates of inter-provincial migration in South Africa*. Meeting of the Population Association of America.

Dorrington R. E., Johnson L. F., Bradshaw D. and Daniel T. (2006) *The Demographic Impact of HIV/AIDS in South Africa. National and Provincial Indicators for 2006*. Centre for Actuarial Research, South African Medical Research Council and Actuarial Society of South Africa. Available: http://www.commerce.uct.ac.za/care

El-Sadr W. M., Lundgren J. D., Neaton J. D., *et al.* (2006) CD4+ count-guided interruption of antiretroviral treatment. *New England Journal of Medicine.* 355(22): 2283-96

Eligibility for ART in lower income countries collaboration (2008) Duration from seroconversion to eligibility for antiretroviral therapy and from ART eligibility to death in adult HIV-infected patients from low and middle-income countries: collaborative analysis of prospective studies. *Sexually Transmitted Infections.* 84 (Suppl i): i31-6

Fairall L. R., Bachmann M. O., Louwagie G. M., *et al.* (2008) Effectiveness of antiretroviral treatment in a South African program: a cohort study. *Archives of Internal Medicine.* 168(1): 86-93

Fatti G., Grimwood A. and Bock P. (2010) Better antiretroviral therapy outcomes at primary healthcare facilities: an evaluation of three tiers of ART services in four South African provinces. *PLoS One.* 5(9): e12888

Fox M. P., Brennan A., Maskew M., *et al.* (2010) Using vital registration data to update mortality among patients lost to follow-up from ART programmes: evidence from the Themba Lethu Clinic, South Africa. *Tropical Medicine and International Health.* 15(4): 405-13

Fox M. P. and Rosen S. (2010) Patient retention in antiretroviral therapy programs up to three years on treatment in sub-Saharan Africa, 2007-2009: systematic review. *Tropical Medicine and International Health.* 15 (Suppl 1): 1-15

Glynn J. R., Sonnenberg P., Nelson G., *et al.* (2007) Survival from HIV-1 seroconversion in Southern Africa: a retrospective cohort study in nearly 2000 gold-miners over 10 years of follow-up. *AIDS.* 21(5): 625-32

Johnson L. F. and Boulle A. (2011) How should access to antiretroviral treatment be measured? *Bulletin of the World Health Organization.* 89: 157-60

Johnson L. F., Dorrington R. E., Bradshaw D., *et al.* (2009) Sexual behaviour patterns in South Africa and their association with the spread of HIV: insights from a mathematical model. *Demographic Research.* 21: 289-340

Johnson L. F., Dorrington R. E. and Matthews A. P. (2007) An investigation into the extent of uncertainty surrounding estimates of the impact of HIV/AIDS in South Africa. *South African Journal of Science.* 103: 135-140

Judson D. H. and Popoff C. L. (2004). Selected general methods. In J. S. Siegel and D. A. Swanson, Eds. *The Methods and Materials of Demography*. San Diego, Elsevier Academic Press.

Kranzer K., Lewis J. J., Ford N., *et al.* (2010) Treatment interruption in a primary care antiretroviral therapy program in South Africa: cohort analysis of trends and risk factors. *Journal of Acquired Immune Deficiency Syndromes.* 55(3): e17-23

Lok J. J., Bosch R. J., Benson C. A., *et al.* (2010) Long-term increase in CD4+ T-cell counts during combination antiretroviral therapy for HIV-1 infection. *AIDS.* 24(12): 1867-76

Maskew M., MacPhail P., Menezes C. and Rubel D. (2007) Lost to follow up: contributing factors and challenges in South African patients on antiretroviral therapy. *South African Medical Journal.* 97(9): 853-7

Nash D., Katyal M., Brinkhof M. W., *et al.* (2008) Long-term immunologic response to antiretroviral therapy in low-income countries: a collaborative analysis of prospective studies. *AIDS.* 22(17): 2291-302

Nglazi M. D., Lawn S. D., Kaplan R., *et al.* (2011) Changes in programmatic outcomes during 7 years of scale-up at a community-based antiretroviral treatment service in South Africa. *Journal of Acquired Immune Deficiency Syndromes.* 56(1): e1-8

Rehle T. M. and Shisana O. (2005) Estimates of eligibility for antiretroviral treatment (ART) and projected ART impact on AIDS mortality among South African educators. *Journal of Social Aspects of HIV/AIDS.* 2(3): 304-10

