Thembisa version 4.3:
A model for evaluating the impact of HIV/AIDS in South Africa

June 2020

Leigh Johnson[1] and Rob Dorrington[2]

1. Centre for Infectious Disease Epidemiology and Research, University of Cape Town
2. Centre for Actuarial Research, 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

Table of contents

1. Introduction

Thembisa is a mathematical model of the HIV epidemic in South Africa. The purpose of this document is to provide a technical description of the model and the methods used to calibrate the model. The focus of this report is limited to the national model; descriptions of the calibration procedures for the provincial models are published separately [1]. Readers who are interested in the outputs of the model are referred to the Thembisa website (www.thembisa.org), which includes various model output files.

This document describes version 4.3 of the Thembisa model, and is similar in structure and content to the previous report that described version 4.2 of the model [2]. Briefly, the main changes to the previous version of the model are as follows:

- Assumptions about antenatal bias (the difference in HIV prevalence between pregnant women attending public antenatal clinics and women in the general population) have been revised, and now depend on the recency of HIV infection.
- Assumptions about the effect of HIV and ART on fertility have been revised based on a recent analysis of pregnancy incidence rates in the Western Cape [3].
- The model of mother-to-child transmission (MTCT) has been updated to allow for the effect of maternal ART interruptions during pregnancy on MTCT.
- The model of condom use has been simplified, with a formal statistical analysis of nationally-representative surveys informing the key parameters.
- Assumptions about male-to-female transmission in sex worker-client relationships have been modified, allowing for changes in average transmission rates over time.
- The model has been extended to include a separation between first-line and second-line ART patients.
- The model has been updated using more recent programmatic data (numbers of individuals tested for HIV, numbers of patients receiving antiretroviral treatment (ART), numbers of medical male circumcision (MMC) operations and numbers of people using pre-exposure prophylaxis (PrEP)).
- The model of viral suppression after ART initiation has been revised, using a Bayesian approach to allow for uncertainty in the extent of the bias due to missing viral load data.
- The model is calibrated to antiretroviral (ARV) metabolite data from the 2012 and 2017 household surveys (proportions of HIV-positive adults who have ART detectable in their blood specimens) [4, 5].
- Several new data sources have been incorporated in the paediatric HIV model calibration process, including data on the proportion of child deaths in which there has been an HIV diagnosis, and data from the National Health Laboratory Services (NHLS) on the age distribution of children receiving ART [6].

The model is deterministic and compartmental. Being compartmental means that the population is divided into various cohorts and sub-cohorts; these are summarized in Table 1.1. The indexing variables for each compartment, which are used throughout this report, are also shown in Table 1.1.

Table 1.1: Index variables and compartments in Thembisa

Symbol	Description	Value	Definition
a	ART status	0	ART-naïve
		1	On ART or previously treated
d	Time since	0	ART-naïve
	ART	1	1st 6 months after ART start
	initiation	2	7-18 months after ART start
		3	19-30 months after ART start
		4	31-42 months after ART start
		5	>42 months after ART start
g	Sex	1	Male
		2	Female
*i**	Risk group	1	High risk
		2	Low risk
j	Partner	1	High risk
	risk group	2	Low risk
l	Marital	0	Unmarried/short-term relationship
	status *or*	1	Married/long-term relationship
	relationship	2	Sex worker/sex worker-client relationship
	type	3	MSM/same-sex short-term relationship
r	Circumcision	0	Uncircumcised
	status	1	Circumcised
s	HIV stage	0	Uninfected
	or baseline	1	Acute HIV
	CD4 count	2	HIV-positive, CD4 ≥500 (after acute infection)
		3	HIV-positive, CD4 350-499
		4	HIV-positive, CD4 200-349
		5	HIV-positive, CD4 <200
t	Year	-	
*v**	HIV testing	0	Never tested for HIV
	history	1	Lasted tested HIV-negative
		2	Diagnosed HIV-positive
x	Age	0-90+	Age at last birthday (at start of the year)
y	Partner age	10-90+	Age at last birthday (at start of the year)

* Note that in section 3.2 we use the symbol *i* instead of *v* to refer to the HIV testing history, in order to avoid confusion with the fraction of pregnant women who are tested for HIV.

The large number of parameters in the Thembisa model makes the calibration of the model challenging. There are many data sources that the model needs to be calibrated to, and there are also many parameters that can be varied in the calibration process. Rather than simultaneously vary all of the parameters to fit the model to all the available data (which would be computationally impractical), we break the calibration down into a number of steps. Each step involves varying a subset of the model parameters to match the model to a subset of the calibration data, the parameter subsets being chosen based on their relative importance in determining the model fit to the relevant data source. The steps in the calibration process are summarized in Table 1.2. Some of the model parameters can only be determined at a national level, as province-specific data are lacking, while in other cases (for example, annual numbers of patients starting ART), the parameters are estimated at a provincial level and are then aggregated to obtain the national model parameters. In a few cases (for example, the

initial HIV prevalence), the model parameter is estimated independently for the national and provincial models. Although the main focus of this report is on the calibration of the model to the adult HIV prevalence and mortality data at a national level (described in more detail in section 7 of the report), the appendices and the provincial modelling report describe the other steps in the calibration in more detail (as shown in the last column of Table 1.2).

Table 1.2: Steps in the Thembisa calibration process

Step	Data sources in calibration	Parameters varied	Description
1. Calibrate to provincial adult HIV data	Total numbers of patients on ART HIV prevalence in household surveys HIV prevalence in antenatal surveys ARV metabolite data % of adult ART patients who are male	Rates of ART initiation Viral suppression on ART Initial HIV prevalence (1985) Sexual mixing Age pattern of sexual contacts Condom use % of adults who are 'high risk'	Provincial modelling report
2. Calibrate to provincial paediatric data	Total numbers of children on ART HIV prevalence in household surveys	Rates of ART initiation Duration of breastfeeding	Provincial modelling report
3. Calibrate to HIV testing data	% of adults ever tested HIV prevalence in adults tested	Rates of HIV testing in adults	Appendix B
4. Calibrate to key population prevalence data	HIV prevalence data in FSWs HIV prevalence data in MSM	Client-to-FSW and male-to-male transmission probabilities Effect of HIV diagnosis on entry into commercial sex	Appendix C
5. Calibrate to paediatric HIV data	HIV prevalence in household surveys Total numbers of children on ART Age distribution of children on ART Numbers of deaths in children Numbers of HIV tests in children HIV prevalence in children tested % of child deaths with HIV diagnosis	Fertility in HIV-diagnosed women Rates of mother-to-child transmission of HIV Paediatric disease progression Mortality after ART initiation in children Paediatric HIV testing	Appendix E
6. Calibrate to adult mortality and HIV prevalence data	HIV prevalence in antenatal and household surveys Recorded numbers of adult deaths ARV metabolite data	Heterosexual transmission probabilities per sex act Initial HIV prevalence (1985) Adult HIV disease progression ART impact on mortality Viral suppression on ART Sexual behaviour	Section 7

ART = antiretroviral treatment. FSW = female sex worker. MSM = men who have sex with men.

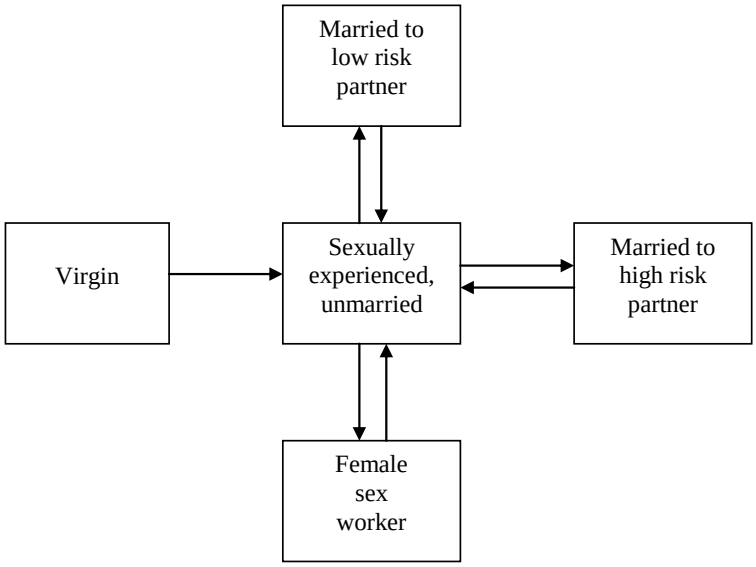
Two versions of the Thembisa model have been developed: one programmed in C++ and one programmed in Excel and Visual Basic. This report describes the C++ version of the model. Although the Excel model is almost identical, there are a few minor technical differences, and the outputs of the Excel model are therefore not exactly the same as those of the C++ model. The Excel version of the model is freely available from the Thembisa website (www.thembisa.org), together with a user guide. The C++ model, although not programmed in a user-friendly format, is available on request.

2. Model of sexual behaviour

The population aged 10 and older is divided into two broad risk groups: a high-risk group and a low-risk group. The high-risk group is defined as all individuals who have a propensity to engage in concurrent sexual partnerships and/or commercial sex, while the low risk group consists of individuals who are serially monogamous (i.e. never having more than one partner at a point in time). Within each risk group individuals are further stratified according to whether they are sexually experienced or virgins, married/cohabiting or unmarried, and (if they are married) the risk group of their married partner. Unmarried women in the high-risk group are further classified according to whether or not they are sex workers, and unmarried men are further stratified according to whether they engage in same-sex activity. There are thus three types of relationship considered in the model: long-term relationships (marital/cohabiting), short-term relationships (non-marital and non-cohabiting) and contacts between sex workers and their clients. The model makes various assumptions about the rates at which people move between different relationship states, and patterns of sexual mixing between different groups. Figure 2.1(a) illustrates the possible transitions for women in the high-risk group (similar transitions are defined for women in the low risk group, but the sex worker state is omitted). It is implicitly assumed that women only engage in heterosexual relationships (although this assumption is obviously incorrect, female same-sex relationships carry negligible HIV transmission risk and are therefore not considered in the model).

Figure 2.1(b) illustrates the possible transitions for men in the high-risk group (the same transitions are defined for men in the low risk group). Unmarried men are divided into those who are heterosexual (having sex only with women) and those who are bisexual (having sex with both men and women). For the sake of simplicity, we do not distinguish ‘gay’ and ‘bisexual’ men, as the vast majority of South African men who have sex with men (MSM) report having had sex with women [7-10]. It is also assumed, in the interests of simplicity, that bisexual men only enter into long-term relationships with female partners, as surveys of South African MSM show that only about 20% report being in marital/cohabiting relationships [8, 11]. Once bisexual men enter into long-term relationships with female partners, they are assumed to cease sexual activity with other men. Although this assumption is unrealistic, it ensures some degree of consistency with the low rates of marriage noted previously.

(a) High risk females



(b) High risk males

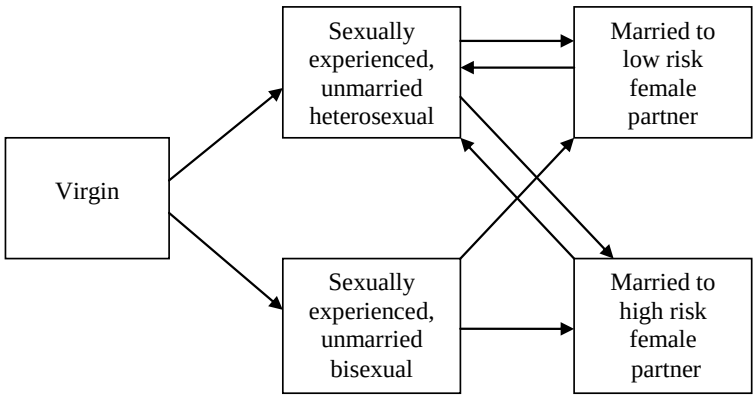


Figure 2.1: Transitions between relationship states

Table 2.1 summarizes the major sexual behaviour parameters. These parameters are explained in more detail in subsequent sections. The assumed initial proportions of men and women in the high risk group have been set at 35% and 25% respectively, based on studies that have estimated proportions of adults ever engaging in concurrent partnerships [12-14].

Table 2.1: Sexual behaviour assumptions

Parameter	Men*	Women	Reference
Initial % of population in high-risk group	35%	25%	[12-14]
Median age at sexual debut: high-risk	17.5	16.5	Calibrated
Log-logistic shape parameter for time to sexual debut	6.0	7.0	
Relative rate of short-term partnership formation in married high risk adults (compared to unmarried high risk)	0.33	0.14	Calibrated (see [15])
Relative rate of short-term partnership formation in unmarried low risk adults (compared to unmarried high risk)	0.37	0.16	Calibrated (Thembisa v2.5)
Mean age difference between partners in short-term relationships	-	3	[16-20]
Standard deviation of age difference in short-term relationships	-	3	
Mean age difference between partners in long-term relationships	-	6	[21]
Standard deviation of age difference in long-term relationships	-	5	

* Male parameters are determined from female parameters in those cases where male parameters are not shown.

2.1 Age at sexual debut

In modelling sexual debut, it is assumed that the youngest age at which sexual activity can begin is age 10, and that the time to starting sexual activity after age 10 follows a log-logistic distribution in high risk individuals. Separate log-logistic parameters are specified for males and females (Table 2.1). We assume that at each age the rate of starting sexual activity in the low risk group is 0.58 times that in the high risk group [22-27]. These parameters were chosen to yield estimates of the proportion sexually experienced at each age roughly consistent with the age-specific data from three national surveys [18, 28, 29], as demonstrated in Figure 2.2. Rates of sexual debut are assumed to be the same for heterosexual and bisexual men [30, 31].

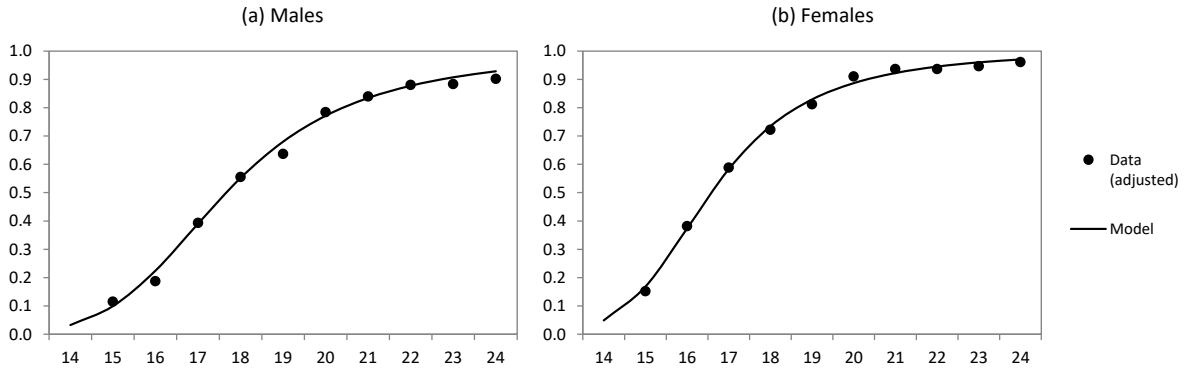


Figure 2.2: Proportion of youth who are sexually experienced, by age and sex
Data in panel (b) have been adjusted to reflect probable under-reporting of sexual experience by young women (assuming that the odds ratio relating true sexual experience to reported average sexual experience is 2 [32]).

2.2 Rates at which non-marital partnerships are formed

We define $c_{g,i,l}(x)$ to be the annual rate of non-marital partnership formation in individuals aged x, of sex g and marital status l, who are in risk group i. The female rates of partnership formation at different ages are modelled using a scaled gamma density of the form

$$c_{2,i,l}(x) = c_{2,i,l}(20)\frac{\lambda^{\alpha}(x-10)^{\alpha-1}\exp(-\lambda(x-10))}{\lambda^{\alpha}10^{\alpha-1}\exp(-10\lambda)}, \quad (2.1)$$

where the λ and α parameters determine the mean and variance of the gamma distribution, and the offset of 10 years is included to prevent sexual activity below age 10. The $c_{2,i,l}(20)$ value is 3.3 for women in the high risk group (i = 1) who are unmarried (l = 0), based on previous modelling of rates of partnership formation in South Africa [33]. The mean and standard deviation of the gamma density are uncertain; in the previous version of Thembisa (version 4.2), different posterior estimates were obtained for each of the 9 provinces. The average of the gamma means was 38.7 years (standard deviation 4.1 years), and the average of the gamma standard deviations was 19.3 years (standard deviation 2.8 years) [34]. We have therefore represented the uncertainty regarding the gamma mean and standard deviation using gamma prior distributions, with means and standard deviations equal to those estimated from the previous provincial fits. For each sampled value of the gamma mean and standard deviation, λ and α parameters are calculated to be consistent with these values.

For unmarried individuals in the low risk group, the rate of non-marital partnership is assumed to be

$$c_{g,2,0}(x) = L_g c_{g,1,0}(x), \quad (2.2)$$

where L_g is the ratio of the rate of non-marital partnership formation in the low risk group to that in the high-risk group. Because the low risk group is defined to consist of individuals who do not engage in concurrent partnerships, it might be expected that the rate of partnership formation would be lower in the low risk group than in the high-risk group. L_1 and L_2 values have been set to 0.3719 and 0.1621, the values previously estimated in the calibration of Thembisa version 2.5 (Table 2.1).

For married individuals in the high-risk group, of sex g, the rate of non-marital partnership formation is assumed to be

$$c_{g,1,1}(x) = R_g c_{g,1,0}(x), \quad (2.3)$$

where R_g is the ratio of the rate of non-marital partnership formation in married high risk individuals to that in unmarried high-risk individuals. Values of R_g have been set to 0.33 for males and 0.14 for females (Table 2.1), based on values previously fitted using the STI-HIV Interaction model [15]. No non-marital partnership formation is modelled in married low risk individuals, as the low risk group would (by definition) not engage in concurrent partnerships.

Finally, male rates of non-marital relationship formation are calculated to be consistent with the assumed rates at which females form new non-marital partnerships. Further mathematical details are provided in Appendix A. Rates of non-marital relationship formation are assumed to be the same in heterosexual and bisexual men, in the absence of reliable data comparing the two. Although this is probably not realistic, it ensures that overall coital frequencies are the same in heterosexual and bisexual men, which is consistent with studies that have found

coital frequencies in MSM [35] to be similar to the coital frequencies we have assumed for heterosexual men.

2.3 Marriage and divorce

The model defines individuals as ‘married’ if they are legally married or living together with their main partner. Rates of marriage and divorce, by age and sex, are assumed to be the same as those assumed in previous modelling work [33], based on proportions of the population reporting that they are married or living with their main partner, in the 1996 and 2001 censuses and 2007 Community Survey. Rates of divorce are estimated from published divorce statistics in 2004 [36], applying a multiple of 2 to the crude rates to reflect known biases in divorce statistics [37]. Age-specific rates of marriage and divorce are shown in Table 2.3.

Table 2.3: Age-specific behavioural parameters

Age	Annual rate of marriage		Annual rate of divorce		Annual rate of sex worker contact in unmarried high risk males	Proportion of sex workers at each age
	Males	Females	Males	Females		
15	0.0000	0.0000	0.0000	0.0000	0.02	0.9%
16	0.0026	0.0073	0.0000	0.0033	0.07	1.4%
17	0.0043	0.0224	0.0009	0.0071	0.20	2.0%
18	0.0058	0.0354	0.0047	0.0104	0.50	2.6%
19	0.0080	0.0465	0.0081	0.0134	1.06	3.2%
20	0.0123	0.0562	0.0112	0.0161	2.02	3.7%
21	0.0197	0.0650	0.0139	0.0183	3.50	4.1%
22	0.0313	0.0730	0.0165	0.0201	5.62	4.5%
23	0.0475	0.0807	0.0188	0.0215	8.44	4.7%
24	0.0674	0.0879	0.0211	0.0226	12.01	4.8%
25	0.0890	0.0943	0.0230	0.0233	16.27	4.9%
26	0.1090	0.0993	0.0246	0.0237	21.14	4.9%
27	0.1235	0.1022	0.0257	0.0241	26.46	4.8%
28	0.1302	0.1025	0.0262	0.0244	32.05	4.7%
29	0.1309	0.1008	0.0262	0.0245	37.68	4.5%
30	0.1297	0.0980	0.0259	0.0246	43.14	4.2%
31	0.1290	0.0949	0.0255	0.0244	48.21	4.0%
32	0.1278	0.0918	0.0250	0.0239	52.72	3.7%
33	0.1268	0.0891	0.0245	0.0230	56.50	3.4%
34	0.1259	0.0868	0.0240	0.0219	59.46	3.2%
35	0.1241	0.0841	0.0233	0.0206	61.52	2.9%
36	0.1215	0.0811	0.0225	0.0193	62.67	2.6%
37	0.1187	0.0780	0.0217	0.0182	62.93	2.4%
38	0.1161	0.0749	0.0210	0.0177	62.36	2.1%
39	0.1135	0.0718	0.0204	0.0174	61.02	1.9%
40	0.1108	0.0686	0.0197	0.0172	59.02	1.7%
41	0.1079	0.0655	0.0190	0.0168	56.47	1.5%
42	0.1051	0.0626	0.0182	0.0163	53.48	1.3%
43	0.1027	0.0601	0.0175	0.0154	50.17	1.2%
44	0.1004	0.0578	0.0167	0.0144	46.64	1.0%
45	0.0982	0.0556	0.0160	0.0133	42.99	0.9%
46	0.0961	0.0535	0.0153	0.0123	39.31	0.8%
47	0.0941	0.0513	0.0145	0.0113	35.67	0.7%
48	0.0924	0.0491	0.0138	0.0105	32.14	0.6%
49	0.0908	0.0469	0.0131	0.0098	28.76	0.5%
50	0.0893	0.0448	0.0124	0.0091	25.57	0.4%
51	0.0879	0.0428	0.0117	0.0083	22.59	0.4%
52	0.0866	0.0408	0.0110	0.0077	19.85	0.3%
53	0.0853	0.0388	0.0102	0.0071	17.34	0.3%
54	0.0842	0.0369	0.0093	0.0065	15.06	0.2%
55	0.0831	0.0351	0.0085	0.0060	13.02	0.2%
56	0.0821	0.0333	0.0077	0.0055	11.20	0.2%
57	0.0812	0.0315	0.0070	0.0050	9.59	0.2%
58	0.0803	0.0297	0.0064	0.0045	8.18	0.1%
59	0.0794	0.0280	0.0058	0.0040	6.94	0.1%

Although the model allows for sexual activity at ages 60 and older, assumptions are not shown.