Sanne I., Orrell C., Fox M. P., *et al.* (2010) Nurse versus doctor management of HIV-infected patients receiving antiretroviral therapy (CIPRA-SA): a randomised non-inferiority trial. *Lancet.* 376(9734): 33-40

Sungkanuparph S., Kiertiburanakul S., Apisarnthanarak A., *et al.* (2007) Rapid CD4 decline after interruption of non-nucleoside reverse transcriptase inhibitor-based antiretroviral therapy in a resource-limited setting. *AIDS Research and Therapy.* 4: 26

Touloumi G., Pantazis N., Antoniou A., *et al.* (2006) Highly active antiretroviral therapy interruption: predictors and virological and immunologic consequences. *Journal of Acquired Immune Deficiency Syndromes.* 42(5): 554-61

Van Cutsem G., Ford N., Hildebrand K., *et al.* (2011) Correcting for mortality among patients lost to follow up on antiretroviral therapy in South Africa: a cohort analysis. *PLoS One.* 6(2): e14684

van Sighem A. I., van de Wiel M. A., Ghani A. C., *et al.* (2003) Mortality and progression to AIDS after starting highly active antiretroviral therapy. *AIDS.* 17: 2227-2236

Williams B. G., Korenromp E. L., Gouws E., *et al.* (2006) HIV infection, antiretroviral therapy, and CD4+ cell count distributions in African populations. *Journal of Infectious Diseases.* 194(10): 1450-8

Wood R., Middelkoop K., Myer L., *et al.* (2007) Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control. *American Journal of Respiratory and Critical Care Medicine.* 175(1): 87-93

World Health Organization (2009) *Rapid advice: Antiretroviral therapy for HIV infection in adults and adolescents*. Available: http://www.who.int/hiv/pub/arv/advice/en/index.html. Accessed 30 July 2010.

Appendix: Estimation of parameters for a four-stage model of CD4 decline

In order to estimate the rates at which individuals become eligible for antiretroviral treatment, it is necessary to estimate the rates of CD4 decline in HIV-infected individuals in the absence of antiretroviral treatment. Figure 3 in the main text shows one possible model that we might wish to fit to South African CD4 data. Note that in this simple model we are ignoring non-AIDS mortality and we are also not allowing for initiation of antiretroviral treatment. Suppose that parameter θ can be estimated from an independent source. For example, a study of mortality rates in HIV-infected individuals without access to antiretroviral treatment in Cape Town (Badri *et al*, 2006), suggests that the ratio of the mortality rate in the CD4 200-349 category to that in the CD4 <200 category is 0.30. Also suppose that we have determined the mean HIV survival time, in the absence of antiretroviral treatment, from an independent source. The mean, μ, is equal to

$$\frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\frac{1}{\lambda_3+\theta\lambda_4}+\frac{\lambda_3}{\lambda_3+\theta\lambda_4}\times\frac{1}{\lambda_4}.$$

Based on an analysis of South African reported death data (Johnson *et al*, 2007) and more recent data on the survival of untreated HIV-infected adults (Glynn *et al*, 2007; Eligibility for ART in lower income countries collaboration 2008), we assume a mean survival time of 12 years. Our objective is to estimate the parameters λ_2, λ_3 and λ_4. Once these parameters have been estimated, the parameter λ_1 can be obtained using the equation

$$\frac{1}{\lambda_1}=\mu-\frac{1}{\lambda_2}-\frac{\lambda_3+\lambda_4}{\lambda_4\left(\lambda_3+\theta\lambda_4\right)}.$$

We aim to estimate these parameters by fitting our model to data from three different South African studies that attempted to determine the proportions of HIV-infected adults in different CD4 categories. The results of these three studies are summarized in Table A1. All three studies were conducted when access to antiretroviral treatment in South Africa was fairly limited, and the proportions can therefore be assumed to be representative of untreated individuals.