2.4 Commercial sex

Sexually experienced heterosexual men in the high-risk group are assumed to visit sex workers at annual rate $w_l(x)$, which depends on their current age (x) and marital status (l). It

is assumed that the rate of visiting sex workers is reduced by a factor of 0.25 in married men (l = 1) [38] and that the effect of age is determined by a gamma scaling function with parameters λ_1 and α_1.The formula used to determine the rate of male contact with sex workers is thus

$$w_l(x) = K \frac{\lambda_1^{\alpha_1}(x-10)^{\alpha_1-1}\exp(-\lambda_1(x-10))}{\lambda_1^{\alpha_1}(21.5-10)^{\alpha_1-1}\exp(-\lambda_1(21.5-10))}0.25^l$$
$$= K \times \left(\frac{x-10}{11.5}\right)^{\alpha_1-1} \exp(-\lambda_1(x-21.5)) \times 0.25^l, \quad (2.4)$$

where K is the rate at which unmarried men aged 21.5 visit sex workers. (The offset of 10 is applied to age x to prevent boys below age 10 from having contact with sex workers, and the age of 21.5 was chosen previously because it corresponded to the average age of male military recruits who were asked about their rate of contact with sex workers [39].) The parameters λ_1 and α_1 are set at 0.37 and 11.1 respectively. With these parameters, the model simulates a client age distribution in 1995 that has a mean of 35.0 years and a standard deviation of 7.9 years, roughly consistent with observed client age distributions in the early stages of South Africa's HIV epidemic [40, 41] as well as a more recent survey (Tim Lane, personal communication [42]). Finally, the K parameter has been set to 3.5, which ensures that the total male demand for commercial sex is roughly consistent with the number of South African sex workers estimated in a recent national study [43], assuming that the average sex worker has 750 client contacts per annum [44-51]. (Some downward adjustment is made to the survey estimate to take into account differences in definitions of commercial sex.) The model estimates substantial age variation in the rate at which men visit sex workers, with the rate reaching as high as 63 contacts per annum in unmarried high-risk males aged 37 (Table 2.3). Bisexual men and men in the low risk group are assumed to have no contact with sex workers.

Women are assumed to enter commercial sex only from the unmarried high-risk group (Figure 2.1), with the rate of entry determined to be sufficient to meet the male demand for commercial sex. The rate of entry into commercial sex is also assumed to vary in relation to age, with the age-specific rates being determined in such a way that the age distribution of the sex worker population remains constant over time. This distribution is assumed to be of gamma form, with mean 29 years and standard deviation 9 years (Table 2.3), based on surveys of South African sex workers [16, 46, 48-55]. Women are assumed to retire from commercial sex at a rate of 0.33 per annum [46, 47, 49].

2.5 Preferences regarding partner risk group

Mixing between the high- and low-risk groups is determined by a sexual mixing parameter, ε. This parameter takes on values between 0 and 1, 0 implying completely assortative sexual mixing (i.e. individuals only choose sexual partners from their own risk group), and 1 implying random sexual mixing (i.e. individuals have no preferences regarding the risk group of their partners and choose partners in proportion to their availability) [56]. The ε parameter is difficult to determine from empirical data, and we have therefore assigned a beta prior distribution to represent the uncertainty around this parameter. The mean and standard

deviation of this prior distribution are 0.53 and 0.12 respectively, based on the distribution of values estimated when the model was previously fitted for each of the 9 provinces [34]. The same mixing parameter is assumed to apply in the selection of heterosexual and same-sex partners.

2.6 Preferences regarding partner age

The symbol $f_{g,l}(y \mid x)$ represents the probability that, for an individual of sex *g* and age *x*, in a relationship of type *l*, the partner's age is *y*. We model female age preferences regarding married partner ages using gamma distributions. For married women aged *x*, the *preferred* age distribution of the marital partners is assumed to have a mean of (*x* + 6), and a standard deviation of 5 years. This gamma distribution is adjusted to take into account relative numbers of available men at different ages. These assumptions yield marital partner age distributions consistent with those observed in the 1998 Demographic and Health Survey (DHS). Mathematically, the probability that a married woman aged *x* has a husband between the ages of *y* and *y* + 1 is

$$f_{2,1}(y \mid x) \equiv \int_{y}^{y+1} \frac{\left(\lambda_2(x)+\xi(x)\right)^{\alpha_2(x)}\left(t-\min(x)\right)^{\alpha_2(x)-1}}{\Gamma\left(\alpha_2(x)\right)\exp\left(\left(\lambda_2(x)+\xi(x)\right)\left(t-\min(x)\right)\right)}dt\,, \qquad (2.5)$$

where $\lambda_2(x)$ and $\alpha_2(x)$ are the parameters of the gamma distribution (calculated from the mean, variance and minimum age), $\xi(x)$ is the average rate of decline in the number of available men per year of increase in age (for women aged *x*), and min(*x*) = 17 + (*x* – 17)/2 for $x \geq 17$ (so that the origin of the gamma distribution is at this minimum age and not at zero, in order to prevent unrealistically low married male ages). The $\xi(x)$ parameters are updated dynamically each year as the population pyramid changes.

For women who are aged *x* and in non-marital relationships, the age distribution of non-marital partners is assumed to have a mean of (*x* + 3), and a standard deviation of 3 years, consistent with partner age distributions reported by young women in various South African studies [16-20]. As for marital relationships, this distribution is adjusted to take into account the actual number of men available at each age. In sex worker-client contacts, clients and sex workers are assumed to have no age preferences.

Proportions of men who choose their female partners from different ages ($f_{1,l}(y \mid x)$) are calculated to be consistent with the distributions specified for women, taking into account the relative rates of partnership formation at different ages, and relative numbers of men and women at different ages. Further mathematical detail is provided in Appendix A. Appendix A also describes the approach to modelling age mixing patterns in MSM ($f_{1,4}(y \mid x)$).

2.7 Coital frequencies

The average number of sex acts per non-spousal relationship is assumed to be 18. This is consistent with an average coital frequency of 3 acts per month in non-spousal relationships

[18, 20, 57, 58] and an average non-marital relationship duration of 6 months [33]. In marital relationships, the frequency of sex is assumed to vary in relation to individuals' age and sex. For married women who are aged 20, the average number of spousal sex acts per month is assumed to be 5, and this number is assumed to halve for each 20-year increase in age [33].

2.8 Condom usage

2.8.1 Condom use in women (excluding commercial sex)

Rates of condom use are assumed to depend on age, sex, type of relationship and knowledge of HIV-positive status. Rates of condom usage are also assumed to change over time; this time-dependency represents the effect of HIV communication programmes and condom promotion campaigns, which were introduced in the 1990s and early 2000s. The parameter $\gamma_{2,l}(x,t)$ represents the probability that an HIV-negative woman aged x uses a condom in an act of sex with a partner of type l at time t (time is measured in years since 1985). This is calculated as

$$\gamma_{2,l}(x,t) = \zeta(t)\, \nu^{(x-20)}\, \beta_l\, \theta_r, \qquad (2.6)$$

where $\zeta(t)$ represents the time trend in condom use, ν is the factor by which condom use decreases per year of increase in age, β_l is the relative rate of condom use in relationship type l, and θ_r is a scaling parameter, which we include to allow for the possibility of bias in self-reported condom use data (the value depends on the type of reporting, r). The 'base rate' of condom use, $\zeta(t)$, relates to women aged 20 who are unmarried ($l = 0$) and reporting on their condom use at last sex ($r = 0$), and the β_0 and θ_0 parameters are therefore both set to 1. The $\zeta(t)$ function is a linear combination of a constant term and two cumulative Weibull distribution functions. The constant term (k_0) represents the initial rate of condom usage, prior to the start of the HIV epidemic in South Africa, the first Weibull distribution corresponds to the increase in condom usage following the introduction of behaviour change communication programmes in the mid-1990s, and the second Weibull distribution represents a possible change in condom usage rates in recent years. In mathematical terms,

$$\zeta(t) = k_0 + k_1\left(1 - 0.5^{(t/m_1)^{\phi_1}}\right) + k_2\left(1 - 0.5^{(t/m_2)^{\phi_2}}\right). \qquad (2.7)$$

The k_1 parameter represents the extent of the increase in condom use following the early phase of the HIV communication programmes, and the m_1 and ϕ_1 parameters represent the median and shape parameters respectively of the first Weibull distribution. The k_2 parameter represents the extent of the change in condom use in recent years (possibly due to changes in funding for behaviour change communication programmes, and possibly due to changes in attitudes towards condom use as ART has become more widely available); the m_2 and ϕ_2 parameters represent the median and shape parameters respectively of the second Weibull distribution. We have fixed $m_2 = 2m_1$ and $\phi_2 = 2\phi_1$, similar to the previous version of Thembisa, in order to simplify the model calibration.

These rates of condom use are adjusted in HIV-diagnosed individuals and individuals on ART. In a diagnosed but ART-naïve individual of sex g, the probability of condom use is calculated as $1 - (1 - \gamma_{g,l}(x,t))(1 - \delta)$, where δ is the proportionate reduction in unprotected sex after HIV diagnosis (as discussed in section 2.10). In ART experienced individuals, the probability of condom use is $1 - (1 - \gamma_{g,l}(x,t))(1 - \delta)(1 - h)$, where h is the proportionate reduction in unprotected sex after ART initiation (as discussed in section 2.11). We have fixed $\delta = 0.61$ (the posterior mean estimated in the previous version of Thembisa) and $h = 0.18$ (see section 2.11) for the purpose of the analyses that follow.

We estimate the parameters in equations (2.6) and (2.7) by fitting the model to self-reported condom usage data from national surveys conducted in South African women. For the purpose of this model fitting we do not include male data on condom usage because it is more difficult to determine the relationship type from male data (for example, if a married man reports that he used a condom the last time he had sex, it is not clear if this occurred in the context of his spousal relationship, another relationship with a non-spousal partner, or a contact with a sex worker). Another reason for not using the male data in calibration is that condoms are perceived to be a male-controlled form of contraception, and thus men might be more inclined to over-report their condom use (due to social desirability bias). Male rates of condom use are instead calculated as a function of female rates of condom use, as described in section 2.8.2.

For the purpose of the model fitting, we consider two types of self-reported condom use data: data on the reported use of condoms at last sex ($r = 0$) and data on use of condoms for contraceptive purposes ($r = 1$). In the case of the former, we set $\theta_r = 1$, so that the 'base rates' correspond to those that would be estimated if we believed reporting of condom use at last sex to be an accurate reflection of the proportion of sex acts that are protected. In the case of the latter, we estimate a value of θ_r less than 1, assuming that there is less likely to be bias towards over-reporting condom use in the context of contraceptive use (since condoms are generally less effective than hormonal methods of contraception).

A total of 71 national survey estimates of women's self-reported condom use were obtained, from 12 surveys conducted between 1986 and 2017. Of these, 49 related to condom use at last sex and 22 related to condom use for contraceptive purposes. All but 5 of the survey estimates combined married and unmarried women; the remaining 5 related only to women with non-cohabiting partners. Data were disaggregated by 5-year age group wherever possible; when this was not possible, broader age groupings were used.

For the purpose of fitting the model in equations (2.6) and (2.7) to the survey data, we use estimates of numbers of women who are married and unmarried, stratified by HIV diagnosis and receipt of ART, by age and by year, as estimated by the previous version of Thembisa (version 4.2) [2]. If $n_l(x, t)$ is the Thembisa estimate of the number of sexually-experienced women of marital status l, aged x in year t, and $d_l(x, t)$ and $a_l(x, t)$ are the corresponding proportions diagnosed positive and ART-experienced respectively, then the model estimate of condom use that is compared to the survey estimate is

$$\frac{\sum_{l=0}^{1} n_l(x,t)\left[\left(1-d_l(x,t)\right)\gamma_{2,l}(x,t)+\left(d_l(x,t)-a_l(x,t)\right)\left\{1-\left(1-\gamma_{2,l}(x,t)\right)\times\left(1-\delta\right)\right\}+a_l(x,t)\left\{1-\left(1-\gamma_{2,l}(x,t)\right)\left(1-\delta\right)\left(1-h\right)\right\}\right]}{n_0(x,t)+n_1(x,t)}. \quad (2.8)$$

For the sake of simplicity, we take x as the mid-point of the age range to which the survey estimate relates (except in the case of survey estimates for the 50+ age range, which we arbitrarily assign to age 60 for calibration purposes).

For the purpose of fitting the model to the data, we defined a simple binomial likelihood to represent the probability of generating the survey data given the model estimate of the 'true' rate of condom use. However, because the sample sizes were not published in many cases, and because most of the published rates of condom use did not include 95% confidence intervals, it was necessary to approximate the likelihood by assuming an effective sample size of 50 for each year in the age group (for example, an effective sample size of 250 for a 5-year age group). This is based roughly on the sample size in the DHSs and a design effect of 1.63 (i.e. taking into account that due to variation across clusters, the effective sample size is smaller than the true sample size). The likelihood was maximized, and the resulting estimates of the model parameters are summarized in Table 2.4. Although previous versions of Thembisa made provision for significant 'risk compensation' in recent years, this updated analysis suggests that the extent of the reduction in condom use in recent years is in fact very modest (k_2 = -0.02).

Table 2.4: Maximum likelihood estimates of condom use parameters

Parameter	Symbol	Value
Self-reported condom use at last sex in unmarried women aged 20		
Initial rate in 1985	k_0	0.044
Increase in rate due to BCC and condom distribution	k_1	0.579
Change in condom use in recent years	k_2	-0.02
Relative rate of condom use per year increase in age	ν	0.982
Relative rate of condom use in marital relationships	β_1	0.242
Relative rate of reporting for contraceptive purposes	θ_1	0.343
Median time to behaviour change (in years after 1985)	m_1	15.8
Shape parameter for time to behaviour change	ϕ_1	6.71

BCC = behaviour change communication.

Figure 2.3 shows the extent of the consistency between the model predictions and the survey estimates of condom use. Figure 2.4 further shows the time trend in condom use, for women aged 15-24, and the difference in condom usage when comparing reporting of condom use at last sex with reporting of condom use for contraceptive purposes.

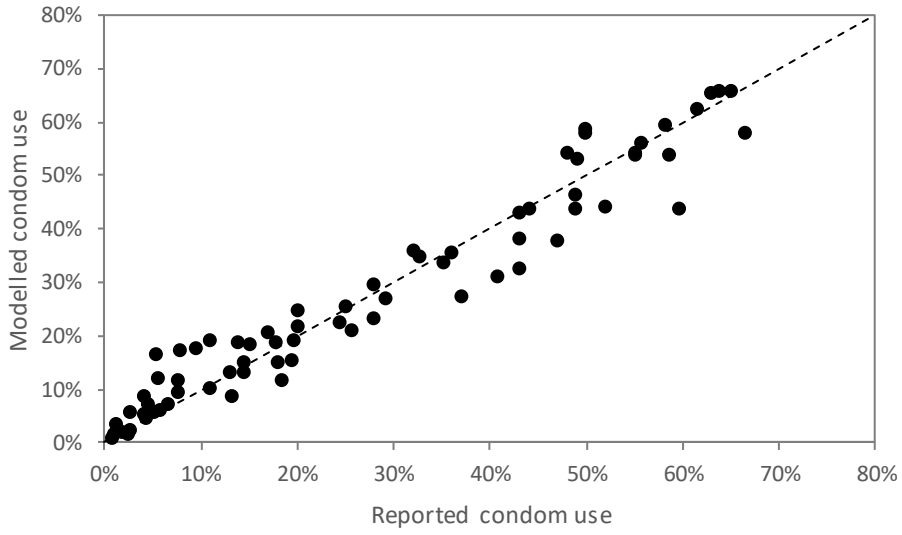


Figure 2.3: Comparison of model and survey estimates of condom use
Each dot represents one of the 71 data points used in model calibration. The dashed line represents the line of equality between the model and survey estimates.

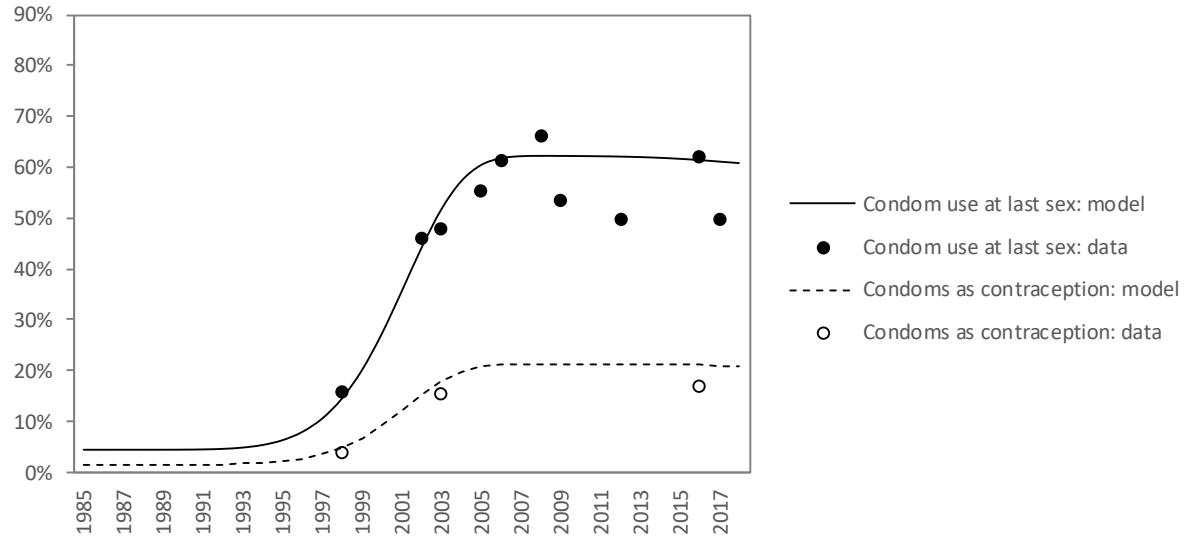


Figure 2.4: Condom use in women aged 15-24
For the sake of comparison, model estimates are shown for women aged 20 who are unmarried and not diagnosed positive (at this age relatively few women are married and few have been diagnosed HIV-positive). Survey estimates combine data for the 15-19 and 20-24 age groups, although these are in many cases entered as separate data points for calibration purposes.

When calibrating the Thembisa model to adult HIV prevalence data and mortality data, we allow for uncertainty regarding the extent of the bias in the self-reported condom use data. We assume that the true rate of condom use is somewhere between the reported rate of condom use at last sex (θ_0) and reported rate of condom use for contraceptive purposes (θ_1). If r represents the extent of the bias in the self-reported condom use at last sex, then the 'true' condom adjustment factor is $\theta = \theta_0(1 - r) + \theta_1 r$, where r lies in the range (0, 1). We assign a uniform (0,1) distribution to represent the uncertainty in r.

2.8.2 Condom use in men (excluding commercial sex)

To ensure that male and female assumptions are consistent, the probability that an HIV-negative man uses a condom in a marital or non-marital relationship is calculated as

$$\gamma_{1,l}(x,t) = \sum_{y} f_{1,l}(y \mid x)\gamma_{2,l}(y,t), \tag{2.9}$$

where $f_{1,l}(y \mid x)$ is the probability that a female partner is aged *y*, if the male partner is aged *x*. It is assumed that the rate of condom use in same-sex relationships is the same as that in heterosexual relationships [30, 31, 59]. The effect of HIV diagnosis and ART initiation on condom use is the same as assumed in women.

2.8.3 Condom use in sex worker-client relationships

We define $\gamma_{2,2}(x,t)$ as the probability that an HIV-negative sex worker aged *x* uses a condom in an act of sex with a client at time *t* (time is measured in years since 1985). This is modelled as

$$\log\left(\frac{\gamma_{2,2}(x,t)}{1-\gamma_{2,2}(x,t)}\right) = \log\left(\frac{\gamma_{2,0}(20,t)}{1-\gamma_{2,0}(20,t)}\right) + \log(\beta_2), \tag{2.10}$$

where β_2 is the ratio comparing the odds of condom use in HIV-negative sex workers to that in HIV-negative women aged 20 in non-marital relationships. (We choose to work on an odds ratio scale rather than a relative risk scale to avoid possible condom usage rates in excess of 100%.) Note that there is no dependence on *x* in the right-hand side of the equation, i.e. we are assuming the rate of condom use in sex workers is the same at all ages, due to lack of age-specific data on condom use in sex worker-client interactions. The effects of HIV diagnosis and ART initiation on condom use are the same as assumed previously (on a relative risk scale).

A challenge in estimating β_2 is that no nationally representative sex worker data are available. However, by comparing sex workers' reporting of condom use at last sex in different settings with that reported by women in the general population, we can approximate the average odds by which condom use increases in the context of commercial sex. Table 2.5 summarizes the available data on sex workers' self-reported condom use at last client contact (we have not included data in which the type of relationship is unspecified, as a substantial proportion of sex workers have regular partners). These rates have been adjusted (based on equation 2.8) to represent the rates that would be expected in HIV-negative sex workers, based on Thembisa estimates of the fraction of sex workers who are HIV-diagnosed and on ART (fourth column). The final column shows the study-specific odds ratio relating the adjusted condom use in HIV-negative sex workers to $\zeta(t)$. These odds ratios are highly variable between studies (range 1.2-44.8), reflecting heterogeneity across samples within South Africa. We have set β_2 to the median of these values, 7.0.

Table 2.5: Condom use by sex workers in contacts with clients

Study	Year	Reported condom use	HIV-negative condom use*	Condom use with short-term partners ($\zeta(t)$)	Odds ratio (β_2)
Peltzer *et al* [49]	2002†	85.7%	84.6%	44.0%	7.0
Richter *et al* [55]	2010	94.5%	92.6%	62.2%	7.6
SWEAT [60]	2012	95%	92.8%	62.1%	7.8
Delva *et al* [51]	2010	99%	98.7%	62.2%	44.8
UCSF, Anova, WRHI [42]	2013	76.4%	65.3%	62.0%	1.2
	2013	89.4%	84.4%	62.0%	3.3
	2013	84.5%	77.2%	62.0%	2.1
Median					7.0

* Corresponding to $\gamma_{2,2}(x,t)$, estimated from the reported rates using a formula similar to that in equation 2.8. † Date not specified – a date two years prior to publication has been assumed.

2.9 Effect of CD4 count on level of sexual activity

The model assumes that coital frequencies in HIV-positive individuals decline as they enter more advanced stages of HIV disease. It is assumed that the frequency of sex in HIV-positive adults with CD4 counts ≥500/μl is the same as would be expected in HIV-negative adults with the same characteristics. The frequency of sex is assumed to be reduced by 8% in individuals with CD4 counts of 350-499/μl, by 24% in individuals with CD4 counts of 200-349/μl, and by 45% in individuals with CD4 counts of <200/μl (relative to individuals with CD4 counts of ≥500/μl in all cases). These assumptions are based on meta-analyses of various studies that have assessed either differences in sexual behaviour or differences in the incidence of pregnancy between CD4 stages [61-67]; results of the individual studies are shown in Figure 2.4.

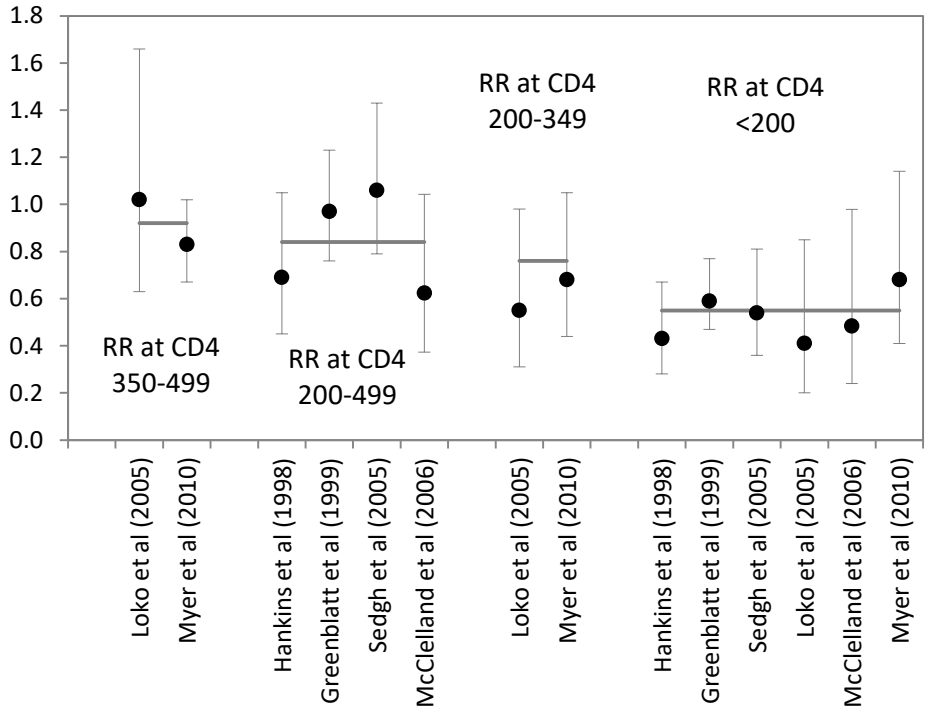


Figure 2.4: Comparison of model assumptions about relative frequency of sex at different CD4 levels and empirical estimates

Model assumptions are represented by horizontal grey lines. Empirical estimates are represented by dots (error bars represent 95% confidence intervals). Note that the model assumption for the CD4 200-499/µl category is taken as the average of that in the 350-499/µl and 200-349/µl categories. For convenience, we have treated the Sedgh *et al* (2005) estimates as if they are based on CD4 cut-offs of 200 and 500 (not 250 and 500).

It is assumed that the frequency of sex is the only sexual behaviour parameter that changes in relation to the CD4 count in HIV-infected adults. In the interests of simplicity, we do not model the possible effect of the CD4 count on rates at which new partnerships are formed, rates of partnership dissolution or rates of condom usage. However, in high risk women, it is assumed that rates of entry into commercial sex are reduced by 12% at CD4 counts of 350-499, by 35% at CD4 counts of 200-349 and by 60% at CD4 counts of <200 cells/µl. Rates of exit from commercial sex are increased by factors that are inversely related to these reduction factors (for example, a sex worker with a CD4 count <200/µl is assumed to cease commercial sex at a rate that is $1/(1 - 0.6) = 2.5$ times that in HIV-negative sex workers). These assumptions are consistent with data from sex workers in Kenya [68], who were found to be significantly more likely to abstain from sex at lower CD4 counts (OR 1.70 for CD4 counts of 200-499 and 2.39 for CD4 counts of <200). It is also assumed that the frequency at which men visit sex workers is reduced by the same factors as those used to reduce coital frequencies in short-term and long-term relationships.

2.10 Effect of knowledge of HIV status on sexual behaviour

Most evidence suggests that HIV testing does not significantly affect sexual behaviour or HIV incidence in individuals who receive negative test results [69-72], and the model therefore assumes no change in behaviour following an HIV-negative test result. However,

studies from developing countries show that HIV-positive diagnoses usually lead to significant declines in unprotected sex, with the reductions in the odds of unprotected sex varying between 10% and 95% (average reduction 57%, based on a random effects meta-analysis of the estimates in Table 2.5).

Table 2.5: Studies evaluating the effect of HIV diagnosis on sexual risk behaviour in developing countries

Study	Location	Definition of risk behaviour	Controls	Effect on risk behaviour in HIV-diagnosed (OR, 95% CI)
Marlow *et al* [73]	South Africa	No condom use at 14 weeks postpartum	HIV-negative women	0.59 (0.48-0.72)
Ngubane *et al* [74]	South Africa	No condom use 0-12 mo postpartum 13-24 mo postpartum	HIV-negative women	 0.58 (0.47-0.72) 0.62 (0.44-0.87)
Morroni *et al* [75]	South Africa	No condom use at last sex	Women at FP/ STI clinics	0.28 (0.16-0.51)
George *et al* [76]	South Africa	Any unprotected sex with recent partner	HIV negative & undiagnosed	0.46 (0.37-0.57) 0.64 (0.54-0.77)
Mwangi *et al* [77]	Kenya	Any unprotected sex with a partner who was HIV-negative or of unknown HIV status	Individuals who were HIV-positive but undiagnosed	0.05 (0.02-0.12)
Voluntary HIV-1 Counselling and Testing Efficacy Study Group [78]	Kenya, Tanzania, Trinidad	Any unprotected sex with primary partner Any unprotected sex with non-primary partner: Women Men	Individuals who tested HIV-negative	0.60 (0.40-0.89) 0.90 (0.49-1.66) 0.19 (0.05-0.81)
Müller *et al* [79]	Thailand	<100% condom use in last 3 sex acts	Individuals who were HIV-positive but undiagnosed	0.15 (0.09-0.24)
Cremin *et al* [71]	Zimbabwe	Inconsistent condom use with regular partners: Women Men	Individuals who were HIV-positive but undiagnosed	 0.53 (0.24-1.16) 0.61 (0.25-1.47)
Pooled OR				0.43 (0.33-0.56)

In all studies, with the exception of Müller *et al*, the odds ratio presented is based on multivariate analysis (Müller *et al* did not employ multivariate analysis, but did select controls who were age- and sex-matched to the cases.)

A challenge is that our model is parameterized in terms of a percentage reduction in unprotected sex after HIV diagnosis. This is effectively a relative risk (RR) measure rather than an odds ratio (OR) measure. RRs will generally be closer to 1 than ORs, especially when the outcome of interest is highly prevalent. For example, in the South African studies cited in Table 2.5, the prevalence of the outcome (unprotected sex) in the control groups varied between 58% and 90% across studies (average of around 77%). If the prevalence of the

outcome is 77% in the control group and 59% in the HIV-diagnosed group, this implies an OR of 0.43 (consistent with the pooled OR in Table 4.3) but an RR of 0.77. This means that the 0.43 can be considered a lower bound on the RR parameter in our model.

Another reason why the 0.43 factor might be considered a lower bound on the RR we seek to estimate is that most of the studies summarized in Table 2.5 do not separate HIV-diagnosed individuals according to whether they are treated or untreated. Since we assume that the effect of ART is to reduce the proportion of sex acts that are unprotected by 18% (see section 2.11), it is possible that much of the observed reduction in unprotected sex is an effect of being on ART rather than an effect of HIV diagnosis. In a worst case scenario, where the RR of 0.77 is calculated with only treated individuals in the numerator, the true effect of diagnosis would be only a 6% reduction in the frequency of unprotected sex ($1 - 0.77/(1 - 0.18)$).

To represent the uncertainty around the reduction in unprotected sex after diagnosis, we assign a beta prior with a mean of 0.18 and a standard deviation of 0.13. This is equivalent to assuming an average relative rate of unprotected sex after HIV diagnosis of 0.82, with 95% confidence interval from 0.50-0.99, i.e. substantially higher than the OR of 0.43. This prior distribution is based on a simulation that takes into account (a) the uncertainty around the OR of 0.43, (b) the uncertainty around the risk reduction after ART initiation, and (c) the uncertainty regarding the fraction of HIV-diagnosed individuals who are on ART in the numerator (for the reported odds ratios).

The model also allows for an effect of women's knowledge of their HIV status on entry into commercial sex work. It is assumed that women who have been diagnosed HIV-positive are less likely to start commercial sex than women who are HIV-positive but undiagnosed. However, as there have been no published studies quantifying the likely magnitude of this reduction, we assign a vague prior (uniform on the interval [0, 1]) to represent the proportionate reduction in the probability of entering sex work for women who have been HIV-diagnosed. Further explanation is provided in Appendix C.

2.11 The effect of ART on sexual behaviour

In our model, ART is assumed to affect the sexual behaviour of treated individuals in two ways. Firstly, by bringing about an improvement in CD4 count and restoring individuals' health [80], ART is assumed to cause an increase in the frequency of sexual activity. Secondly, because of their greater contact with health services and greater exposure to prevention messages, sexually active ART patients are assumed to have a higher level of condom usage when compared with sexually active ART-naïve patients who are HIV-diagnosed.

Coital frequencies after ART initiation are assumed to depend only on current CD4 count, as described in section 2.9, with no effect of ART after controlling for current CD4 count. This is because most African studies show that after controlling for measures of disease severity, ART does not significantly affect frequency of sexual activity [81-83].

The assumed proportion of sex acts that are protected in year t, in an HIV-treated adult of age x and sex g, in relationship type l, is

$$1-\left(1-\gamma_{g,l}(x,t)\right)\left(1-\delta(t)\right)\left(1-h\right), \qquad (2.11)$$

where $\gamma_{g,l}(x,t)$ is the corresponding rate of condom use in HIV-negative individuals (discussed in section 2.8), $\delta(t)$ represents the reduction in unprotected sex following diagnosis (discussed in section 2.10), and h represents the additional reduction in unprotected sex following ART initiation.

The h parameter has been set to 0.18. This is based on a meta-analysis [84], which found that in high-quality studies receipt of ART was associated with a significant reduction in unprotected sex (OR 0.68, 95% CI: 0.58-0.79). (Low-quality studies were excluded, as these tend not to control for time since diagnosis and thus tend to conflate the effects of HIV diagnosis and ART on levels of condom usage.) As in the previous section, it is important to consider that the OR is likely to be different from the RR, which is what we require to parameterize Thembisa. If the prevalence of unprotected sex is 45% in South African ART patients [85] and 55% in HIV-diagnosed individuals who are untreated, this implies an OR of 0.67 (consistent with the systematic review) but an RR of 0.82. We have therefore set the h parameter to 0.18 (1 – 0.82) rather than 0.32 (1 – 0.68).

2.12 Same-sex relationships

It is assumed that at the time of beginning sexual activity, 5% of men enter the 'bisexual, unmarried' group and the remainder enter the 'heterosexual, unmarried' group. This 5% assumption is based on the results of a household survey conducted in two South African provinces, which used computer-assisted interview techniques to obtain estimates of the fraction of men who had ever engaged in sex with other men [7]. Bisexual men are assumed to form 70% of their short-term relationships with other men and the remaining 30% with women. 69% is the average fraction of MSM who report *only* engaging in sex with men in the last 6 months, across three South African studies [8-10]. Although this may be an under-estimate of the fraction of partners who are male (since those who reported having sex with women were also having sex with men), it could also be an over-estimate if MSM who have less frequent sex with men are less likely to be included in the sample (recent sex with other men was a condition for inclusion in the three cited studies).

3. Model of HIV disease progression and mortality in adults

HIV-infected adults are assumed to progress through five stages of HIV infection in the absence of ART. An initial acute infection phase, lasting for three months, is followed by four stages of increasing immunosuppression (CD4 count ≥500 cells/μl, 350-499 cells/μl, 200-349 cells/μl and <200 cells/μl). Individuals are further classified according to whether they have been diagnosed HIV-positive, with rates of diagnosis changing over time and varying in relation to age, sex and CD4 stage. Adults who have been diagnosed HIV-positive are assumed to start ART at a rate that changes over time, as ART rollout expands and treatment eligibility criteria change [86]. Once individuals have started ART, they are stratified by their time since ART initiation and baseline CD4 category. The model of HIV disease progression, diagnosis and ART initiation is illustrated in Figure 3.1.

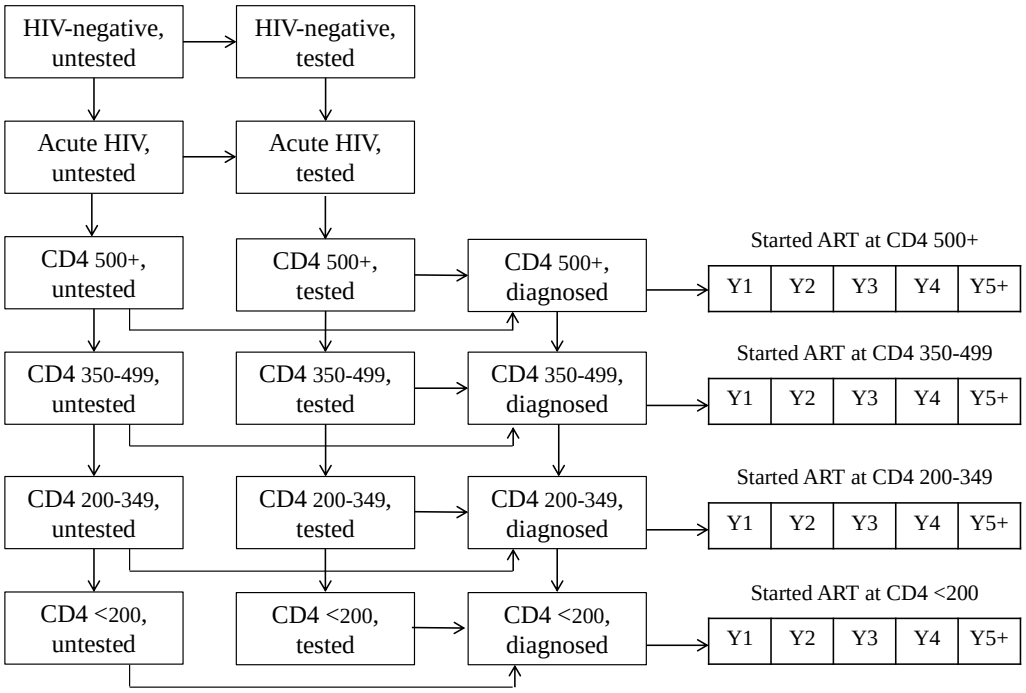


Figure 3.1: Multi-state model of survival in HIV-positive adults

It is worth noting that the model does not define a separate state to represent individuals who have interrupted ART. However, the model does calculate, for each of the times since ART initiation, the probability that the individual is on ART versus interrupting ART, and these duration-specific probabilities are used in calculating the number of adults currently on ART at any point in time. A more detailed explanation of the assumptions about rates of ART interruption is provided in Appendix G.

3.1 HIV disease progression and mortality prior to ART initiation

In untreated individuals, we define the symbol $\lambda_{g,s}(x)$ to be the annual rate of transition from HIV state *s* to state (*s* + 1) in untreated HIV-positive individuals of sex *g* (1 = males, 2 = females) who are aged *x*. This is calculated as

$$\lambda_{g,s}(x) = \lambda_s \varpi^{g-1} (1+k)^{(x-30)/10} E^{t-1999}, \quad (3.1)$$

where λ_s is the rate that applies in men aged 30 in 1999, ϖ is the factor by which HIV disease progression is adjusted in women, *k* is the proportional increase in the rate of disease progression per 10-year increase in age, and *E* is the factor by which the rate is adjusted per year as a result of changes in HIV virulence. Similarly, we define the symbol $\mu_{g,s}(x)$ to be the annual HIV-related mortality rate in HIV state *s* in untreated individuals of sex *g* who are aged *x*. This is calculated as

$$\mu_{g,s}(x) = \mu_s \varpi^{g-1} (1+k)^{(x-30)/10}, \quad (3.2)$$

where μ_s is the HIV mortality rate that applies in men aged 30. The adjustment factors for the effects of age and sex on HIV disease progression are thus the same as the adjustment factors for the corresponding effects on HIV-related mortality (except in respect of the HIV evolution parameter). HIV-positive women tend to have lower viral loads [87-89] and lower rates of CD4 decline [90] than HIV-positive men, and studies suggest a lower mortality rate in HIV-positive women than in HIV-positive men in the pre-ART era [89, 91-93]. To represent the uncertainty regarding the ϖ parameter, a gamma prior distribution has been assigned, with a mean of 0.96 and standard deviation of 0.05 [94].

Evidence suggests that increasing age is associated with both increasing rates of CD4 decline [95, 96] and increasing mortality in HIV-positive adults [97-100]. To represent the uncertainty around the *k* parameter, a gamma prior with a mean of 0.18 and standard deviation of 0.06 has been assigned [94]. A gamma prior has also been assigned to represent the uncertainty regarding the overall mean HIV survival time (mean 12 years, standard deviation 1 year), and this is used to determine λ_s and μ_s parameters (corresponding prior means are shown in Table 3.1) [94]. Assumptions about the relative lengths of time spent in different CD4 stages were determined by calibrating the model to cross-sectional surveys of CD4 distributions in HIV-positive adults [101-108], and assumptions about relative rates of mortality by CD4 stage were based on the assumption of negligible HIV-specific mortality at CD4 counts >350 cells/μl and a mortality hazard ratio of 0.13 for individuals with CD4 counts of 200-349, when compared to individuals with CD4 counts <200 cells/μl [109].

HIV virulence may be changing as a result of HIV evolution; studies from Uganda and Botswana suggest that there have been substantial reductions in HIV virulence over time [110, 111]. It is also possible that HIV virulence may have changed as a result of improvements in tuberculosis prevention, screening and treatment (for example, isoniazid preventive therapy for HIV-positive individuals and tuberculosis case finding in HIV-positive individuals [112]). However, evidence from high-income countries generally suggests a shift towards *increased* HIV virulence over time [113, 114]. Given the inconsistent estimates from the literature, a gamma prior was previously assigned to represent the uncertainty in the *E*

parameter, with a mean of 1 and a standard deviation of 0.0065 [94]. Although our original analysis estimated the value of E to be significantly less than one [94], a subsequent analysis found that after allowing more realistically for the effect of ART interruptions in the model, the E estimate was not significantly different from one [115]. We have therefore fixed $E = 1$ for the purpose of this analysis.

Table 3.1: Parameters by HIV disease stage

Parameter	Acute HIV	CD4 range 500+	350-499	200-349	<200	Source
Average time (in years) to next stage, in absence of ART* ($1/\lambda_s$)	0.25	3.16†	2.13†	3.20†	-	Calibrated
Annual HIV mortality rate, in absence of ART* (μ_s)	0.00	0.00	0.00	0.033†	0.254†	Calibrated
Annual incidence of OIs, in absence of ART						[116, 117]
All WHO stage III and IV OIs	0.05	0.05	0.12	0.27	0.90	
WHO stage IV	0.01	0.01	0.02	0.06	0.28	
Pulmonary TB	0.01	0.01	0.015	0.04	0.07	
Relative infectiousness if untreated (I_s)	10	1	1	2	7	[118-120]
Annual male HIV mortality after ART initiation, by baseline CD4‡						
1st 6 months of ART	-	0.0002	0.0016	0.0146	0.2554	[121]
Months 7-18	-	0.0009	0.0050	0.0132	0.0613	
Months 19-30	-	0.0027	0.0085	0.0116	0.0306	
Months 31-42	-	0.0042	0.0076	0.0076	0.0202	
Months 43+	-	0.0049	0.0063	0.0063	0.0166	
Annual female HIV mortality after ART initiation, by baseline CD4‡						
1st 6 months of ART	-	0.0001	0.0016	0.0159	0.2072	[121]
Months 7-18	-	0.0008	0.0045	0.0101	0.0490	
Months 19-30	-	0.0020	0.0057	0.0057	0.0235	
Months 31-42	-	0.0027	0.0034	0.0034	0.0141	
Months 43+	-	0.0025	0.0025	0.0025	0.0103	

* Parameters are specified for 30-year old males, and adjustments for age and sex are made in the process of calibrating the model to reported death data. † Prior means corresponding to average untreated survival of 12 years. ‡ Parameters are adjusted to take into account age effects, and effects of increasing baseline CD4 counts over time. OI = opportunistic infection.

3.2 HIV testing and diagnosis

As shown in Figure 3.1, the population aged 10 and older is divided into three HIV testing history groups (never tested, previously tested negative and previously tested positive). Three types of HIV testing are modelled: testing in antenatal clinics, testing of HIV patients with opportunistic infections (OIs), and testing for other reasons. The annual rate at which sexually-experienced individuals get tested is assumed to depend on their HIV stage (s), age (x), sex (g), HIV testing history (i) and the calendar year (t):

$$\tau_{g,i,s}(x,t) = b(t)A_g(x,t)r_i(t) + \Omega_s d_i(t) + F_{g,s}(x,t)v_i(t) \qquad (3.3)$$

where $b(t)$ is the base rate of HIV testing in year t, in individuals who do not have any HIV symptoms and are not pregnant; $A_g(x,t)$ is an adjustment factor to represent the effect of age and sex on the base rate of test uptake; $r_i(t)$ is an adjustment factor to represent the effect of testing history; Ω_s is the annual incidence of OIs in CD4 stage s; $d_i(t)$ is the fraction of OI

patients who are tested for HIV in year t; $F_{g,s}(x,t)$ is the fertility rate in sexually experienced women aged x, in HIV stage s, during year t (set to zero for men); and $v_i(t)$ is the proportion of pregnant women who receive HIV testing in year t. The function used to represent the effect of age and sex on the uptake of HIV testing is

$$A_g(x,t) = B_g(t)\left(\frac{x}{25}\right)^{\alpha_g - 1} \exp\left(-\sigma_g\left(x-25\right)\right), \quad (3.4)$$

where $B_g(t)$ is a time-dependent sex adjustment factor, and α_g and σ_g are coefficients for the effect of age on the rate of HIV test uptake. The testing history (i) is classified as never tested (0), last tested negative (1), tested positive but never started ART (2), or initiated ART (3). It is assumed that the effect of testing history changes over time [122], with the relative rate of testing in previously-tested individuals increasing as the frequency of testing increases. For previously-tested individuals who last tested negative, the relative rate of testing in year t is calculated as

$$r_1(t) = r_1(0) + (r_1(\max) - r_1(0)) \times b(t-1) / 0.5, \quad (3.5)$$

where $r_1(0)$ and $r_1(\max)$ are the baseline and 'maximum' relative rates of retesting (compared to individuals who have never tested). The factor of 0.5 is somewhat arbitrary, but is based on previous fits of the model to provincial HCT statistics [123]; fitted values of $b(t)$ were seldom found to increase above 0.5. For previously diagnosed individuals who are either ART-naïve (i = 2) or on ART (i = 3), the relative rate of testing in year t is calculated as

$$r_i(t) = r_i \times r_1(t) / r_1(0), \quad (3.6)$$

where r_i is the relative rate that would be expected at low rates of HIV screening, when access to HIV testing is very limited.