Table A1: Empirical estimates of proportions of HIV-infected adults in different CD4 stages

Study	Year of survey	Population sampled	# HIV+ adults	% with CD4 of >500	350-500	200-349	<200
Auvert *et al* (2004)	2002	Households in Orange Farm	196	46.0	25.6	18.9	9.5
Rehle and Shisana (2005)	2004	Teachers	444	27.9	19.8	30.0	22.3
Connelly *et al* (2007)	2005	Health workers	74	35.1	17.6	28.4	18.9

To fit the model to the cross-sectional CD4 data, it is necessary to make assumptions about the annual numbers of new HIV infections in South African adults, in each year since the start of the South African HIV/AIDS epidemic – these are obtained from the ASSA2003 AIDS and Demographic Model (Dorrington *et al*, 2006). The model shown in Figure 3 is used to project the change in the number of HIV-infected adults in each CD4 category at the

middle of each year, taking as entrants to the CD4 >500 stage the number of new HIV infections. (It is reasonable to assume that all adults have CD4+ counts above 500 at the time of acquiring HIV, as South African CD4 data (Williams *et al*, 2006; Coutsoudis *et al*, 2010) suggest that the proportion of HIV-negative adults with CD4 <500 is less than 3%.) The model is then fitted to the data shown in Table A1 using maximum likelihood (the method used to define the likelihood is explained at the end of the appendix). The resulting maximum likelihood estimates are summarized in Table A2, and the maximum likelihood fit to the CD4 data is shown in Figure A1.

Table A2: Maximum likelihood estimates of model parameters

Stage s	Definition	Maximum likelihood estimate	
		λ_s	$1/\lambda_s$
1	CD4 >500	0.306	3.27
2	CD4 350-500	0.490	2.04
3	CD4 200-349	0.248	4.04
4	CD4 <200	0.224	4.47

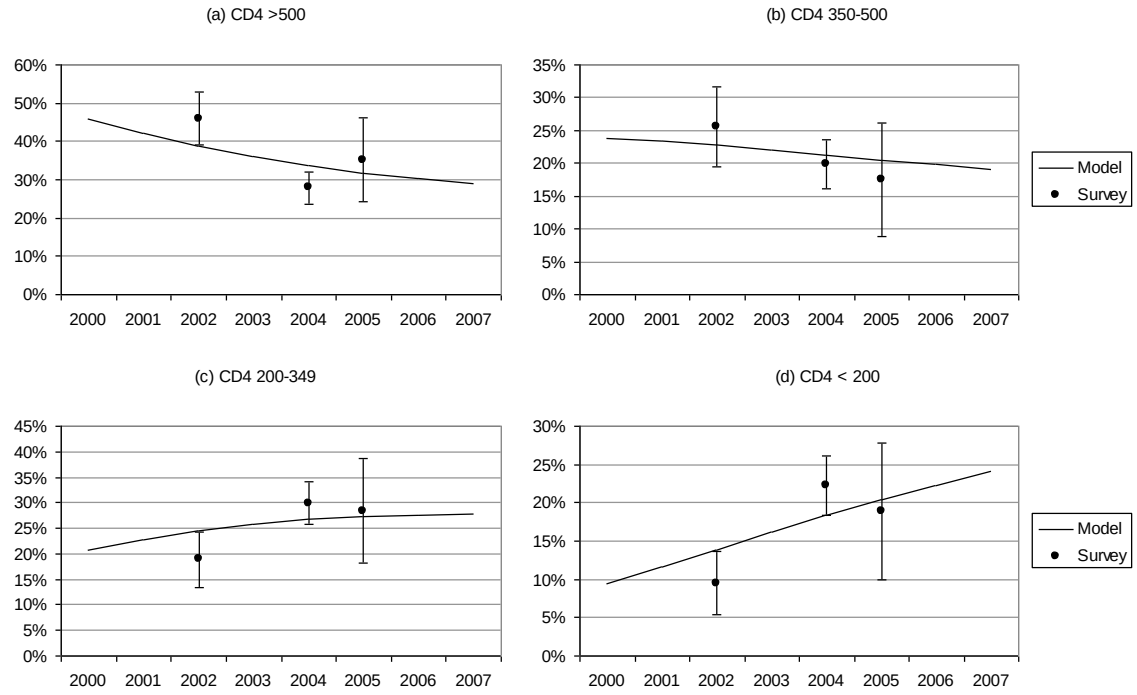

Figure A1: Maximum likelihood fit to cross-sectional CD4 data
Dots represent results from surveys, and error bars represent 95% confidence intervals around these estimates. Solid black line represents model estimate when the maximum likelihood parameters (Table A2) are entered into the model.

Three independent checks on the validity of the maximum likelihood estimates were performed. Firstly, the model estimates of the proportions of infected adults surviving at each integer duration following HIV acquisition were compared with estimated rates of survival from a cohort of South African gold miners who did not have access to antiretroviral treatment (Glynn *et al*, 2007). (The effect of non-AIDS mortality was removed, so the survival rates effectively represent the survival that would be expected in the absence of non-

AIDS mortality). As Figure A2 shows, the model estimates are reasonably consistent with the estimates from the cohort of gold miners.

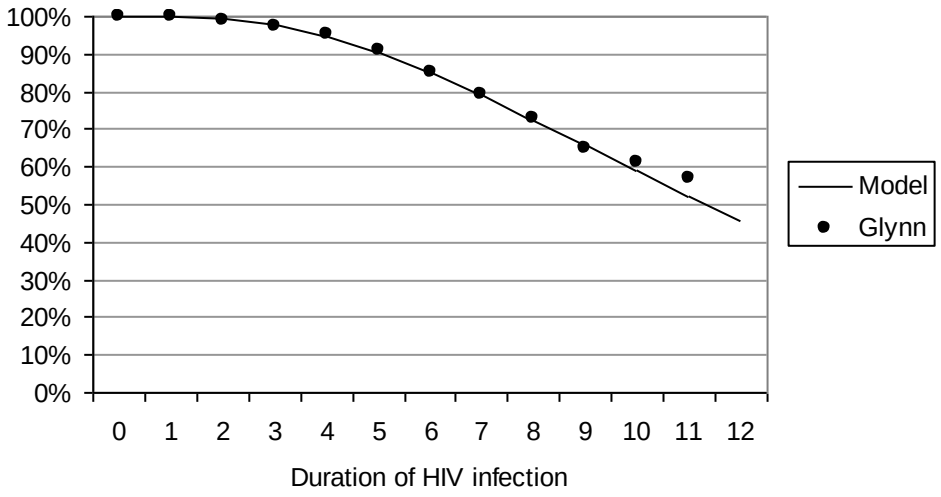

Figure A2: Comparison of model estimates and empirical estimates of proportions of HIV-infected adults surviving, in the absence of antiretroviral treatment

The second independent check was to compare the model estimates of the proportion of HIV-positive adults in each CD4 category in 2004 with corresponding estimates of another model of CD4 decline that has been applied to South Africa (Adam and Johnson 2009). The latter model estimates proportions in the CD4 >500 and CD4 350-500 categories of 36.7% and 24.6% respectively, which are roughly consistent with our model estimates (33.7% and 21.3% respectively). In the CD4 <350 category, the model estimates are not directly comparable because the latter model has an 'AIDS' stage that is not defined in terms of CD4 count, and a small proportion of individuals in this stage would have CD4 counts above 200. As a result, the estimated proportion of individuals who have CD4 counts of 200-349 *and* have not progressed to AIDS (19.9%) is lower than our model estimate of the proportion of infected adults with CD4 counts of 200-349 (26.7%). The estimated proportion of adults with AIDS *or* a CD4 count <200 (18.8%) is consistent with our model estimate of the proportion of adults with a CD4 count <200 (18.3%).

The third check was to compare the estimated mortality rate in untreated patients with CD4 count <200 with that estimated empirically, in patients from Cape Town without access to ART (Badri *et al,* 2006) and in pregnant women in KwaZulu-Natal without access to ART (Coutsoudis *et al,* 2010). The model estimate (0.224 in Table A2) is somewhat lower than the estimate of 0.27 from the Cape Town data. It is important to note that in the Cape Town cohort, individuals were not tracked from seroconversion, but from the date at which they first were diagnosed HIV-positive, and we would therefore expect some bias towards sicker individuals with a higher mortality rate. This might explain why our model estimate is lower than the observed mortality rate. The model estimate of 0.224 is higher than the mortality rate of 0.171 observed in pregnant women in KZN, but this is probably because the CD4 counts in these women were reduced by haemodilution during pregnancy (i.e. the women appear to be at a more advanced stage of disease than is actually the case).