The parameterization of the model was originally based on data collected up to 2012 [124], but has been updated to take into account more recent HIV testing data (as described in Appendix B). Briefly, the model parameters have been estimated using a Bayesian procedure that incorporates three data sources: total numbers of HIV tests performed in the South African public and private health sectors (2002-2018), proportions of individuals testing for HIV who test positive, and proportions of adults who report previous HIV testing in five national surveys [4, 18, 19, 125, 126], stratified by age, sex and HIV status. The calibration procedure allows for potential bias in self-reporting of previous HIV testing. The prior distributions in the Bayesian analysis are based on observed patterns of HIV testing by age and sex [127, 128] and observed increases in rates of testing in previously-tested individuals [53, 129, 130]. The assumed incidence of OIs by HIV stage (Ω_s) is shown in Table 3.1, and the assumed proportions of OIs tested for HIV are shown in Table 3.2. The assumed fractions of pregnant women tested for HIV are also shown in Table 3.2, and assumptions regarding fertility rates and the effect of HIV on fertility are described in section 6.2.

Table 3.2: Assumed proportions of patients tested for HIV and linked to ART

Year	Antenatal testing ($v_i(t)$)		Testing of OI patients ($d_i(t)$)*		Linkage to ART in pregnancy ($l_2(5,t)$)	
	Rate	Sources	Rate	Sources	Rate	Sources
Pre-1999	0.0%		12%		0.0%	
1999-00	0.9%		12%		0.0%	
2000-01	2.9%		12%		1.9%	
2001-02	7.5%	[131]	12%		2.5%	
2002-03	15.6%	[132, 133]	12%		2.5%	
2003-04	31.3%		12%		3.6%	
2004-05	42.0%	[134]	18%	[135]	12.6%	
2005-06	54.5%	[136]	39%	[137]	22.6%	[138]‡
2006-07	72.2%	[139]	53%	[137]	29.1%	
2007-08	84.0%	[140]	61%	[137, 141]	35.5%	[142]‡
2008-09	89.0%	[143]	62%	[137, 141]	44.5%	[144]‡
2009-10	93.0%		39%	[137]	55.0%	
2010-11	97.0%	[145]	61%	[146] [147]	64.1%	[148]
2011-12	98.0%	[149]	85%	[147]	75.4%	[148]
2012-13	98.0%		88%	[147]	75.9%	
2013-14	98.0%		91%	[147]	76.3%	[147]
2014-15	98.0%		95%	[147]	91.2%	[147]
2015-16	98.0%		95%	[147]	93.0%	[147]
2016-17	98.0%**		95%†		95.0%†	[150]

* Rates are generally higher than those in the sources cited because it is assumed that some OI patients would be referred for HIV testing at other health services or would independently seek HIV testing even if not referred (for more detail, see Appendix B). ** Rates are assumed to remain constant at 98% after 2016. † Rates are assumed to remain constant at 95% after 2016. ‡ Adjusted to take into account differences in access to ART between provinces.

In calibrating the model to historic data, the average estimate for the base rate, $b(t)$, over the five-year period from mid-2014 to mid-2019, is 0.22. This represents the average annual rate of testing in women aged 25 who are asymptomatic and not pregnant, who have not previously been tested for HIV. This rate is assumed to apply in each future year from mid-2018 onward.

3.3 Adult ART initiation

We model ART initiation as occurring either in the month of HIV diagnosis, or else at longer durations since HIV diagnosis. (In reality relatively few adults start ART within a month of being diagnosed, but we use 'in the same month' as a convenient model approximation to represent individuals who link to care and start ART shortly after HIV diagnosis.) Table 3.3 summarizes the assumed proportions of HIV-positive adults in different categories who are eligible to receive life-long ART, and shows how this has changed over time. 'Eligibility to receive ART' here means only that the relevant guidelines recommended ART initiation in these patients [151-155] – this does not reflect the actual proportion of patients who started ART when they became eligible. In some of the periods the assumed eligible proportion has been set to 50% because the change in guideline occurred midway through the relevant period. For patients with CD4 counts of 200-349 cells/μl, the model allows for non-zero access to ART prior to official guideline changes, as some NGO-supported programmes and private sector programmes applied higher CD4 eligibility thresholds [156-158], and these adjustments are necessary to bring the model estimates in line with reported fractions of ART initiators in the CD4 200-349 category [159, 160].

Table 3.3: Proportions of adult patients assumed to be eligible to receive lifelong ART

	2000-2003*	2003-2009	2009-2010	2010-2011	2011-2012	2012-2014	2014-2015	2015-2016	Post-2016
WHO stage IV or CD4 <200	100%	100%	100%	100%	100%	100%	100%	100%	100%
Pulmonary TB, CD4 200-349	10%	10%	50%	100%	100%	100%	100%	100%	100%
WHO stage III, CD4 350+	0%	0%	0%	0%	0%	100%	100%	100%	100%
Pregnant women, CD4 200-349	10%	10%	50%	100%	100%	100%	100%	100%	100%
Pregnant women, CD4 350+	0%	0%	0%	0%	0%	0%	50%	100%	100%
Other patients, CD4 200-349	10%	10%	10%	20%	80%	100%	100%	100%	100%
Other patients, CD4 350-499	0%	0%	0%	0%	0%	0%	50%	100%	100%
Other patients, CD4 500+	0%	0%	0%	0%	0%	0%	0%	0%	100%†

Calendar periods are defined to run from the middle of the first year quoted to the middle of the second year.
* Applies only to rollout in private sector and NGO-run programmes. † Except in 2016/17 (80%).

The number of adults of sex *g* who initiate ART in the same month as diagnosis, in year *t*, is modelled as

$$S_g^0(t) = \sum_{i=0}^{1}\sum_{s=2}^{5}\sum_{x=15}^{90} N_{g,i,s}(x,t)\left\{b(t)A_g(x,t)r_i(t)l_0(s,t) + \Omega_s d_i(t)l_1(s,t)\right\} + \sum_{i=0}^{2}\sum_{s=2}^{5}\sum_{x=15}^{49} N_{g,i,s}(x,t)F_{g,s}(x,t)v_i(t)l_2(s,t) \quad , \qquad (3.7)$$

where $N_{g,i,s}(x,t)$ is the number of individuals of age *x* and sex *g*, in HIV stage *s* and with HIV testing history *i* at the start of year *t*; $l_0(s,t)$ is the fraction of newly-diagnosed, asymptomatic, non-pregnant individuals in HIV stage *s* who start ART within a month of being diagnosed; and $l_1(s,t)$ and $l_2(s,t)$ are the corresponding fractions of OI patients and pregnant women respectively who start ART immediately after diagnosis (other symbols are defined in the same way as in equation (3.3)). Although the calculation is presented as an annual total for ease of comparison with equation (3.3), the actual model calculations of numbers starting ART are performed at monthly time steps, using monthly rates of HIV testing in place of annual rates of testing. Note that the summation excludes individuals in the acute phase of HIV infection (*s* = 1), since it is assumed that most rapid tests would return negative results during this disease stage. The summation also excludes individuals who were previously diagnosed (*i* = 2), although previously-diagnosed pregnant women are included if they are retested and are ART-eligible.

3.3.1 Linkage to ART after diagnosis during pregnancy

The assumed fractions of ART-eligible pregnant women who start ART during pregnancy are shown in Table 3.2. Assumptions for the early years are based on studies in the Western Cape [138, 142, 144], but are adjusted downward to take into account the lower rate of access to ART in other provinces in the early stages of the ART programme (applying the ratio of the ART initiation rate in women in the Western Cape to that estimated nationally). Assumptions for the more recent years are obtained from national statistics, which showed the proportion increasing to 75.4% in 2011/12 [148]. This proportion increased in subsequent periods, following the introduction of WHO option B at the start of 2013, which eliminated the need for CD4 testing prior to ART initiation and thus simplified the ART initiation process. Based

on data from the DHIS [150], it is assumed that coverage increased to 95% in 2016/17 and remains at this level in subsequent years.

3.3.2 Linkage to ART after HIV diagnosis in OI patients

Few studies have reported on rates of linkage to ART specifically in those patients who are diagnosed in the course of management of an OI. However, relatively high rates of linkage might be expected, given that (a) such patients are sicker and thus likely to be fast-tracked through the patient preparation process, and (b) symptomatic patients are likely to be more motivated to start ART [161-163]. Data from the 2015-16 District Health Barometer indicate that 85% of TB patients with HIV were on ART [147], though this is likely to be an over-estimate of the rate of linkage to ART after TB diagnosis, given that the indicator includes patients who were already on ART prior to developing TB (Katherine Hildebrand, personal communication). In a Cape Town study, the fraction of TB patients diagnosed with HIV who received a CD4 count within 6 months after diagnosis was 16.5% lower than among pregnant women diagnosed with HIV [144]. We have therefore set the assumed value of $l_1(s,t)$ in 2015/16 to 78% (lower than the 93% assumed for pregnant women), and in all other years the value of $l_1(s,t)$ is set to $l_2(s,t) \times 0.78/0.93$, i.e. scaling down the values assumed for pregnant women (Table 3.2).

3.3.3 Linkage to ART after HIV diagnosis in non-pregnant, asymptomatic adults

In a review of sub-Saharan African studies that have examined linkages between HIV diagnostic services and ART services, half of studies included were from South Africa [164]. Restricting attention to those studies conducted in South Africa, the median proportion of patients who received CD4 testing following HIV diagnosis was around 75% and the median proportion of those receiving CD4 testing who collected their test results was around 80%. Of those who were found to be ART-eligible, the average proportion who started ART was around 67%. This suggests that of all individuals who are newly diagnosed and ART-eligible, the proportion who actually start ART within a few months of diagnosis is only about 40% (0.75 × 0.80 × 0.67). We have therefore set the rate of linkage to care in 2012/13 ($l_0(s,2012)$) to 0.4, for all individuals who are ART-eligible. This is roughly half of the rate assumed for pregnant women. We have therefore set the $l_0(s,t)$ parameters for all years prior to 2012 to be half of the corresponding rates assumed for pregnant women (Table 3.2). The rate of linkage of 0.4 estimated in 2012/13 is assumed to also apply in subsequent years. Although it might be expected that a higher rate of linkage would apply after the move to universal ART eligibility in 2016 (since CD4 testing is no longer a requirement prior to ART initiation), the South African experience has suggested that removing the delay due to CD4 testing makes little difference to the proportion of patients who initiate ART soon after diagnosis [165].

3.3.4 ART initiation in previously-diagnosed adults who did not link to ART soon after diagnosis, up to mid-2019

In the period up to mid-2019, the modelled rates of ART initiation in previously-diagnosed adults are calculated from reported numbers of adults starting ART in each period. Suppose that in the period up to mid-2019, $S_g(t)$ is the estimated number of adults of sex g starting ART in month t. Further suppose that $S_g^0(t)$ is the number who started ART immediately after HIV diagnosis in month t, calculated as shown in equation (3.7) but converting the annual total into a monthly number. Let $N_{g,s}(x,t)$ be the number of HIV-diagnosed individuals in CD4 category s, who are ART-naive at time t, of age x and sex g. Let $\mu_{g,s}(x,t)$ be the monthly HIV mortality rate that applies in these individuals, and let $J_s(t)$ be the relative rate of ART initiation in stage s relative to that in the CD4 <200/μl category (s = 5). In most periods $J_s(t)$ will be zero for s < 5, since South African ART guidelines have only recently changed to allow for ART initiation at higher CD4 counts. When all individuals are eligible for ART, we set $J_s(t)$ to 0.40 for CD4 of 500 or higher, 0.50 for CD4 of 350-499, 0.70 for CD4 of 200-349 and 1 for CD4 <200. (These assumptions are based primarily on the observed relative rates of ART initiation in ART-eligible individuals in different CD4 categories [166, 167], and are consistent with the relative rates at which individuals enrolled in pre-ART care return for regular CD4 testing [90, 156].) We wish to estimate the monthly rate at which previously-diagnosed individuals in the CD4 <200/μl category initiate ART, $\rho_g(t)$. We estimate this by noting that

$$\begin{aligned} S_g(t) - S_g^0(t) &= \sum_{x=15}^{90}\sum_{s=1}^{5} N_{g,s}(x,t)\int_0^1 \rho_g(t) J_s(t) \exp\left(-\left(\mu_{g,s}(x,t) + \rho_g(t) J_s(t)\right)u\right) du \\ &\approx \sum_{x=15}^{90}\sum_{s=1}^{5} N_{g,s}(x,t)\rho_g(t) J_s(t)\left(1 - 0.5\left(\mu_{g,s}(x,t) + \rho_g(t) J_s(t)\right)\right) \end{aligned} \qquad (3.8)$$

This is a quadratic in $\rho_g(t)$, and the smaller of the two roots is the rate of ART initiation that we wish to estimate.

The assumed values of $S_g(t)$, expressed as annual totals, are summarized in Table 3.4. These are estimated by combining data from the public sector, private sector and NGO programmes. Surveys of private sector and NGO programmes were conducted every two years up to 2014, to determine total numbers of patients currently receiving ART [168]. In more recent years private sector data have been collected by the Council for Medical Schemes through the South African National AIDS Council (Billia Luwaca, personal communication), and these data have been validated using private sector drug sales [169]. Reporting of patient totals in the public sector has changed over time; early reporting systems provided information only on numbers of patients cumulatively enrolled into ART programmes, but since late 2009 most provinces have switched to reporting numbers of patients *currently* receiving ART [148, 170]. To estimate the number of new initiates in each period from the reported numbers of current patients, we have modelled the change over time in the rate of ART initiation in HIV-diagnosed individuals using piecewise-linear functions over 5-year intervals, with the model being fitted to the reported totals; a more detailed description is provided elsewhere [1]. The model has been fitted separately for each province, and the results presented in Table 3.4 are the aggregated totals for the whole country.

Table 3.4: Assumed annual numbers of patients starting ART in South Africa, and implied average treatment delays in previously-diagnosed patients with CD4 counts <200 cells/μl

	Men (15+)	Women (15+)	Children (<15)	Implied ART delay ($1/\rho(t)$) Men	Women	Children
Pre-2000	0	0	0	-	-	-
2000-01	3071	6526	931	244.0	156.6	472.2
2001-02	4088	8687	1217	234.4	156.4	471.1
2002-03	5270	11198	1459	223.3	157.2	453.9
2003-04	6849	14554	1788	211.6	158.4	463.7
2004-05	11303	24020	3028	201.2	151.5	675.4
2005-06	22296	47378	11205	208.9	132.5	98.7
2006-07	56870	120849	16380	65.9	47.2	67.9
2007-08	89393	189960	23201	41.2	29.2	46.9
2008-09	113942	242127	32256	31.5	21.2	35.0
2009-10	126049	267854	49452	24.9	17.4	24.2
2010-11	164304	349146	43256	21.4	14.4	20.8
2011-12	223176	474248	31531	21.2	13.6	26.9
2012-13	204201	433928	33558	29.0	14.3	19.7
2013-14	159441	338811	25773	37.0	14.5	26.3
2014-15	188477	400513	24427	33.0	16.9	29.6
2015-16	177706	377625	19898	36.0	23.0	44.4
2016-17	193401	410977	20270	34.7	22.3	86.9
2017-18	164248	349026	18553	38.2	21.4	94.5
2018-19	140541	298651	17224	40.6	18.8	83.3

3.3.5 ART initiation in previously-diagnosed adults who did not link to ART soon after diagnosis, after mid-2019

Because we do not yet have data on the absolute numbers starting ART after mid-2019, we specify the $\rho_g(t)$ parameters directly for this period. These parameters can also be expressed in terms of average delays (in months) between diagnosis and ART initiation, if the individual does not link to ART soon after ART initiation ($1/\rho_g(t)$). The average delays implied by our assumed absolute numbers are shown in the last three columns of Table 3.4. Our estimates suggest that in both men and women with CD4 counts of <200 cells/μl, this average delay has increased since 2011/12, possibly as a result of 'crowding out' of sicker patients as ART eligibility criteria have expanded to include healthier patients. However, these results should be interpreted with caution, as the estimates are sensitive to assumptions about linkage to care after diagnosis (sections 3.3.1-3.3.3), which are difficult to determine precisely. For the period after 2023, we assume an average treatment delay of 360 months and 20 months in men and women respectively who have CD4 counts <200 cells/μl, roughly consistent with the average delay over the period from mid-2014 to mid-2019. In the period up to 2022, the delays are interpolated between the rates shown in Table 3.4 and these ultimate rates.

3.4 Mortality after ART initiation in adults

HIV-related mortality after ART initiation is assumed to depend on age, sex, baseline CD4 category and time since ART initiation. The mortality rates specified in Table 3.1 relate to individuals who are aged 35, and these mortality rates are assumed to increase by factors of 1.12 and 1.09 per 10-year increase in age, in men and women respectively. For the most part these parameters have been determined from a model fitted to data from the IeDEA Southern Africa collaboration [121]. However, the IeDEA-SA data relate mainly to individuals who start ART with CD4 counts below 350 cells/μl, and the few patients starting ART at higher CD4 counts are mostly patients who started ART because they qualified on the basis of HIV-related symptoms. Although we lack South African data on mortality in asymptomatic patients starting ART at higher CD4 counts, observational data from high income countries suggest that untreated patients with CD4 counts above 250 cells/μl have similar long-term mortality rates, as long as they start ART before their CD4 count declines below 250 cells/μl [171]. We have therefore set the mortality rates of patients starting ART at higher CD4 counts in such a way that the predicted long-term mortality rate in untreated patients with CD4 counts above 500 cells/μl is roughly the same regardless of whether they start ART immediately, defer ART to when their CD4 count drops below 500, or defer ART to when their CD4 count drops below 350.

Within the group of patients starting ART at CD4 counts <200 cells/μl there is substantial heterogeneity in mortality depending on the exact baseline CD4 value. Although the model does not explicitly model variation in mortality rates by CD4 count below the 200 cells/μl cut-off, mortality rates are adjusted to take into account the rate of ART initiation, since high rates of ART initiation would imply that (a) most individuals starting ART at CD4 <200 cells/μl do so soon after their CD4 count falls below 200, and (b) most untreated individuals with CD4 <200 cells/μl have CD4 counts close to 200. We therefore calculate the theoretical minimum mortality rates that would be expected (both in untreated individuals with CD4 <200 and in treated individuals starting ART with CD4 <200) if ART was started soon after the CD4 count dropped below the 200 threshold. The difference between the mortality rate in Table 3.1 and the theoretical minimum is reduced by a factor of $\exp\left(-m\rho_g(t^-)\right)$ in year *t*, where $\rho_g(t^-)$ is the average rate of ART initiation in the 3 years prior to year *t*, in adults of sex *g* with CD4 <200 cells/μl, and *m* is a scaling factor. This scaled difference is added to the minimum mortality rate to determine the modelled mortality rate in year *t*. To represent the uncertainty regarding the *m* scaling parameter, a gamma prior has been assigned, with a mean of 7.5 and standard deviation of 3.5 [94]. The adjustments are made only to those ART-naïve adults with CD4 counts <200 cells/μl and those treated adults with baseline CD4 counts <200 cells/μl.

In addition, the mortality assumptions have previously been adjusted to take into account potential bias in the IeDEA-SA data. This bias arises mainly because the IeDEA-SA cohorts do not constitute a representative sample of all ART services in South Africa; IeDEA-SA cohorts are almost all located in urban areas, and most have support from academic partners and NGOs. This means that the mortality rates in those cohorts may be lower than the national average. However, the bias may become less substantial at longer treatment durations, as patients frequently discontinue ART or move to different services, though their vital status can still be tracked through the South African vital registration system. In

addition, the IeDEA-SA data may *over-state* the true mortality rate in the longest duration category, i.e. durations >42 months (Table 3.1). This is because the average follow-up duration in the IeDEA-SA cohorts is short, which means that follow-up times in the >42 month category are likely to be biased towards those individuals with relatively short follow-up, who are likely to have higher mortality. We therefore specify parameter I_d to represent the ratio of the true mortality rate to the IeDEA-SA mortality estimate at duration *d* after ART initiation. Prior distributions were previously assigned to represent the uncertainty around I_0 and the ratio I_4/I_0. The I_d values at other durations are calculated by interpolating between the I_0 and I_4 values. Although our original analysis suggested that the I_0 parameter was significantly greater than one [94], a more recent analysis found that after modelling ART interruptions more realistically, the I_0 parameter was not significantly different from one [115]. We have therefore fixed $I_0 = 1$, and have fixed $I_4/I_0 = 0.741$, the same value as estimated in this more recent analysis.

3.5 ART interruptions and switches to second-line ART

The model assumes that ART patients interrupt ART and resume ART at a constant rate, which is independent of age, sex and time since ART initiation. The annual rates of ART interruption and resumption after an interruption are set to 0.25 and 0.90 respectively (for a more detailed explanation of the methods used to estimate these parameters, see Appendix G). The assumptions about ART interruption and resumption are used for the purpose of calculating the numbers of people currently on ART, by time since first ART initiation, but do not influence the calculation of mortality rates after ART initiation.

The model also assumes that ART patients switch from first-line ART to second-line ART at a fixed rate that depends on their HIV disease stage at the time they first initiate ART. No switching is assumed to occur during the first 6 months of ART, as switching can only occur after the first viral load test (which is scheduled at 6 months after ART start). The annual switch rates have been set to 0.0230 in adults who start ART at CD4 <200 cells/µl, 0.0092 in adults with a baseline CD4 count of 200-349 cells/µl, 0.0040 in adults with a baseline CD4 count of 350-499 cells/µl, and 0.0015 in adults starting ART at CD4 counts of 500 cells/µl or higher. A more detailed explanation of these switching assumptions, and the data sources on which they are based, is included in Appendix H. The switching assumptions are used only for the purpose of calculating the modelled numbers of ART patients on first- and second-line ART, and these assumptions have no effect on the model mortality and viral suppression estimates.

4. Model of sexual transmission of HIV

HIV transmission probabilities per act of sex are difficult to determine with a high degree of precision. We therefore specify prior distributions to represent the uncertainty regarding average transmission probabilities from untreated adults to their HIV-susceptible partners, then specify various adjustments to represent the effects of known cofactors on HIV transmission.

4.1 The effect of sex and relationship type

The symbol $\beta_{g,l}$ represents the average HIV transmission probability, in a single act of unprotected sex, from an untreated HIV-positive individual of sex *g*, to an HIV-negative partner in relationship type *l*. Table 4.1 summarizes the assumed prior distributions for these parameter values. Although empirical estimates suggest high female-to-male transmission probabilities per act of unprotected sex in unmarried men [172, 173], these are likely to be over-estimates, as they may be inflated by male acquisition of HIV infection through sex worker contact, which is often substantially under-reported [174]. The prior distribution for the $\beta_{1,0}$ parameter has therefore been set in such a way that the mean (0.008) is below the empirical estimates (0.016 and 0.0128) but the 97.5 percentile of the distribution (0.015) is close to the empirical estimates. Beta distributions are used for all of the specified priors. A justification for the prior distributions on the male-to-male and client-to-sex worker transmission probabilities is provided in Appendix C.

Table 4.1: Assumed probabilities of HIV transmission per act of sex

Relationship type	Symbol	Susceptible female			Susceptible male[c]		
		Mean	Std dev.	Ref.	Mean	Std dev.	Ref.
CSW-client relationships	$\beta_{g,2}$	0.001	0.0005	[175, 176]	0.008[b]	-	-
Short-term relationships	$\beta_{g,0}$	0.012	0.005	[177, 178]	0.008	0.003	[172, 173]
Long-term relationships	$\beta_{g,1}$	0.0043[a]	-	[88, 179, 180]	0.0010[a]	-	[88, 179, 180]
MSM relationships	$\beta_{1,4}$	-	-		0.020	0.005	[181-183]

CSW = commercial sex worker.
[a] Fixed parameter, not included in Bayesian analysis, based on previous model calibrations. [b] Parameter value is assumed to be the same as in short-term relationships. [c] For a male partner who is uncircumcised.

4.2 The effect of risk group

Sexually transmitted infections (STIs) have been shown to have a significant effect on HIV transmission probabilities, both when present in the HIV-susceptible partner [184, 185] and when present in the HIV-infected partner [186]. Although Thembisa does not model other STIs explicitly, we would expect the prevalence of other STIs to be higher in the high-risk group than in the low-risk group, and for this reason, some adjustment to the previously-stated HIV transmission probabilities is appropriate, depending on the risk groups of the HIV-infected partner and the HIV-susceptible partner. The transmission probabilities

specified in Table 4.1 are assumed to apply to partnerships in which both partners are in the low risk group (except in the case of interactions between sex workers and clients, in which both partners are by definition high risk). The parameter $\Theta_{g,i,l,j}$ is defined to represent the ratio of the transmission probability from an infected individual of sex *g* and risk group *i* to a partner of type *l* in risk group *j*, to the transmission probability that would be expected if both partners were low risk. These parameter values have been estimated from a previously-published model of STI-HIV interactions in South Africa [187], and are shown in Table 4.2.

Table 4.2: Assumed multiples by which HIV transmission probabilities are increased depending on partner risk groups

	Short-term contacts		Marital contacts	
	HIV+ male partner	HIV+ female partner	HIV+ male partner	HIV+ female partner
High risk male, high risk female	1.23	1.20	1.62	1.38
High risk male, low risk female	1.25	1.14	1.57	1.35
Low risk male, high risk female	1.08	1.09	1.33	1.24
Low risk male, low risk female	1.00	1.00	1.00	1.00

4.3 The effect of HIV stage and antiretroviral treatment

Table 3.1 shows how relative levels of HIV infectiousness are assumed to differ by CD4 count in untreated adults. Although we do not express these assumptions in terms of differences in viral load between CD4 stages, we do make assumptions about viral load distributions and HIV infectiousness as a function of viral load for the purpose of calculating average levels of infectiousness after ART initiation. Suppose that random variable $X_{a,s}$ is the difference between the maximum viral load and the actual viral load, on the logarithmic scale, in individuals with ART status *a* (0 = untreated, 1 = treated) and CD4 stage *s* (in untreated individuals, *s* refers to the current CD4 stage, while in treated individuals *s* refers to the CD4 stage at the time of ART initiation). The maximum viral load is set to 6 on the $\log_{10}$ scale (although higher values are possible, these have little effect on the HIV transmission dynamics in which we are interested). Variable $X_{a,s}$ is assumed to be Weibull-distributed, with parameters $\omega_{a,s}$ and ϕ. The probability of viral suppression (a viral load of less than 400 copies/ml) in treated individuals is thus

$$\exp\left(-\omega_{1,s}\left(6-\log 400\right)^{\phi}\right), \tag{4.1}$$

from which it follows that if $V_s(t)$ is the probability of viral suppression in year *t*, at a threshold of <400 copies/ml, then

$$\omega_{1,s} = \frac{-\ln\left(V_s(t)\right)}{\left(6-\log 400\right)^{\phi}}. \tag{4.2}$$

In fitting Weibull distributions to viral load data from both treated [104, 107] and ART-naïve South Africans [188], we have found that a ϕ parameter of 1.5 produces reasonable fits. For a given level of viral suppression, $V_5(t)$, it is then possible to calculate $\omega_{1,5}$. For example, if the rate of viral suppression in patients starting ART with CD4 <200 cells/μl is set to 0.77,

substituting $V_5(t) = 0.77$ into equation (4.2) yields a $\omega_{1,5}$ estimate of 0.042. For ART-naïve patients, a different approach is adopted in estimating $\omega_{a,s}$. Based on fitting the Weibull model to the median and inter-quartile range of viral loads prior to ART initiation in South Africans who almost all had CD4 counts of <200 cells/μl [188], we estimate the $\omega_{0,5}$ parameter to be 0.635.

The $V_5(t)$ parameters have been estimated from reported levels of viral suppression, allowing for a change in the rate of viral suppression over time; a more detailed explanation of the data sources and assumptions is provided in Appendix F. Very briefly, estimates of viral suppression are derived from both IeDEA-SA data [189] and TIER reporting [159], allowing for uncertainty in the representativeness of the IeDEA-SA data and the extent of the bias due to missing viral load data. For the purpose of the calibration of Thembisa (described in section 7), we specify a parameter to represent the ratio of the 'true' (unobserved) odds of viral suppression nationally to the odds of viral suppression measured in IeDEA-SA cohorts (the model input up to 2018). To represent the uncertainty around this odds ratio we assign a gamma prior with a mean of 0.844 and standard deviation 0.092 (the 2.5 and 97.5 percentiles of this distribution are 0.67 and 1.03 respectively, i.e. we allow for substantial uncertainty regarding the bias in the IeDEA-SA data).

Since 2019, dolutegravir has replaced efavirenz as the standard first line drug in South Africa, and most patients who are virally suppressed are being switched to dolutegravir. Dolutegravir is associated with significantly greater viral suppression than efavirenz; in a recent network meta-analysis, the odds of viral suppression on dolutegravir was estimated to be 1.87 times that on efavirenz [190]. Based on this, the odds of viral suppression $V_5(t)$ in 2020 is assumed to be 1.87 times that in 2018, and the odds of viral suppression is assumed to remain constant at this level in future years. (The odds of viral suppression in 2019 is assumed to be half of that in 2018 and 2020, as this was a year of transition from dolutegravir to efavirenz.)

We assume that if x is the difference between the maximum viral load and the actual viral load (on the logarithmic scale), the HIV transmission risk per act of sex is

$$c\exp\left(-\theta x^{\phi}\right), \tag{4.3}$$

where c is the maximum HIV transmission risk (when $x = 0$) and parameter θ determines the extent of the association between viral load and HIV transmission risk. Including $\phi > 1$ in the above equation ensures that the effect of viral load is less substantial at higher viral load levels than at lower viral load levels [191]. For reasons of mathematical convenience, explained below, we use the same value of $\phi = 1.5$ as estimated in the model of viral load distributions. The θ parameter is estimated by noting that if the factor by which infectiousness increases, per unit increase in viral load, is of the order of 2.5 [88, 192, 193], this implies that

$$\frac{-\frac{d}{dx}\left[c\exp\left(-\theta x^{\phi}\right)\right]}{c\exp\left(-\theta x^{\phi}\right)} = \ln(2.5). \tag{4.4}$$

From this it follows that $\theta\phi x^{\phi-1} = \ln(2.5)$. Substituting ϕ = 1.5 and x = 2 [88, 192] yields θ = 0.432. The average HIV transmission probability, for patients with ART status a and CD4 stage s, is then

$$\int_0^\infty \omega_{a,s}\phi x^{\phi-1}\exp\left(-\omega_{a,s}x^{\phi}\right)c\exp\left(-\theta x^{\phi}\right)dx = c\int_0^\infty \omega_{a,s}\phi x^{\phi-1}\exp\left(-\left(\theta+\omega_{a,s}\right)x^{\phi}\right)dx$$
$$= \frac{c\omega_{a,s}}{\omega_{a,s}+\theta}. \qquad (4.5)$$

The advantage of using the same value of ϕ = 1.5 in the modelled relationship between viral load and HIV transmission risk is thus that it ensures a simple mathematical expression for the average probability of HIV transmission. From equation (4.5), the ratio of the infectiousness after ART initiation to that prior to ART initiation is

$$R_s = \frac{\omega_{1,s}}{\omega_{1,s}+\theta}\Big/\frac{\omega_{0,s}}{\omega_{0,s}+\theta}. \qquad (4.6)$$

Substituting the values of $\omega_{1,5}$ = 0.042 and $\omega_{0,5}$ = 0.635 into this equation, for example, yields an R_5 estimate of 0.149. This is somewhat higher than the relative risk estimates of 0.04-0.08 estimated from randomized controlled trials [119, 194], but lower than the relative risk of 0.36 estimated in a meta-analysis of observational studies [195]. It is important to note, however, that the value of R_s changes over time, as the $\omega_{1,s}$ parameter in the numerator changes as the rate of viral suppression changes.

Patients who start ART at higher CD4 counts (>200 cells/μl) have lower rates of virological failure after ART initiation [196]. As explained in Appendix F, we assume that the odds of viral suppression in the CD4 200-349, 350-499 and ≥500 categories are 1.49, 1.73 and 1.92 times those in the <200 category respectively; these assumptions determine the $\omega_{1,s}$ values, for given values of $V_5(t)$. In untreated patients with CD4 >200 cells/μl, we assume average viral load levels decrease by 0.16 for each 100-cell increase in the CD4 cell count [101] (which determines the $\omega_{0,s}$ values). Reductions in infectivity in patients who start ART at CD4 counts >200 cells/μl are then calculated from equation (4.6),

Although most South African data sources report viral suppression at a threshold of <400 copies/ml, global reporting standards recommend a threshold of <1000 copies/ml [197, 198]. For the sake of consistency with these global reporting standards, equation (4.1) is used to estimate viral suppression at this threshold (replacing 400 with 1000), although we have few direct South Africa data to inform the levels of viral suppression at this threshold.

4.4 Condom effectiveness

Condoms are assumed to be 90% effective in preventing HIV transmission. Although this is slightly higher than the rates of around 80% that have been estimated empirically in the context of heterosexual intercourse [88, 199], it is likely that empirical estimates are biased downward due to over-reporting of condom usage [200, 201]. Levels of condom efficacy close to 90% have also been estimated in MSM [182].

4.5 Age and year effects

Young women are at a biologically increased risk of HIV acquisition due to the high prevalence of cervical ectopy in adolescence and young adulthood [202-204], and their relatively low levels of protective lactobacilli [205]. The model makes allowance for this heightened susceptibility by assuming that the HIV transmission risk per act of sex is increased by a factor of $Z_g(x)$, relative to adults aged 25 and older, in individuals of sex g and age x. The function $Z_g(x)$ is defined as

$$Z_g(x) = \begin{cases} (1+Z_g)^{25-x} & \text{for } x < 25 \\ 1 & \text{for } x \geq 25 \end{cases}. \qquad (4.7)$$

The Z_2 parameter (for females) is set equal to 0.15, based on studies that have quantified the effect of age on HIV transmission probabilities in women [206-208]. For males, there does not appear to be strong evidence of age variation in the risk of HIV acquisition per sex act [88, 207], and the Z_1 parameter has therefore been set to zero.

As described in section 3.1, the model allows for changes in HIV virulence over time through the parameter E, which represents the factor by which the rate of CD4 decline changes per year, in untreated adults. These changes in virulence are likely to be associated with changes in set point viral load (SPVL), which in turn are likely to cause changes in HIV transmission probabilities. The model therefore allows for an annual change in the transmission probability, which depends on the annual change in the rate of CD4 decline. We define the transmission probability in year t to be

$$\beta_{g,l}(t) = \beta_{g,l} E^{(t-1999)\times 2.5\alpha}, \qquad (4.8)$$

where 2.5α is the scaling factor for the relationship between HIV virulence and HIV transmissibility. As explained in more detail elsewhere [94], the α parameter can be interpreted as the ratio of the increase in infectivity to the increase in HIV disease progression (on a natural log scale), for a given change in SPVL. However, as noted in section 3.1, the E parameter is currently fixed at 1, and the model therefore assumes HIV transmission probabilities in untreated individuals are constant over time.

4.6 Mathematical model of heterosexual transmission

We define $\Gamma(s)$ to be the frequency of sex in untreated HIV disease stage s, relative to that in uninfected individuals (these parameters are estimated in section 2.9). The previously-defined $\beta_{g,l}$ transmission probabilities are assumed to be weighted averages of the probabilities from all untreated disease stages, where the weights are calculated from the expected numbers of unprotected sex acts in each stage. If we define $\beta^*_{g,l}$ to be the transmission probability from chronically-infected individuals who have CD4 counts ≥500 cells/μl (s = 2), then

$$\beta_{g,l} = \beta^*_{g,l} \frac{\sum_{s=1}^{3} \frac{I_s \Gamma(s)}{\lambda_s} + \frac{I_4 \Gamma(4)}{\lambda_4 + \mu_4} + \frac{\lambda_4}{\lambda_4 + \mu_4} \times \frac{I_5 \Gamma(5)}{\mu_5}}{\sum_{s=1}^{3} \frac{\Gamma(s)}{\lambda_s} + \frac{\Gamma(4)}{\lambda_4 + \mu_4} + \frac{\lambda_4 \Gamma(5)}{(\lambda_4 + \mu_4)\mu_5}}, \quad (4.9)$$

where the I_s factors are the relative levels of infectiousness (Table 3.1), and the CD4 decline parameters (λ_s) and mortality parameters (μ_s) are those specified in section 3.1. We define I^*_s to be the ratio of infectiousness in stage *s* to average infectiousness, from which it follows that $I^*_2 = \beta^*_{g,l} / \beta_{g,l}$, and hence

$$I^*_2 = \frac{\sum_{s=1}^{3} \frac{\Gamma(s)}{\lambda_s} + \frac{\Gamma(4)}{\lambda_4 + \mu_4} + \frac{\lambda_4 \Gamma(5)}{(\lambda_4 + \mu_4)\mu_5}}{\sum_{s=1}^{3} \frac{I_s \Gamma(s)}{\lambda_s} + \frac{I_4 \Gamma(4)}{\lambda_4 + \mu_4} + \frac{\lambda_4}{\lambda_4 + \mu_4} \times \frac{I_5 \Gamma(5)}{\mu_5}}. \quad (4.10)$$

For other values of *s*, $I^*_s = I^*_2 \times I_s$. Lastly, we define $I^*_s(a)$ to be the relative infectiousness for individuals with ART status *a* (0 implying ART-naïve and 1 implying ever treated), where *s* is either the current HIV stage (for *a* = 0) or the HIV stage at the time ART was initiated (for *a* = 1). For ART-naïve individuals $I^*_s(0) = I^*_s$. For ART-experienced individuals who started ART in HIV disease stage *s*, the relative infectiousness is $I^*_s(1) = I^*_s\left(\iota_d + (1-\iota_d)R_s\right)$, where ι_d is the proportion of ART-experienced adults surviving to duration *d* after ART initiation, who are interrupting ART, and R_s is the relative infectivity after ART initiation (as defined in equation (4.6)). The ι_d parameters have been set to 0.053 for the first 6 months after ART initiation, 0.142 for months 7-18, 0.186 for months 19-30, 0.199 for months 31-42 and 0.204 for longer ART durations, based on a model of ART interruptions in South Africa described in Appendix G.

We define $G(v,a)$ to be the ratio of the proportion of sex acts that are unprotected in individuals with testing history *v* and ART status *a*, to that in individuals who are HIV-negative. As in section 3.2, the HIV testing history *v* is coded as 0 if the individual has never been tested, 1 if the individual has been tested but not diagnosed positive, and 2 if the individual has been diagnosed positive. For all values of *v* < 2, we set $G(v,a)$ = 1, while for *v* = 2 we set

$$G(v,a) = (1-\delta(t))(1-h)^a, \quad (4.11)$$

where the $\delta(t)$ and *h* parameters represent the reductions in unprotected sex due to HIV diagnosis and ART initiation respectively (see sections 2.10 and 2.11).

We define $Y(a,s,d)$ to be the ratio of the frequency of sex in individuals with ART status *a* and CD4 stage *s*, with duration *d* since first ART initiation, to the frequency of sex in HIV-

negative individuals. In the case of ART-naïve individuals (a = 0 and d = 0), $Y(0,s,0) = \Gamma(s)$. In the case of ART-experienced individuals, we define

$$Y(1,s,d) = \iota_d \Gamma(s) + (1-\iota_d)\sum_{s'=2}^{5} \psi_d(s' \mid s)\Gamma(s'), \qquad (4.12)$$

where $\psi_d(s' \mid s)$ is the proportion of surviving ART patients with current CD4 count in category s', in the cohort of patients who started ART with a CD4 count of s and who are in ART duration category d. Individuals who interrupt ART are assumed to experience a return to baseline CD4 levels [87], and the frequency of sex is thus assumed to be a function only of the *current* CD4 count. The $\psi_d(s' \mid s)$ values are estimated from studies of CD4 distributions after ART initiation [209-211], and the assumed values are shown in Table 4.3.

Table 4.3: Proportions of treated patients in different CD4 categories

	Time since ART initiation (months)				
	0-6	7-18	19-30	31-42	43+
Patients starting ART with CD4 <200					
Proportion with current CD4 500+	0.00	0.08	0.22	0.30	0.41
Proportion with current CD4 350-499	0.01	0.22	0.29	0.30	0.28
Proportion with current CD4 200-349	0.20	0.44	0.35	0.30	0.24
Proportion with current CD4 <200	0.79	0.26	0.13	0.10	0.06
Patients starting ART with CD4 200-349					
Proportion with current CD4 500+	0.00	0.28	0.57	0.66	0.81
Proportion with current CD4 350-499	0.26	0.58	0.35	0.25	0.16
Proportion with current CD4 200-349	0.73	0.14	0.08	0.08	0.04
Proportion with current CD4 <200	0.01	0.00	0.00	0.00	0.00
Patients starting ART with CD4 350-499					
Proportion with current CD4 500+	0.28	0.69	0.81	0.84	0.93
Proportion with current CD4 350-499	0.65	0.29	0.17	0.13	0.06
Proportion with current CD4 200-349	0.06	0.02	0.02	0.03	0.01
Proportion with current CD4 <200	0.00	0.00	0.00	0.00	0.00
Patients starting ART with CD4 500+					
Proportion with current CD4 500+	0.91	0.93	0.94	0.94	0.98
Proportion with current CD4 350-499	0.09	0.07	0.06	0.05	0.02
Proportion with current CD4 200-349	0.00	0.00	0.00	0.01	0.00
Proportion with current CD4 <200	0.00	0.00	0.00	0.00	0.00

For the purpose of calculating average transmission probabilities, we define $N^{r}_{g,i,l,j}(x)$ to be the total number of individuals aged x and of sex g, who are in risk group i, in relationship state l (0 for unmarried heterosexual, 1 for married/cohabiting, 2 for sex workers and 3 for MSM) with a partner in risk group j (the j subscript is omitted in the case of unmarried individuals, i.e. for l = 0, 2 or 3) and circumcision status r (1 for circumcised males, 0 otherwise). Within this group we define $X^{r}_{g,i,l,j}(x,a,s,v,d)$ to be the proportion who are in HIV stage s, with ART status a, HIV testing history v and ART duration d. In total there are 35 possible HIV-positive states, summarized in Table 4.4.

Table 4.4: Definitions of HIV-positive states

ART status (a)	HIV stage (s)	Testing history (v)	ART duration (d)	Description
0	1	0	0	Acutely infected, never tested
0	2	0	0	CD4 ≥500, never tested
0	3	0	0	CD4 350-499, never tested
0	4	0	0	CD4 200-349, never tested
0	5	0	0	CD4 <200, never tested
0	1	1	0	Acutely infected, previously tested but undiagnosed
0	2	1	0	CD4 ≥500, previously tested but undiagnosed
0	3	1	0	CD4 350-499, previously tested but undiagnosed
0	4	1	0	CD4 200-349, previously tested but undiagnosed
0	5	1	0	CD4 <200, previously tested but undiagnosed
0	1	2	0	Acutely infected, diagnosed but not yet treated*
0	2	2	0	CD4 ≥500, diagnosed but not yet treated
0	3	2	0	CD4 350-499, diagnosed but not yet treated
0	4	2	0	CD4 200-349, diagnosed but not yet treated
0	5	2	0	CD4 <200, diagnosed but not yet treated
1	2	2	0	Started ART with CD4 ≥500 in current year
1	2	2	1	Started ART with CD4 ≥500 in previous year
1	2	2	2	Started ART with CD4 ≥500 2 years previously
1	2	2	3	Started ART with CD4 ≥500 3 years previously
1	2	2	4	Started ART with CD4 ≥500 4 years previously or earlier
1	3	2	0	Started ART with CD4 350-499 in current year
1	3	2	1	Started ART with CD4 350-499 in previous year
1	3	2	2	Started ART with CD4 350-499 2 years previously
1	3	2	3	Started ART with CD4 350-499 3 years previously
1	3	2	4	Started ART with CD4 350-499 4 years previously or earlier
1	4	2	0	Started ART with CD4 200-349 in current year
1	4	2	1	Started ART with CD4 200-349 in previous year
1	4	2	2	Started ART with CD4 200-349 2 years previously
1	4	2	3	Started ART with CD4 200-349 3 years previously
1	4	2	4	Started ART with CD4 200-349 4 years previously or earlier
1	5	2	0	Started ART with CD4 <200 in current year
1	5	2	1	Started ART with CD4 <200 in previous year
1	5	2	2	Started ART with CD4 <200 2 years previously
1	5	2	3	Started ART with CD4 <200 3 years previously
1	5	2	4	Started ART with CD4 <200 4 years previously or earlier

* Only relevant in the case of individuals who seroconvert while receiving PrEP – all other infections are assumed to be diagnosed following acute infection.