Method to derive likelihood function

Suppose that our model estimates that the proportions of individuals in stage *i* of HIV infection, in year t_j, is $\pi_i(t_j)$ (i = 1, 2, 3, 4). Further suppose that the observed numbers of infected individuals in the j^{th} study, conducted in year t_j, who are in stage *i* is n_{ij}. It could then be assumed that the n_{ij} terms are multinomially distributed, so that the likelihood function in respect of the j^{th} study is equal to

$$\binom{n_j}{n_{1j}\ n_{2j}\ n_{3j}\ n_{4j}} \pi_1(t_j)^{n_{1j}} \pi_2(t_j)^{n_{2j}} \pi_3(t_j)^{n_{3j}} \pi_4(t_j)^{n_{4j}}, \qquad \text{(A1)}$$

where n_j is the total number of HIV-positive adults in the j^{th} study. This is a fixed effects model, i.e. it is assumed that the CD4 distribution is the same for all South African populations that we might choose to sample. The assumption of a fixed effects framework is probably unrealistic, since some populations may have better access to healthcare, some may have better nutrition, some populations may be experiencing more advanced epidemics than others, etc. To account for variation in proportions between sub-populations, define ρ_{ij} as the true proportion of infected individuals in stage *i*, in the j^{th} sub-population. Then it would be natural to assume that the ρ_{ij} terms are Dirichlet-distributed, i.e.

$$p\left(\boldsymbol{\rho}_j \mid \boldsymbol{\pi}(t_j), \varphi\right) = \Gamma(\varphi) \prod_{i=1}^{4} \left(\rho_{ij}^{\ \varphi\pi_i(t_j)-1}\right) \Big/ \Gamma(\varphi\pi_i(t_j)), \qquad \text{(A2)}$$

where $\boldsymbol{\rho}_j$ and $\boldsymbol{\pi}(t_j)$ represent the vectors of ρ_{ij} and $\pi_i(t_j)$ values respectively. Note that from the properties of the Dirichlet distribution,

$$\mathrm{E}\left[\rho_{ij}\right] = \pi_i(t_j)$$
$$\mathrm{Var}\left[\rho_{ij}\right] = \frac{\pi_i(t_j)\left(1-\pi_i(t_j)\right)}{\varphi+1}$$

so that the φ variable controls the variance of the random effects. The likelihood function is then

$$p\left(\mathbf{n}_j \mid \boldsymbol{\pi}(t_j), \varphi\right) = \int_{\boldsymbol{\rho}_j} p\left(\mathbf{n}_j \mid \boldsymbol{\rho}_j\right) p\left(\boldsymbol{\rho}_j \mid \boldsymbol{\pi}(t_j), \varphi\right) d\boldsymbol{\rho}_j \qquad \text{(A3)}$$

where $\mathbf{n}_j$ represents the vector of n_{ij} values. The likelihood in equation (A3) can be more fully expressed as

$$\int_{\rho_{ij}} \int_{\rho_{2j}} \int_{\rho_{3j}} \int_{\rho_{4j}} \Gamma(\varphi) \binom{n_j}{n_{1j}\ n_{2j}\ n_{3j}\ n_{4j}} \prod_{i=1}^{4} \frac{\rho_{ij}^{\ n_{ij}+\varphi\pi_i(t_j)-1}}{\Gamma\left(\varphi\pi_i(t_j)\right)} d\rho_{4j} d\rho_{3j} d\rho_{2j} d\rho_{1j}.$$

Note that after factoring out the terms that are independent of ρ_{ij} in the above equation, the integral is itself of a Dirichlet form, and therefore integrates to 1 with the multiplication of an appropriate constant term. Hence

$$p\left(\mathbf{n}_j \mid \boldsymbol{\pi}(t_j), \varphi\right) = \frac{\Gamma(\varphi)}{\Gamma(n_j + \varphi)} \binom{n_j}{n_{1j}\ n_{2j}\ n_{3j}\ n_{4j}} \prod_{i=1}^{4} \frac{\Gamma\left(n_{ij} + \varphi\pi_i(t_j)\right)}{\Gamma\left(\varphi\pi_i(t_j)\right)}. \qquad \text{(A4)}$$

The total likelihood is obtained by multiplying the values of the likelihood for each individual study (equation A4). We maximize the natural log of this likelihood with respect to φ, λ_2, λ_3 and λ_4 (since the λ_i terms determine the $\pi_i(t_j)$ estimates), and hence the factorial term can be omitted from equation (4). The expression we attempt to maximize is thus

$$\sum_{j=1}^{3}\left(\ln\left(\Gamma(\varphi)\right) - \ln\left(\Gamma(n_j + \varphi)\right) + \sum_{i=1}^{4} \ln\left(\Gamma\left(n_{ij} + \varphi\pi_i(t_j)\right)\right) - \ln\left(\Gamma\left(\varphi\pi_i(t_j)\right)\right)\right).$$

The maximum likelihood estimate of φ (not shown in Table A2) is 188.5.