As in recent modelling of herpes transmission [212], the model is parameterized in terms of a force (or 'hazard') of transmission per sex act, which is then converted into a cumulative hazard, given the expected number of sex acts in the relationship. For an HIV-positive individual with state covariates (a, s, v, d), the cumulative hazard for HIV transmission, per short-term partnership with a partner in risk group j, is

$$\begin{aligned} &n_{g,0}(x)Y(a,s,d)\beta_{g,0}(t)I_s^*(a)\Theta_{g,i,0,j}\left(1-\left[1-\left(1-\gamma_{g,0}(x,t)\right)G(v,a)\right]E\right) \\ &= n_{g,0}(x)Y(a,s,d)\beta_{g,0}(t)I_s^*(a)\Theta_{g,i,0,j}\left\{\left(1-\gamma_{g,0}(x,t)\right)G(v,a)E+\left(1-E\right)\right\} \end{aligned} \qquad (4.13)$$

where $n_{g,0}(x)$ is the average number of sex acts per short-term relationship, $\gamma_{g,0}(x,t)$ is the probability of condom use by HIV-negative individuals (as defined in section 2.8), and E is

the condom efficacy parameter. From this we can calculate the probability of HIV transmission per short-term partnership:

$$1-\exp\left(-n_{g,0}(x)Y(a,s,d)\beta_{g,0}(t)I_s^*(a)\Theta_{g,i,0,j}\left\{\left(1-\gamma_{g,0}(x,t)\right)G(v,a)E+\left(1-E\right)\right\}\right) . \quad (4.14)$$

The rate at which individuals transmit HIV, per short-term partnership with a partner in risk group *j*, averaged across the HIV disease stages defined in Table 4.4, is defined as

$$T_{g,i,l,k}^{0,r}(j,x)=\sum_{a,s,v,d}X_{g,i,l,k}^{r}(x,a,s,v,d)n_{g,0}(x)Y(a,s,d)\beta_{g,0}(t)I_s^*(a)\Theta_{g,i,0,j} \times\left\{\left(1-\gamma_{g,0}(x,t)\right)G(v,a)E+\left(1-E\right)\right\} . \quad (4.15)$$

For the sake of simplicity, we consider here only the case where the susceptible partner is uncircumcised and is not receiving PrEP or microbicides, but allowing for these factors involves only a multiplicative adjustment to the $T_{g,i,l,k}^{0,r}(j,x)$ variable. It is also worth noting here that although we have expressed these equations in terms of rates of transmission per short-term partnership, the approach is the same for long-term partnerships (replacing 0 with 1 in the above equations), except that $n_{g,1}(x)$ is defined as the number of sex acts *per month*, and hence $T_{g,i,l,k}^{1,r}(j,x)$ represents the average transmission rate per month rather than per partnership. The same approach is also followed in interactions between sex workers and their clients (replacing 0 with 2 in the above equations), except that these interactions are assumed to comprise a single act, meaning that the $n_{g,l}(x)$ factor is 1 and $T_{g,i,l,k}^{2,r}(j,x)$ represents the average transmission probability per sex act. The same approach is again followed in MSM relationships (replacing 0 with 3 in the above equations). Finally, it should be noted that the relationship type in the superscript is not necessarily the same as the marital status indicator (*l*) in the subscript, as some married individuals may engage in extramarital or commercial sex activity. Similarly, the risk group of the long-term partner (*k*) is not necessarily the same as the risk group of the partner under consideration (*j*).

The average probability of transmission per short-term relationship is calculated as

$$1-\exp\left(-T_{g,i,l,k}^{0,r}(j,x)\right), \quad (4.16)$$

and the average probability that an individual aged *x*, of sex *g* and risk group *i*, transmits HIV to a short-term partner in risk group *j* is

$$U_{g,i}^{0}(j,x)=1-\frac{\sum_{r,l,k}N_{g,i,l,k}^{r}(x)c_{g,i,l}(x)\exp\left(-T_{g,i,l,k}^{0,r}(j,x)\right)}{\sum_{r,l,k}N_{g,i,l,k}^{r}(x)c_{g,i,l}(x)}, \quad (4.17)$$

where $c_{g,i,l}(x)$ is the annual rate at which new non-spousal relationships are formed (as defined in section 2.2). Although MSM are included in the above equation (l = 3), their rate of partner acquisition is scaled down in proportion to the fraction of their partners who are female (see Appendix A).

Following the same approach, the average monthly probability that an individual aged x, of sex g and risk group i, transmits HIV to a long-term partner in risk group j is

$$U^1_{g,i}(j,x) = 1 - \frac{\sum_r N^r_{g,i,1,j}(x)\exp\left(-T^{1,r}_{g,i,1,j}(j,x)\right)}{\sum_r N^r_{g,i,1,j}(x)}, \tag{4.18}$$

and the average probability that a client transmits HIV to a sex worker is

$$U^2_{1,1} = \frac{\sum_{r,l,k,x} N^r_{1,1,l,k}(x) w_l(x) T^{2,r}_{1,1,l,k}(1)}{\sum_{r,l,k,x} N^r_{1,1,l,k}(x) w_l(x)}, \tag{4.19}$$

where $w_l(x)$ is the annual rate at which high risk men visit sex workers if they are aged x and of marital status l. Note that MSM are excluded from this equation, i.e. MSM are assumed not to have sex with female sex workers.

Now consider a sexually experienced HIV-*negative* individual of sex g in risk group i, aged x and with marital status l. The probability that this individual acquires HIV from a short-term heterosexual partner in the next month is

$$P^0_{g,i,l}(x) = 1 - \exp\left(-\frac{c_{g,i,l}(x)}{12} Z_g(x) \sum_{y=10}^{90} f_{g,0}(y \mid x) \times \left[\rho_{g,i,0}(1,t) U^0_{3-g,1}(i,y) + \rho_{g,i,0}(2,t) U^0_{3-g,2}(i,y)\right]\right) \tag{4.20}$$

where $f_{g,0}(y \mid x)$ is the proportion of short-term partners who are aged y, $\rho_{g,i,0}(j,t)$ is the proportion of partners who are in risk group j, and $(3 - g)$ is the sex opposite to g.

For a man who has sex with other men, a similar approach is adopted in calculating their probability of HIV acquisition through a same-sex relationship, with the inclusion of a factor $\Omega(x)$ to represent the fraction of partners who are of the same sex:

$$P^0_{1,i,3}(x) = 1 - \exp\left(-\frac{c_{1,i,3}(x)\Omega(x)}{12} Z_1(x) \sum_{y=10}^{90} f_{1,3}(y \mid x) \times \left[\rho_{1,i,3}(1,t) U^0_{1,1}(i,y) + \rho_{1,i,3}(2,t) U^0_{1,2}(i,y)\right]\right) \tag{4.21}$$

If the individual is married to an individual in risk group j, the probability that they acquire HIV from their marital partner in the next month is

$$P^1_{g,i,1,j}(x) = 1 - \exp\left(- Z_g(x) \sum_{y=10}^{90} f_{g,1}(y \mid x) U^1_{3-g,j}(i,y) \right). \quad (4.22)$$

If the individual is a high-risk man who has sex only with women, then the probability that they acquire HIV from a sex worker in the next month is

$$P^2_{1,1,l}(x) = 1 - \exp\left(- \frac{w_l(x)}{12} Z_1(x) \sum_{y=10}^{90} N^0_{2,1,2}(y) T^{2,0}_{2,1,2}(1,y) \Big/ \sum_{y=10}^{90} N^0_{2,1,2}(y) \right), \quad (4.23)$$

and if the individual is a female sex worker her probability of HIV acquisition in the next month is

$$P^2_{2,1,2}(x) = 1 - \exp\left(- \frac{C}{12} Z_2(x) U^2_{1,1} \right), \quad (4.24)$$

where C is the average annual number of clients a sex worker has.

4.7 Extensions to represent effect of male circumcision

Men who are circumcised are assumed to have a 60% lower probability of HIV acquisition than uncircumcised men, per act of sex with an HIV-positive partner (the transmission probabilities in Table 4.1 relate to uncircumcised men) [213-216]. Male circumcision is assumed to have no effect on male-to-female rates of HIV transmission [217] or male-to-male transmission [218, 219].

4.7.1 Male circumcision prior to the promotion of MMC as an HIV prevention strategy

The rate at which men get circumcised is assumed to be composed of two parts: the 'background' rate of male circumcision that would be expected in the absence of any efforts to promote male circumcision as an HIV prevention strategy, and the rate of male circumcision due to medical male circumcision (MMC) campaigns. In modelling the former, a cumulative Weibull distribution is used to represent the age-related changes in the prevalence of male circumcision prior to 2008. It is assumed that the prevalence of male circumcision at age x is determined by the function

$$p(x) = a + (b - a)\left(1 - 0.5^{(x/m_1)^{\phi}} \right), \quad (4.25)$$

where a is the proportion of males who are circumcised soon after birth, b is the maximum cumulative uptake of male circumcision in the absence of MMC promotion, m_1 is the median age at circumcision in men who get circumcised after birth, and ϕ is the shape parameter that determines the concentration of the distribution of circumcision ages (post-birth) around the

median. Since surveys usually report the median age at circumcision for all men (including those who are circumcised at the time of birth), it is useful to parameterize the model in terms of this overall median circumcision age, m_2, noting that

$$m_1 = m_2\left(\frac{\ln\left(b/\left(2(b-a)\right)\right)}{\ln(0.5)}\right)^{-1/\phi} \qquad \text{for } \frac{b}{2} > a\,. \tag{4.26}$$

Parameters a and b are set at 0.105 and 0.42 respectively. The shape parameter ϕ is set at 4.5, and the median age at circumcision m_2 is set at 18, the median age at circumcision reported by Africans in the 2002 HSRC survey [220]. Most of these parameters have been set so that the model is consistent with reported rates of male circumcision by age in national surveys [220-222], after correcting the self-reported data to take into account known biases in the reporting of male circumcision [223-229]. These corrections also take into account that many men who report being circumcised are only partially circumcised (i.e. we treat partially circumcised men as if they are uncircumcised and assume that partial circumcision provides no protection against HIV). The two national surveys used in the parameterization were conducted in 2002 and 2003, and thus represent the situation prior to the promotion of male circumcision as an HIV prevention strategy. Figure 4.1 shows the model calibration.

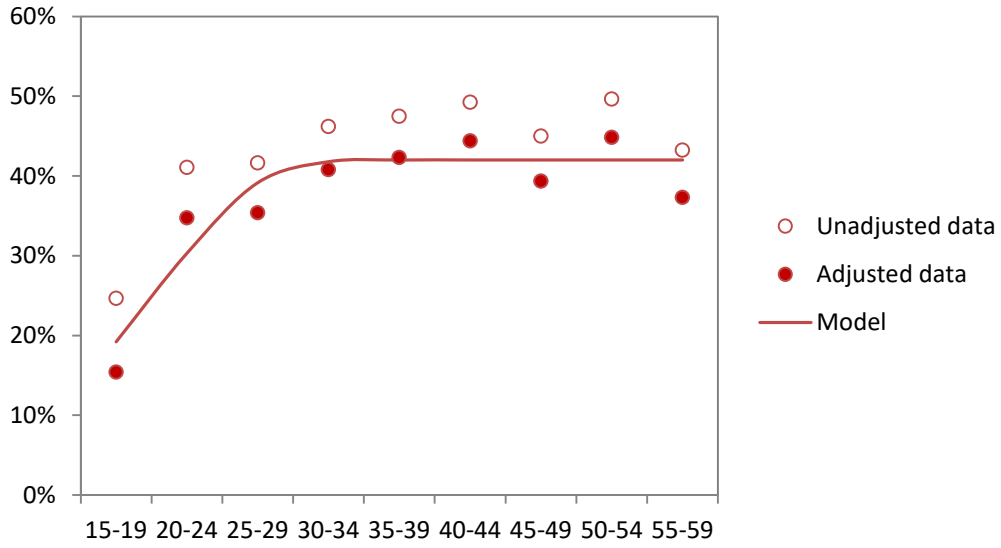


Figure 4.1: Fraction of men who are circumcised, by age, prior to MMC campaigns
Unadjusted data represent the average of the results from national surveys in 2002 and 2003 [220-222]. Adjusted estimates are calculated on the assumption that the sensitivity and specificity of self-reported male circumcision status (relative to true status) are 96.4% and 88.4% respectively [223-228].

The annual probability that uncircumcised men aged x would get circumcised in the absence of MMC campaigns is calculated from the $p(x)$ values defined previously using the equation

$$\psi(x) = 1 - \frac{1-p(x+1)}{1-p(x)}. \tag{4.27}$$

4.7.2 Male circumcision in the MMC campaign era

Extending the model to include MMC in response to MMC promotion campaigns requires that we define the symbol $p^*(x, t)$ as the proportion of men aged x, at time t, who are circumcised. Of those men who are uncircumcised at age x in year t, the proportion who intend to get traditionally circumcised (i.e. they would want to get circumcised even in the absence of MMC promotion campaigns) is calculated as

$$\frac{b - p(x)}{1 - p^*(x,t)}, \tag{4.28}$$

and the proportion who do not intend to get traditionally circumcised is

$$\frac{1 - p^*(x,t) - \left(b - p(x)\right)}{1 - p^*(x,t)}. \tag{4.29}$$

The implicit assumption is that the men who intend to get traditionally circumcised would not accept MMC, i.e. the demand for traditional MC and the demand for MMC are mutually exclusive. This is different from the assumption made in Thembisa version 4.1 [115], in which it was assumed that the demand for MMC was independent of the individual's desire for traditional male circumcision. This change was made because the previous model produced estimates of circumcision coverage that appeared implausibly low relative to the levels reported in recent surveys (even after correcting for misreporting), and because data from the 2017 HSRC household survey suggest there has been no reduction in the prevalence of traditional male circumcision since the start of the MMC rollout (in contrast to what would be expected if some of the men who would previously have been traditionally circumcised instead chose MMC) [5]. With the revised assumption, the model yields a slightly higher estimate of male circumcision coverage, more consistent with recent survey data.

Men are assumed to undergo MMC only if they are HIV-negative, as HIV testing is conducted prior to most MMC operations [230, 231], and although men who are HIV-positive are not excluded from getting circumcised, there would be little incentive to undergo the procedure if they were already HIV-positive. The symbol $\eta(x, t)$ is defined as the probability that HIV-negative men who are aged x, uncircumcised at the start of year t, and not intending to get traditionally circumcised, get medically circumcised through MMC campaigns. This is calculated as $\eta(x, t) = \theta(t) \times R(x)$, where $\theta(t)$ is the maximum probability in year t and $R(x)$ is the relative rate of MMC uptake in men aged x, compared to boys aged 10-14. The relative rates of MMC uptake in the 15-19, 20-24, 25-49 and 50+ age groups have been set to 0.59, 0.27, 0.14 and 0.012 respectively; these rates were chosen to ensure the model matches the age profile of MMC operations provided from PEPFAR-supported MMC programmes in South Africa. The $\theta(t)$ values are estimated from the reported number of MMC operations in year t, $\Lambda(t)$, which are shown in Table 4.5. Mathematically,

$$\Lambda(t) = \sum_x N(x,t)\left(1 - \frac{b - p(x)}{1 - p^*(x,t)}\right)\eta(x,t), \tag{4.30}$$

where $N(x, t)$ is the number of uncircumcised, HIV-negative men who are aged x at the start of year t. From the above equation, it follows that

$$\theta(t) = \Lambda(t) \Big/ \sum_x N(x,t)\left(1 - \frac{b - p(x)}{1 - p^*(x,t)}\right) R(x). \qquad (4.31)$$

Combining traditional and medical male circumcision, the net probability of male circumcision in an HIV-negative male aged x at the start of year t is

$$\psi(x,t) = \frac{p(x+1) - p(x) + \left(1 - p^*(x,t) - \left(b - p(x)\right)\right)\theta(t)R(x)}{1 - p^*(x,t)}. \qquad (4.32)$$

For HIV-positive men, the rate of circumcision is calculated using equation (4.27), i.e. assuming that they would only get circumcised traditionally.

Table 4.5: Annual numbers of MMC operations performed through MMC campaigns

Year	Operations	Source
Pre-2008	0	-
2008/09	5190	[232]
2009/10	9168	[232]
2010/11	131117	[232]
2011/12	347973	[233]
2012/13	422262	[233]
2013/14	331668	[234]
2014/15	508404	[235]
2015/16	518130	[236]
2016/17	446678	[237]
2017/18	540327	[238]
2018/19	595006	[239]

Based on equation (4.30), the model estimates the average annual probability of MMC in boys aged 10-14, over the period from mid-2014 to mid-2019, to be 0.25. This value is assumed to continue to apply in all years after 2019.

Figure 4.2 shows that with these model assumptions, the model estimates of the proportion of the male population circumcised are roughly consistent with national survey data, after taking into account uncertainty regarding the sensitivity and specificity of self-reported circumcision status (the model estimates of the fraction circumcised have been adjusted, allowing for variation in sensitivity and specificity within plausible ranges [223-229]). However, the model estimates of the fraction of men circumcised appear slightly too low when compared against the data from the 2017 HSRC household survey [5].

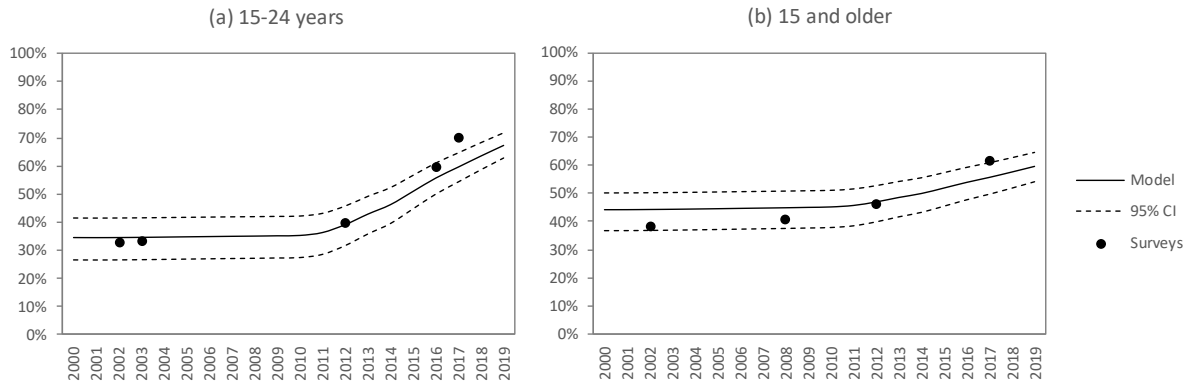


Figure 4.2: Proportion of men who report being circumcised
Model estimates of the 'true' prevalence of male circumcision have been adjusted to reflect inaccuracies in self-report. We use bootstrapping of sensitivity and specificity estimates from seven previously published studies [223-229] to derive the average model estimate and the 95% confidence intervals around the model estimates. Survey estimates are taken from the 2003 and 2016 DHSs and the 2002, 2008, 2012 and 2017 HSRC surveys.

4.8 Extensions to represent effect of pre-exposure prophylaxis (PrEP)

4.8.1 Effectiveness of PrEP

Randomized controlled trials published to date have yielded conflicting estimates of the effectiveness of PrEP, mostly because of differences in PrEP adherence across trials. Although a recent meta-analysis estimated that PrEP reduced heterosexual transmission and transmission between MSM by 46% and 66% respectively [240], these estimates are probably under-estimates, as most of the evidence included in the meta-analysis came from randomized trials that were conducted prior to the effectiveness of PrEP being established. More recent studies, conducted in the context of known PrEP efficacy, have generally found much higher levels of adherence and efficacy [241-243], suggesting that individuals are more motivated to use PrEP consistently when they know that it works. The assumed efficacy of PrEP is therefore set to 65% in heterosexuals and 85% in MSM. The assumed efficacy of 65% in heterosexuals is based on a meta-analysis that found an average 65% reduction in women's HIV risk in studies in which average PrEP adherence was at least 50% [244], and the assumed effectiveness of 85% in MSM is based on the results of the PROUD and IPERGAY studies, which both found 86% efficacy in MSM [241, 242]. The assumed greater efficacy of PrEP in MSM is supported by in vitro evidence of greater drug concentration in rectal tissue when compared to female genital tract tissue [245].

4.8.2 Risk compensation

Although data from randomized trials generally do not show evidence of risk compensation in PrEP recipients [246-248], it is difficult to extrapolate from the data collected in these randomized trials, as trial participants would have been counselled on the uncertainty regarding the efficacy of the products that were being evaluated, and even if they believed the study products to be effective, would not have known whether they were receiving the study

drug or the placebo. In an analysis of changes in behaviour after the unblinding of the Partners PrEP trial in heterosexual couples, a statistically significant 10% increase was noted in unprotected extramarital sex, amongst individuals who were receiving open-label PrEP [249]. A recent meta-analysis of PrEP studies conducted in MSM also found that PrEP use was associated with increased STI diagnosis (OR 1.26, 95% CI: 0.99-1.54) and increases in condomless sex [250]. Based on these two studies, we assume a 10% reduction in condom use among PrEP users.

4.8.3 PrEP discontinuation

Rates at which individuals discontinue PrEP are highly variable between studies, ranging from rates of 0.23 per annum in American MSM [251] to rates of 0.45 and 0.80 per annum in studies that have followed individuals following the completion of randomized controlled trials of PrEP [249, 252]. In our model we assume an average PrEP duration of 6 months in women and 1 year in men. This is based on the limited programme data available in South Africa for female sex workers (FSWs) and MSM respectively (Sarah Jenkins, personal communication). We fit simple Weibull models to the data to estimate the time from initiating PrEP to stopping PrEP; in the case of FSWs, a Weibull distribution with a mean of 4.8 months and a shape parameter of 0.45 provides an adequate fit to the data, while in the case of MSM, a Weibull distribution with a mean of 11.1 months and a shape parameter of 0.60 provides an adequate fit to the data (Figure 4.3).

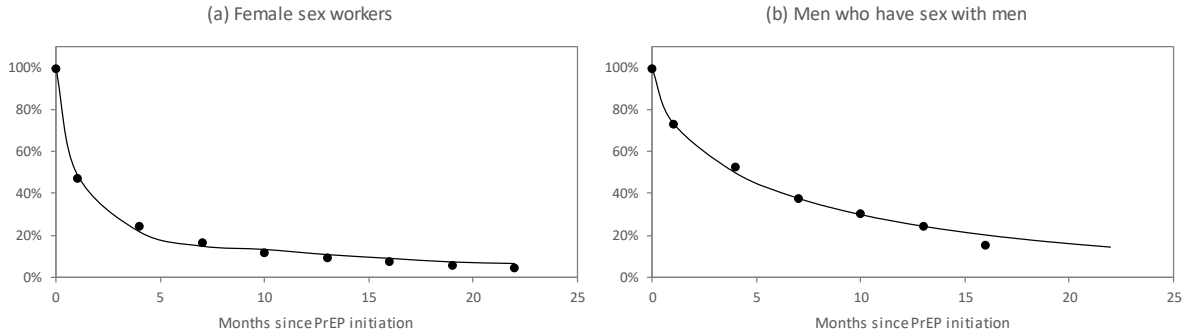


Figure 4.3: Retention in South African PrEP programmes
Data (represented by dots) are from South African PrEP programmes, as at November 2018 (Sarah Jenkins, personal communication). The solid lines represent Weibull fits to the data.

4.8.4 Adoption of PrEP

Model assumptions about rates of PrEP uptake in HIV-negative FSWs and MSM have been set in such a way that the model matches roughly the recorded total numbers of PrEP recipients from routine programme data (Sarah Jenkins and Natsai Shoko, personal communication). These data have been adjusted to exclude individuals who have stopped PrEP (in contrast to most of the cited statistics, which relate to the number of individuals who have *ever* been on PrEP). The assumed annual rates of PrEP uptake in different groups are shown in Table 4.6. Consistent with our previous modelling work [253], we assume that in low-risk sexually active AGYW, the rate of PrEP uptake is 0.33 times that in high-risk

sexually active AGYW, based on early data on correlates of PrEP uptake in pregnant Kenyan women [254].

Table 4.6: Annual rates of PrEP uptake

Year	Sex workers	MSM	Sexually active AGYW	
			High risk	Low risk
2016-17	0.0400	0.0010	0.0000	0.0000
2017-18	0.0500	0.0070	0.0002	0.0001
2018-19	0.0800	0.0090	0.0060	0.0020
2019+	0.0900	0.0090	0.0140	0.0046

AGYW = adolescent girls and young women. MSM = men who have sex with men.

Figure 4.3 shows that with these assumptions the model produces estimates of numbers of PrEP users roughly consistent with routine data up to January 2020. In the absence of more recent data, the rates of initiation assumed for 2019-20 are assumed to apply in all future years.

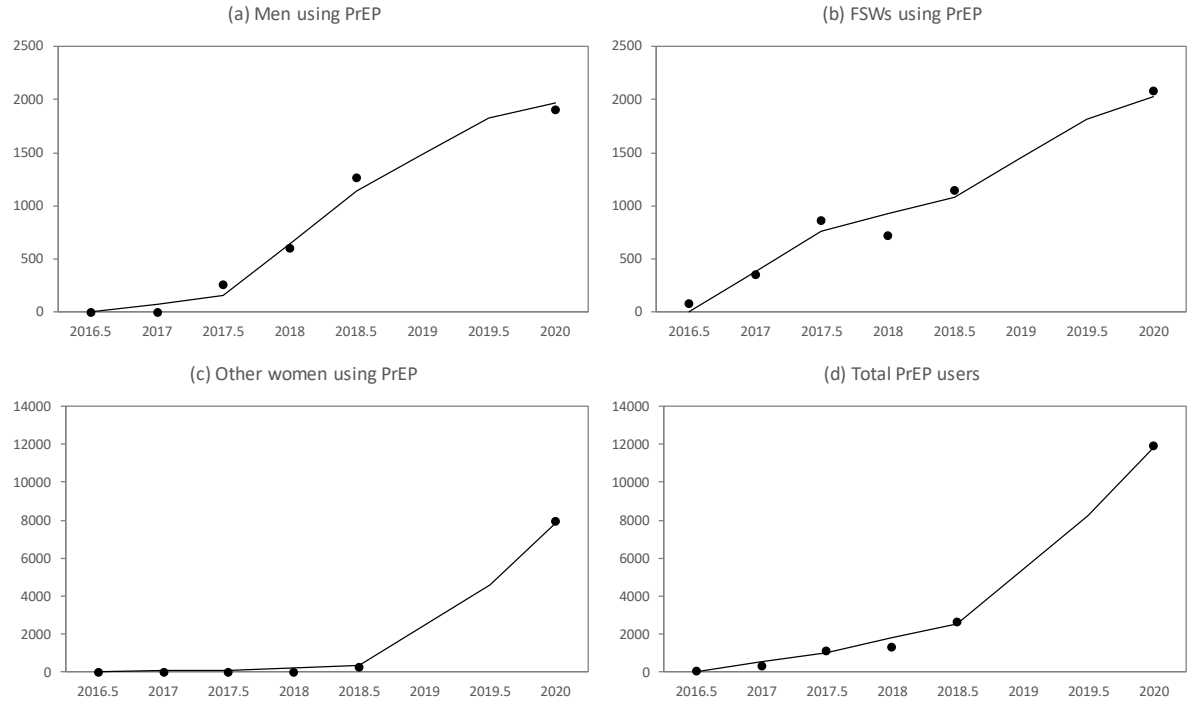


Figure 4.3: Total numbers of individuals receiving PrEP
Data (represented by dots) are from South African PrEP programmes (Sarah Jenkins and Natsai Shoko, personal communication). The solid lines represent Thembisa estimates.

A limitation of the data is that the 'risk group' classification is based on the type of facility that provides PrEP rather than the actual profile of PrEP users. For example, some of the individuals classified in the data as receiving PrEP through FSW services might not be FSWs.

5. Model of mother-to-child transmission and paediatric HIV

The model allows for two types of mother-to-child transmission (MTCT): perinatal transmission (at or before the time of birth, i.e. intrapartum or intrauterine) and postnatal transmission (transmission occurring due to breastfeeding). HIV survival rates in HIV-infected children are assumed to depend on whether infection is acquired perinatally or postnatally.

5.1 Perinatal transmission

The model of mother-to-child transmission has been described elsewhere [255], and key parameters are summarized in Table 5.1. Perinatal transmission probabilities are assumed to depend on the mother's HIV disease stage and the type of antiretroviral prophylaxis that she receives. Proportions of women who receive testing for HIV and proportions of women who start long-term ART (if they are ART-eligible) have both been presented in Table 3.2.

5.1.1 Short-course antiretroviral prophylaxis

Of women who test positive during pregnancy but do not start long-term ART, 71% are assumed to have received single-dose nevirapine (sd NVP) in the period up to 2011/12. A fraction $D(t)$ of these women also receive short-course AZT (dual therapy), and the fraction of women not receiving sd NVP who receive short-course AZT is assumed to be proportional to $D(t)$. The fraction of diagnosed women not starting long-term ART, who receive some form of short-course ARV prophylaxis, is thus $0.71 + (1 - 0.71) \times D(t) \times 0.79$, where 0.79 is the assumed constant of proportionality (Kate Kerber, personal communication, based on national survey data [256]). The $D(t)$ parameters are assumed to increase from zero in 2002/3 up to 90% in the 2010-2012 period [256, 257]. However, $D(t)$ parameters are assumed to decline to zero in 2014, following the introduction of WHO option B, which recommended triple-drug prophylaxis for all HIV-positive women, regardless of CD4 count. It is nevertheless assumed that even after the introduction of WHO option B, 71% of HIV-diagnosed mothers who do not start triple-drug therapy prior to ART initiation would receive sd NVP as an emergency prophylaxis (typically in situations where HIV is diagnosed only in labour).

Table 5.1: Mother-to-child transmission assumptions

Parameter	Value	S.D.*	Source
Transmission rate at/before birth, from chronically-infected women with no ARV prophylaxis, with			
CD4 >500	13.4%	-	Meta-analysis of published studies [258]
CD4 350-500	15.2%	-	
CD4 200-349	25.8%	-	
CD4 <200	35.0%	-	
Transmission rate at/before birth, from acutely-infected women with no ARV prophylaxis	25.4%	-	[259-264] and previous calibration [115]
% of HIV-diagnosed women who receive single-dose nevirapine, if not starting ART	71.0%	-	Kate Kerber (pers. comm.), based on national survey data [256]
% reduction in perinatal MTCT if mother receives single-dose nevirapine only	40.0%	-	[265]
% reduction in perinatal MTCT if mother receives short-course zidovudine only	65.0%	-	[266]
% reduction in perinatal MTCT if mother receives single-dose nevirapine + short-course zidovudine	85.8%	-	[267, 268] and previous calibration [115]
Transmission rate at/before birth, from women on long-term ART pre-conception	0.3%	-	[269-274]
Probability of MTCT from chronically-infected mothers, per year of mixed feeding	14.0%	2.5%	Meta-analysis [275], adjusted to reflect effect of excluding EBF
Probability of MTCT from acutely-infected mothers, per month of mixed feeding	16.0%	3.0%	Derived from meta-analysis [255]*
Ratio of postnatal transmission risk per month of EBF to postnatal transmission risk per month of mixed feeding	0.50	0.15	[276, 277]
% reduction in monthly postnatal MTCT risk if child receives extended nevirapine prophylaxis	60.0%	-	[278-280]
% reduction in monthly postnatal MTCT risk if mother receives long-term ART			1 - average MTCT rate per month of BF divided by the rate in women not on ART [275]
ART initiated during pregnancy	78%	-	[281-290]
ART initiated before conception	96%	-	[271, 274, 291]

* Standard deviation (SD) is specified only for those parameters that are considered in the uncertainty analysis; the corresponding values specified in the previous column represent the prior means (see Appendix E for more detail). EBF = exclusive breastfeeding; MTCT = mother-to-child transmission.

5.1.2 Effectiveness of long-term maternal ART

For women who start ART during pregnancy, in CD4 stage *s*, the probability of perinatal transmission is assumed to be of the form

$$a + b_s R^x, \tag{5.1}$$

where a is the minimum transmission risk (the risk that might be expected in women who started ART prior to conception), b_s is the difference between the maximum and minimum transmission risk (the maximum being that which applies if ART is initiated just prior to delivery), R is the factor by which the difference reduces per week of ART prior to delivery, and x is the number of weeks of ART received prior to delivery. If $g(x)$ is the probability density function describing the distribution of ART durations in the baseline scenario (before any interventions to improve ART initiation during pregnancy), and this density is assumed

to be of gamma form, then the average probability of perinatal transmission in the baseline scenario is

$$\int_0^\infty g(x)\left(a + b_s R^x\right) dx = a + b_s \int_0^\infty \frac{\lambda(t)^\alpha x^{\alpha-1} \exp\left(-\lambda(t)x\right)}{\Gamma(\alpha)} R^x dx$$
$$= a + b_s \left(\frac{\lambda(t)}{\lambda(t) - \ln(R)} \right)^\alpha . \qquad (5.2)$$

where α and $\lambda(t)$ are the parameters of the gamma distribution. Based on South African data sources [269, 270, 292, 293], the mean and standard deviation of the gamma distribution in the period before 2010 have been set to 10.6 weeks and 8 weeks respectively (α = 1.7556 and $\lambda(t)$ = 0.1656 for t < 2010), and the R parameter has been set to 0.9. Parameter a has been set to 0.003, the average transmission risk from studies that evaluated the perinatal transmission rate from mothers who started ART prior to conception [269-274] (Table 5.1).

The remaining b_s parameter is estimated by equating expression (5.2) to the known average perinatal transmission probability that existed in the baseline scenario. This is calculated separately for women who started ART during pregnancy with CD4 <200 (s = 5) and women who started ART in pregnancy at higher CD4 counts (s < 5); based on previous research these average transmission probabilities are assumed to be 0.036 and 0.013 respectively [269, 270, 272, 281-283, 285, 286, 289, 290, 292, 294, 295]. The resulting estimates of the b_s parameter are 0.078 and 0.024 respectively.

In the period since 2010, there has been substantial improvement in the average duration of ART. The South African 2010 PMTCT guidelines recommended integration of ART provision into PMTCT services [296], which led to more rapid initiation of ART during pregnancy. For example, Van Schalkwyk *et al* [293] found that the median duration of ART prior to delivery increased from 7.7 weeks in the 2008-9 period to 13.1 weeks in 2010 following the introduction of the new guidelines. A similar median of around 12 weeks has been observed in the period following 2010 in the Eastern Cape, and even higher rates of ART uptake were measured from 2012 [297]. Stinson *et al* [298] documented a more substantial difference (about 7 weeks) in the median time to ART initiation when comparing the ART referral model to the integrated ART model. There have also been steady improvements over time in the mean gestational age at first antenatal booking; for example, the Department of Health [233] reports that the proportion of mothers who had their first antenatal visit before 20 weeks gestation has increased from 37.5% in 2010/11 to 50.6% in 2013/14. It is therefore assumed that the mean duration of ART increased by 50% in 2010-12 (relative to the mean duration in the pre-2010 period). This means setting $\lambda(t)$ = 0.1104 over the 2010-2012 period, which leads to a 22% reduction in the probability of perinatal transmission from mothers with initial CD4 counts <200 cells/μl. Following the introduction of WHO option B at the start of 2013, it is likely that the delay in ART initiation would have been reduced even further, since the removal of the CD4 restriction would have eliminated the delay associated with CD4 testing. We assume that after 2013, the average ART duration before delivery increases by 70% (relative to baseline), which is roughly consistent with what would be expected if all pregnant women starting ART during pregnancy did so soon after their first antenatal visit.

In Thembisa 4.3, we make provision for the possibility that some women who initiated ART prior to conception might not be on ART at the time of conception. The assumed proportions of women who are interrupting ART depend on the time since ART initiation, and are specified in section 3.5. Women who have interrupted ART are assumed to have the same probability of starting ART during pregnancy as women who are ART-naïve, and their risk of mother-to-child transmission (if they restart ART) is modelled in the same way as for ART-naïve pregnant women.

5.1.3 HIV incidence in pregnancy and retesting in late pregnancy

The first antenatal visit is assumed to occur at 23 weeks gestation [21, 299, 300] and delivery at 39 weeks [300], on average, so that the average time in which a woman seronegative at her first visit can acquire HIV before delivery is 20 weeks if a 4-week window period is assumed [301]. The probability that a pregnant woman seronegative at her first antenatal visit acquires HIV before delivery is therefore calculated as the annual HIV incidence rate in pregnant women multiplied by a factor of 0.38 (20/52). The probability that a woman who acquires HIV in late pregnancy transmits HIV perinatally is difficult to determine precisely, and a value of 25.4% has been assumed (Table 5.1). This probability applies if the woman receives no antiretroviral prophylaxis.

In the period up to 2006, there is assumed to have been no retesting prior to delivery of mothers HIV-negative at their first antenatal visit. Recent studies suggest that the proportion of women testing negative who get tested again in late pregnancy has been steadily increasing over time [302, 303], with the 2016/17 DHIS data suggesting a proportion close to 100%. A retesting frequency of 95% is therefore assumed from 2016 onward. Women who are diagnosed HIV-positive following retesting are assumed to be as likely to receive short-course ARV prophylaxis and long-term ART as women who are diagnosed at their first antenatal visit.

5.2 Postnatal HIV transmission

5.2.1 Infant feeding practices up to 2011

Among HIV-negative mothers and undiagnosed HIV-positive mothers, 86.7% are assumed to breastfeed, and in those who breastfeed the duration of breastfeeding is modelled using a Weibull distribution with a median of 18 months and a shape parameter of 2 [21]. All of these women are assumed to practise mixed feeding, as exclusive breastfeeding (EBF) is usually of very short duration in HIV-negative (or undiagnosed positive) mothers [21, 304]. Of women who were diagnosed HIV-positive antenatally in the period up to 2011, it is assumed 56% avoided breastfeeding completely [145], 30% practised EBF and 14% practised mixed feeding [305]. HIV-diagnosed women who practised EBF are assumed to have done so for a median of 2 months (up to a maximum of 6 months), after which 30% are assumed to have discontinued breastfeeding completely and the remainder practised mixed feeding (i.e. continued breastfeeding while introducing complementary feeds), for a median of 7 months [305-307]. The median duration of mixed feeding in HIV-diagnosed mothers is assumed to

be the same regardless of whether mixed feeding was provided from birth or following a period of EBF.

5.2.2 Infant feeding practices after 2011

The benefits of EBF have been increasingly emphasized following the Tshwane declaration [308], with guidelines recommending 6 months of EBF for all mothers (as well as continued mixed feeding after 6 months) and the phasing out of the free provision of formula milk for HIV-positive mothers. The proportion of HIV-diagnosed women who avoid breastfeeding is assumed to have declined from 56% in 2010/11 to 20% in 2013/14, in line with data from a series of national PMTCT surveys [309].

5.2.3 Postnatal transmission probabilities

Table 5.1 summarizes the assumptions regarding postnatal HIV transmission probabilities, per month of breastfeeding. Exclusive breastfeeding is assumed to be associated with a reduced risk of transmission relative to mixed feeding, while women who seroconvert during breastfeeding are assumed to be at a significantly increased risk of transmitting HIV to their infants. The duration of this period of increased postnatal transmission risk is assumed to be the same as the duration of acute infection.

Following the revision to the South African PMTCT guidelines in 2010 [296], HIV-positive mothers who breastfed but did not start ART were provided with extended nevirapine prophylaxis to administer to their infants during the breastfeeding period. Although there is a lack of data on the uptake of this prophylaxis, it is assumed that 80% of all breastfed children whose HIV-positive mothers were not on ART received this prophylaxis. After the introduction of Option B in 2013, it is assumed that this provision of extended nevirapine (throughout the breastfeeding period) was phased out.

The modelling of the uptake of long-term ART in pregnant HIV-positive women has been described in section 3.3. In addition to this, in the period between the start of 2013 and the end of 2014 (prior to adoption of WHO Option B+), women who were not eligible for long-term ART were eligible for short-term ART (triple-drug therapy) for the duration of pregnancy and the breastfeeding period. The rate of short-term ART uptake during pregnancy is assumed to have been the same as the rate of long-term ART uptake in the corresponding year.

For women who are on ART while breastfeeding, the monthly HIV transmission risk is assumed to depend on whether they started ART before conception or during pregnancy (Table 5.1). The assumption of a 78% reduction in postnatal transmission rates in women who started ART during pregnancy, relative to breastfeeding mothers who are untreated, is calculated as one less the ratio of the average monthly postnatal transmission risk in various studies (0.0017) to the average monthly transmission risk of 0.0077 for untreated mothers in a meta-analysis [275]. Similarly, the 96% reduction in postnatal transmission risk from mothers who started ART prior to their pregnancy is calculated as one less the ratio of the average monthly transmission risk from these mothers (0.0003) to that in untreated mothers (0.0077).

5.3 Paediatric HIV survival

The structure of the paediatric HIV survival model is illustrated in Figure 5.1. The model is an adaptation of a previous model of paediatric HIV survival [310], which has been extended to include HIV diagnosis after infancy. HIV-infected children are assumed to progress from an early disease stage to a late disease stage in the absence of ART (late disease is defined as having met the immunological or clinical criteria that were previously used to determine ART eligibility under the 2006 WHO paediatric ART guidelines [311]). HIV-related mortality in untreated children is assumed to occur only in the late disease stage. Children who are infected postnatally are assumed to have a slower rate of progression from early disease to late disease, but after progression to late disease and after ART initiation, age-specific mortality rates are assumed to be the same regardless of timing of transmission. Although the model distinguishes perinatally-infected children according to whether or not they were PMTCT-exposed, this is assumed to have no effect on their rate of HIV disease progression.

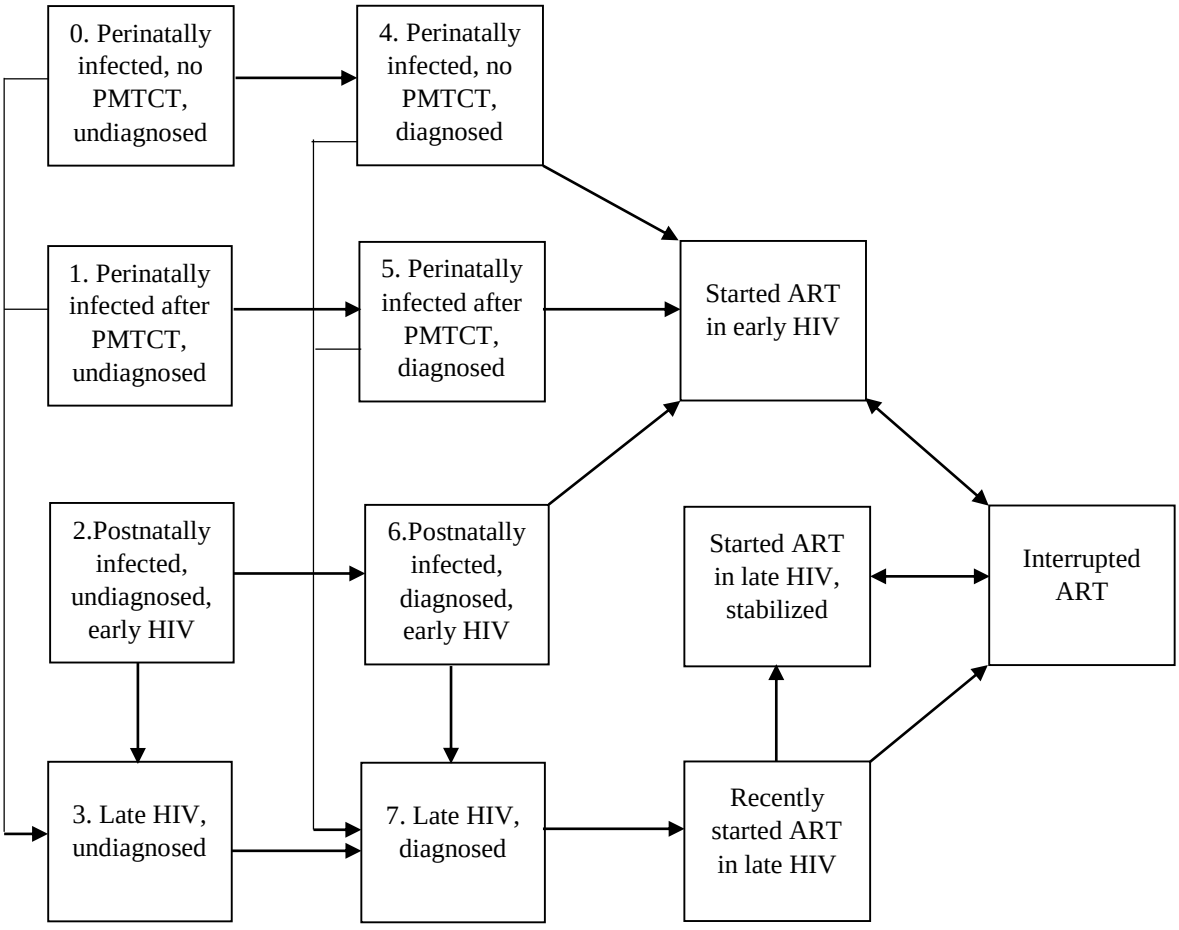


Figure 5.1: Multi-state model of HIV survival in HIV-positive children
All children are assumed to experience non-AIDS mortality rates that vary by age and sex, and all children who are in the late disease stage or are ART-experienced are assumed to be at risk of AIDS mortality (not shown).

5.3.1 Untreated disease progression and mortality

Since the rate of progression to late disease declines as children age, the time to reaching late disease is assumed to follow a Makeham distribution, with the hazard rate in perinatally-infected children aged x being

$$\eta(x) = G_p + \left(H_p \times c^x\right), \tag{5.3}$$

where G_p is the annual rate of progression in older children, H_p is the excess rate of progression in neonates, and c is the factor by which the excess rate of progression is reduced per year of age. Children who acquired HIV postnatally are assumed to progress to late disease at rate $\theta G_p + H_p (\theta c)^x$, where θ is a constant scaling factor. This functional form was chosen to ensure that there is relatively less of a reduction in disease progression in postnatally infected infants soon after birth (when $x = 0$) than at older ages, in line with data showing poorer HIV survival in HIV-infected infants who acquire HIV soon after birth compared to those acquiring HIV at older ages [312]. The assumed parameter values and the data sources on which they are based are summarized in Table 5.2.

Table 5.2: Paediatric HIV survival assumptions (ages <10)

Parameter	Symbol	Value	S.D.*	Source
Children infected at/before birth				
Annual rate of progression to late disease in older children	G_p	0.40	0.10	$\theta G_p = 0.14$ is consistent with rates of progression observed by Charlebois *et al* [313] in children aged ≥1 year
Excess annual rate of progression to late disease in neonates	H_p	2.00	0.50	[314, 315]
Excess progression reduction factor, per year of age	c	0.40	0.10	[314-317]
Relative rate of progression to late disease if infected after birth	θ	0.35	0.15	[318-321]
Children in late disease, untreated				
Annual rate of AIDS mortality in older children	G_m	0.12	0.03	[322, 323]
Excess annual rate of AIDS mortality in neonates	H_m	3.50	0.70	Based on fitting model to mortality data from children diagnosed with
Excess mortality reduction factor, per year of age	d	0.20	0.10	HIV-related symptoms at different ages [323]
Relative rate of AIDS mortality in 'stabilized' children who started ART in late disease	Φ_1	0.10	0.05	Based on fitting model to mortality data from IeDEA Southern Africa Collaboration [324]
Reduction in mortality (on log scale) per unit increase in rate of ART initiation over last 3 years	m	7.5	3.5	See Appendix D
Children who started ART while in early disease				
AIDS mortality at age 0	β	0.06	-	See Appendix D
Relative rate of AIDS mortality per year of age	P	0.20	-	See Appendix D

* Standard deviation (SD) is specified only for those parameters that are considered in the uncertainty analysis; the corresponding values specified in the previous column represent the prior means (see Appendix E for more detail).

In the absence of ART, children in the late disease stage are assumed to die from AIDS at rate $\mu(x)$ at age x. As this mortality rate appears to decline with increasing age [322, 323], a Makeham distribution is again used to model the time from reaching late disease to death. It is therefore assumed that the AIDS-related mortality rate is of the form

$$\mu(x) = G_m + \left(H_m \times d^x\right), \tag{5.4}$$

where G_m is the annual rate of mortality that would be expected in older children in late disease, H_m is the excess AIDS mortality rate in neonates, and d is the factor by which this excess mortality risk declines per year of age. Assumed parameter values are summarized in Table 5.2.

5.3.2 Survival after ART initiation

Children who start ART after having progressed to late disease are assumed to remain in a 'high risk' phase for an average period of three months after starting ART, if they do not die. After 'stabilizing' on ART, these children are assumed to experience lower mortality rates. The rates of AIDS mortality in the 'high risk' and 'stabilized' states are assumed to be $\Phi_0\mu(x)$ and $\Phi_1\mu(x)$ respectively, and are thus higher in children receiving ART at young ages than in children on ART at older ages. A prior distribution is assigned to represent the uncertainty around the Φ_1 parameter, and for simplicity we assume that $\Phi_0 = 0.5 \times (1 + \Phi_1)$, i.e. the reduction in mortality in the 'high risk' phase is half of that in the stabilized phase. These rates are also adjusted to take into account changes over time in the relative severity of untreated late disease. This is to compensate for the selective nature of ART initiation; in the early stages of the ART rollout, it is the sickest of the children with late disease who start ART, but as ART uptake expands, the average disease severity among children starting ART declines. For example, in the 'high risk phase' after starting ART in late disease in year t, the mortality rate is calculated as

$$\Phi_0\,\mu(x)\,[0.43 + (1 - 0.43)\,\exp(-m\,r_{t\text{-}})], \tag{5.5}$$

where $r_{t\text{-}}$, is the average rate of paediatric ART initiation over the previous three years, m is a scaling parameter, and 0.43 is the assumed ratio of the minimum ART mortality rate (the rate that might be expected if all children started ART in the early phase of late disease) to the mortality rate that applied in the earliest phase of the paediatric ART rollout. This implies that the mortality rate in the 'high risk' phase declines exponentially towards a minimum rate as the rate of ART initiation in children increases, with the m parameter determining the pace of this exponential decay. The adjustment is analogous to that described for adults in section 3.4, and is described more fully in Appendix D.

The AIDS mortality rate in children who start ART in early disease, $\psi(x)$ at age x, was previously calculated as an adjustment to the rate assumed in children who started ART in late disease and who were 'stabilized' [310]. However, this was found to produce implausibly high rates of mortality, especially in the context of birth testing. In the new version of Thembisa, we have therefore estimated these mortality rates directly from South African ART programmes participating in the International Epidemiology Databases to Evaluate AIDS (IeDEA) collaboration [324]. The rate of AIDS mortality in children starting ART in

early disease is calculated as $\psi(x) = \beta P^x$, where β is the HIV mortality rate that applies at age 0, and P is the factor by which the AIDS mortality rate is reduced per year of age. These parameters have been set to 0.06 and 0.2 respectively, based on attempts to fit the model to IeDEA data (for a more detailed explanation, see Appendix D).

Although the model allows for children to interrupt ART and resume ART (Figure 5.1), there is currently substantial uncertainty around the rates of ART interruption in children, and we therefore assume the same rates of ART interruption and resumption as for adults (see Appendix G). There is also substantial uncertainty regarding the extent to which mortality changes while children are interrupting ART, and we therefore do not attempt to differentiate mortality rates in interrupters from those currently on ART.

5.3.3 HIV diagnosis and ART initiation

The model assumes that a proportion of children born to HIV-positive mothers receive PCR testing for HIV soon after birth (until 2015, guidelines recommended PCR screening at 6 weeks and since then screening has been done both at birth and at 10 weeks). Of these screened infants, a proportion of those eligible for ART are assumed to start ART, which is assumed to occur either at birth or at 2 months of age (the latter being a crude approximation to the timing that might be expected if screening occurs at 6 weeks or 10 weeks). Mathematically, the number of perinatally-infected infants who start ART at birth or at 2 months, following PCR screening, is calculated as

$$S^0(t) = \left(\sum_{s=0}^{1} \left(N_s(0,t)V(0,t)\pi_s(0) + N_s(2,t)V(2,t)\pi_s(2) \right) E_0(t) + N_3(2,t)V(2,t)E_3(t) \right) l, \quad (5.6)$$

where $N_s(x, t)$ is the number of infected infants at the age of x months, in stage s of infection; $V(x, t)$ is the fraction of children born to HIV-positive mothers who receive PCR testing at age x in year t; $\pi_s(x)$ is the sensitivity of the PCR in infants in stage s aged x; $E_0(t)$ is the fraction of infants in early disease who are eligible to receive ART in year t; $E_3(t)$ is the fraction of children in advanced disease who are eligible to receive ART in year t; and l is the fraction of ART-eligible diagnosed infants who link to ART care soon after diagnosis. As shown in Figure 5.1, stages 0 and 1 correspond to infants in early disease who were antenatally PMTCT-unexposed and PMTCT-exposed respectively, and stage 3 corresponds to infants in the late stage of HIV disease (all ART-naïve). The time-dependent parameters are summarized in Table 5.3. Rates of PCR testing at 6 weeks are based on public sector statistics [148, 325], adjusted to reflect under-count due to late immunization [326, 327] and over-count due to non-return of test results to caregivers [328-330]. After birth testing was introduced in 2015, birth screening coverage increased to 68.7% in 2015-16 [147, 331], and to around 90% thereafter [332]. Limited information is available on the rate of screening at 10 weeks since the introduction of the new screening policy, but data suggest that screening coverage at 10 weeks may be lower than has historically been observed at 6 weeks [331, 333]. For example, Kalk *et al* [331] found that the fraction of infants receiving HIV testing at 6-10 weeks dropped from 93% in the period before birth testing to 80% after the introduction of universal birth testing. We have assumed 80% coverage from 2015 onward; this is conservative because the model treats the probabilities of birth and 6-10 week testing as independent, when in reality there is a negative association that is likely to result in a high

overall fraction of infants screened (either at birth or at 6-10 weeks). PCR sensitivity levels at 2 months have been set at 76%, 81% and 100% for stages 0, 1 and 2 respectively, based on a previous model of perinatal transmission [334], assuming that all infants who are tested for HIV would at least have received NVP prophylaxis postnatally [296]. Sensitivity levels at birth have been set to 38% and 75% for stages 0 and 1 respectively (no infants are assumed to be already in advanced disease at birth). Although children in late disease have been eligible for ART since 2004 [151], ART eligibility for infants in early disease only became official policy in 2010 [335], with some earlier provision following the 2008 WHO guideline revision [336]. The fraction of eligible, diagnosed infants who link to care and start ART (l) is set to 80%, based on evidence of high infant uptake of ART after EID in South African settings [330, 333, 337].

Table 5.3: HIV diagnosis and ART eligibility in HIV-positive children

	Fraction tested at 6 or 10 weeks ($V(2,t)$)	Fraction tested at birth ($V(0,t)$)	Early ART eligibility			Late ART eligibility ($E_3(t)$)
			Infants ($E_0(t)$)	Ages 1-4 ($E_1(t)$)	Ages 5-12 ($E_2(t)$)	
Pre-2004	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2004-2006	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
2006-2007	8.5%	0.0%	0.0%	0.0%	0.0%	100.0%
2007-2008	19.1%	0.0%	0.0%	0.0%	0.0%	100.0%
2008-2009	29.5%	0.0%	20.0%	0.0%	0.0%	100.0%
2009-2010	40.1%	0.0%	60.0%	0.0%	0.0%	100.0%
2010-2011	53.0%	0.0%	100.0%	0.0%	0.0%	100.0%
2011-2012	60.8%	0.0%	100.0%	0.0%	0.0%	100.0%
2012-2013	68.9%	0.0%	100.0%	100.0%	0.0%	100.0%
2013-2014	84.8%	0.0%	100.0%	100.0%	0.0%	100.0%
2014-2015	92.0%	0.0%	100.0%	100.0%	0.0%	100.0%
2015-2016	92.0%	68.7%	100.0%	100.0%	0.0%	100.0%
Post-2016	80.0%	90.0%	100.0%	100.0%	100.0%	100.0%

The model also makes provision for other HIV testing in children (independently of the screening programmes at birth and 6-10 weeks) and resulting ART initiation. The number of children who start ART in month t as a result of this other HIV testing is calculated as

$$S^1(t) = \left[\sum_{s=0}^{2} \left(\sum_{x=18}^{59} N_s(x,t)\,\tau_s(x,t) E_1(t) + \sum_{x=60}^{179} N_s(x,t)\tau_s(x,t)\,E_2(t) \right) + \sum_{x=0}^{179} N_3(x,t)\,E_3(t)\tau_3(x,t) \right] l_3(t) \quad (5.7)$$

where $E_1(t)$ and $E_2(t)$ are the fractions of children in early disease who are eligible to receive ART in year t, for the 1-4 and 5-14 age groups respectively (Table 5.3); $\tau_s(x, t)$ is the monthly probability of HIV testing in stage s in year t; and $l_3(t)$ is the fraction of newly-diagnosed children who link to ART after diagnosis. The $E_1(t)$ and $E_2(t)$ parameters have been set to reflect the changes in ART eligibility criteria in children over time, which in 2012 included children in early disease aged 1-4 [154], and which were extended to all children in 2016.

HIV testing rates in older children in early HIV disease ($\tau_s(x, t)$ for $s < 3$ and x aged 60 months or older) are assumed to be a constant multiple of those in sexually experienced girls aged 15, $\varphi(t)$. The multiple is φ_1 in the period up to 2005, φ_2 in the period following 2010, and the multiple is linearly interpolated between the φ_1 and φ_2 values over the 2006-2009 period. This change in multiple over time is allowed for in the model as it is hypothesized that there may have been substantial differences in testing patterns following the introduction of national HIV testing campaigns in South Africa in 2010. Prior distributions have been specified to represent the uncertainty around the φ_1 and φ_2 values (see Appendix E). HIV testing rates in children in advanced HIV disease ($\tau_3(t)$) are assumed to be a constant multiple of those in early disease, i.e. $\tau_3(x, t) = \tau_0(x, t) \times Q$; a uniform (0, 1) prior is specified to represent the uncertainty around the $1/Q$ parameter (see Appendix E). In younger children (aged 19-59 months), the rate of testing is assumed to be 1.8 times that in the 60-179 month age group, based on routine data on total numbers of tests performed in children over the 2015-18 period (Tshepo Molapo, personal communication). In children aged 18 months, it has been policy to conduct HIV screening since 2008 [338], although implementation has been variable, with some provinces conducting universal testing and others limiting testing to HIV-exposed children (Ameena Goga, personal communication). Based on the routine data for the 2015-2018 period, we assume that 20% of children aged 18 months get tested in every year after 2008, although we lack data on the extent of 18-month testing in the period before 2008. Mathematically, the testing rate in children (ignoring screening at birth and 2 months) is

$$\tau_s(x,t) = \begin{cases} \varphi(t)k(t) & \text{for } 60 \le x < 179 \text{ and } s < 3 \\ \varphi(t)k(t)Q & \text{for } 60 \le x < 179 \text{ and } s = 3 \\ \varphi(t)k(t)J & \text{for } 18 < x < 60 \text{ and } s < 3 \\ \varphi(t)k(t)JQ & \text{for } 0 \le x < 179 \text{ and } s = 3 \\ F(t) & \text{for } x = 18 \\ 0 & \text{for } x < 18 \text{ and } s < 3 \end{cases} \quad (5.8)$$

where $k(t)$ is the rate of HIV testing in sexually-experienced non-pregnant girls aged 15, $\varphi(t)$ is the relative rate of testing in virgins, J is the relative rate of testing at ages 19-59 months (relative to 60-179 months) and $F(t)$ is the fraction of children who are tested for HIV at 18 months in year t. Because HIV testing in children under the age of 18 months requires a PCR test rather than a standard rapid test, and because PCR testing is more complex logistically, HIV testing below age 18 months is assumed to occur only if there is a clinical suspicion of HIV (i.e. the child is in advanced disease, as reflected in equation 5.7) or because the child receives HIV screening at the standard birth/2 month screening (equation 5.6).

Due to lack of information on rates of paediatric linkage to ART after diagnosis, outside of the context of early infant diagnosis, the rates of linkage soon after diagnosis ($l_3(t)$) are assumed to be the same as those assumed for newly diagnosed adults with OIs (see section 3.3.2). If children do not initiate ART at the time of HIV diagnosis, the model allows for later ART initiation, provided they are eligible. Similar to the approach adopted in modelling ART initiation in adults, the approach is to calculate the rate of ART initiation in children in late disease from the reported total numbers of children starting ART in month t ($S(t)$), after subtracting the model estimate of the number of children starting ART immediately after diagnosis. Similar to equation (3.6),

$$S(t) - S^0(t) - S^1(t) \approx \rho(t)\sum_{x=0}^{179} N_7(x,t) + \rho(t) \times \delta \times \sum_{s=4}^{6}\left(\sum_{x=0}^{11} N_s(x,t)E_0(t) + \sum_{x=12}^{59} N_s(x,t)E_1(t) + \sum_{x=60}^{179} N_s(x,t)E_2(t)\right) \quad (5.9)$$

where $S(t)$ is the total number of children (aged <15) starting ART in month t; stages 4-7 correspond to the HIV-diagnosed but ART-naïve stages (Figure 5.1); $\rho(t)$ is the monthly probability of ART initiation in month t, in children who are in late disease; and δ is the relative rate of ART initiation in early disease compared to advanced disease. The relative rate of ART initiation in early disease compared to advanced disease is uncertain, and a value of 0.5 has been assumed. By rearranging the terms in equation (5.9), $\rho(t)$ can be estimated on a monthly basis, for those periods in which absolute numbers of children starting ART are specified. The assumed total numbers of children starting ART are shown in Table 3.4 for each year up to mid-2019 (monthly numbers are calculated by dividing these annual totals by 12).

In the period after mid-2019, the rate of ART initiation is calculated based on assumed average times to ART initiation (in months) after progressing to late disease. The baseline results suggest that over the 2014-19 period the average treatment delay for children with advanced disease ($1/\rho(t)$) was approximately 51 months (Table 3.4), and this same parameter value has been assumed in the post-2023 period. (In the period between 2019 and 2023, the $\rho(t)$ parameter is interpolated linearly between the estimated rate in 2018/19 and the assumed ultimate rate of 1/51 per month). The long average treatment delay relative to adults (around 20 months in women and 36 months in men, as discussed in section 3.3.5) is partly because of differences in disease staging in adults and children (the adult rates apply at CD4 counts <200 cell/μl, whereas the child rates apply in 'late disease').

5.3.4 Transition from paediatric to adult HIV staging

Transitions from the paediatric HIV disease categories to the adult HIV disease stages are modelled on the assumption that late disease is equivalent to a CD4 count of <200 cells/μl in older children, to be consistent with the definition of late disease [311]. This means that on reaching age 10, children who are in the late HIV stage and ART-naïve get moved into the CD4 <200/μl category, and children who are ART-naïve and in early disease on reaching age 10 are divided equally between the CD4 500+, 350-499 and 200-349 categories. HIV survival in 10-14-year olds is modelled according to the disease progression and mortality assumptions specified for adults, although the model of ART initiation in 10-14-year olds remains consistent with that in children under the age of 10. For the sake of simplicity, equations (5.7) and (5.9) do not reflect the changes in HIV disease stage definitions that occur after age 10.

6. Demographic assumptions

6.1 Base population

The initial population numbers in 1985, by sex and individual age from 0 to 89 and open interval 90+, were set the same as those from the ASSA2008 national model [339].

6.2 Fertility

Total fertility rates (TFRs) for the 12 months prior to each census (1996, 2001 and 2011) and Community Survey (CS, 2007 and 2016) were estimated by rescaling the age-group specific fertility rates (ASFRs) derived from the number of births in the 12 months prior to each census/survey and children ever born reported by women in the censuses [340], so that the number of births in a particular year was equal to the number estimated by projecting the number of survivors at the time of the 2011 census who were born in the province/country, backward to the time of birth. TFRs for the years between the censuses/surveys were produced in a similar way, assuming that the proportion of births to each 5-year age group of women changed linearly over time between the censuses/surveys. These estimates were then replaced by estimates for the period following the 2001 census as described below.

For the period 2002 to 2016, the approach was to derive the fertility rates from the best estimate of the number of births in each year. The best estimates were produced by averaging estimates derived from several sources. The first, covering the years from 2002 onwards, were derived by applying estimates of under-reporting to correct the vital registration birth statistics. These estimates of under-reporting were originally estimated by comparison of the numbers of registered births to those implied by the numbers counted in the 2011 census and/or the school enrolment numbers in 2011 if the various sources produced significantly different estimates. The second, covering the years from 2004 onwards, were derived from births recorded by the DHIS, plus 20% to allow for those not captured by the system (estimated from a comparison with the estimates from the first source of estimates). The third source were estimates from the CARe_3.2 projection model[1], which applies estimates of fertility derived from the responses to the fertility questions in the censuses and community surveys.

The average of these estimates (the first up to 2003, the first and second from 2004 to 2010, and all three after that) were then used to measure the completeness of the registration of births by duration since birth. These estimates were adjusted to ensure (to the extent that it is possible) that the completeness of the births reported up to and including a given year of registration was plausible.

[1] The CARe workbook is a simplified model, with the same demographic assumptions as the Thembisa model, which concentrates on the demographic impact of HIV/AIDS.

These numbers were converted into fertility rates by assuming that the proportion of births at each age remained constant at the levels estimated from the 2011 census and 2017 CS data.

TFRs for the projection years (from the middle of one year to the middle of the next) were linearly interpolated from the estimates by census year (i.e. from census anniversary in one year to the census anniversary in the next) for the period prior to 2002, and from the calendar year after that.

TFRs for the period from the middle of 1985 to middle of 1996 were estimated to be the linear trend from the ASSA2008 estimate for 1985 to the estimate for the year starting at the middle of 1996, produced above. Kinks in the estimates in the period 1998 to 2003 were removed to produce a smooth trend over time.

The age-specific fertility rates were set by applying proportions of the TFR at each individual age to TFRs for each year from 1985 to 2011.

The proportions of the TFR attributable to single ages for 1996, 2001, 2006, 2011 and 2015 were derived from the proportions attributable to five-year age groups reported by women in each census and the 2007 and 2016 Community Surveys using Beers subdivision. The proportions for individual years between the census/survey years were derived by linear interpolation. For the projection years 1985-1995, the ASSA2008 ASFRs were rescaled to the TFRs estimated above.

Beyond 2016, age-specific fertility rates are assumed to decline to a set of ultimate fertility rates at annual rates of decline. The assumed ultimate rates and annual rates of decline are both the same as assumed in the ASSA2008 models.

Fertility rates in HIV-positive women are specified as multiples of corresponding fertility rates in HIV-negative women of the same age. Mathematically, the fertility rate in HIV-positive women aged x in year t, in CD4 stage s, with HIV testing history v, ART status a and ART duration d, is calculated as

$$F(x, t)\,\Gamma(s, v, a, d), \qquad (6.1)$$

where $F(x, t)$ is the fertility rate in HIV-negative women, and $\Gamma(s, v, a, d)$ is the HIV-positive multiplier. The indices used in defining the multiplier are specified as follows:

- CD4 stage (s) is defined as 1 if the woman is in the acute phase of HIV infection, and 2, 3, 4 or 5 if the woman has progressed out of the acute phase of infection and has a CD4 count of ≥500, 350-499, 200-349 or <200 cells/μl respectively.
- HIV testing history (v) is defined as 0 if the woman has never tested for HIV, 1 if the woman has tested for HIV but not been diagnosed positive, and 2 if the woman has been diagnosed positive.
- ART status (a) is defined as 0 if the woman is ART-naïve and 1 if the woman has started ART.
- ART duration (d) is defined based on time since *first* ART initiation. The index is 0 for women who have started ART for the first time in the current year (or who have never started ART), 1 for women who started ART in the previous year, 2 or 3 for women who started ART 2 or 3 year previously respectively, and 4 for women who started ART 4 or more years previously.

One complication to note is that the model compartments for treated women are defined in terms of baseline CD4 count at ART initiation (*s*') rather than current CD4 count (*s*). As described in section 4.6, the variable $\psi_d(s|s')$ represents the fraction of surviving ART patients with current CD4 count in category *s*, in the cohort of patients who started ART with a CD4 count of *s'* and who are in ART duration category *d*. Another complication is that ART-experienced women are assumed to interrupt ART at a constant rate (and can also resume ART after an interruption). It is assumed that CD4 counts return to those just before ART initiation in the women who interrupt ART [87, 341], and the CD4 effects that apply during ART interruptions are thus the same as the baseline CD4 effects.

The $\Gamma(s, v, a, d)$ multiplier is calculated as the product of a number of adjustment factors:

$$\Gamma(s, v, a, d) = B_0\, B_1(s)\, B_2(v)\, B_3(a). \qquad (6.2)$$

The $B_1(s)$ adjustment factor represents the relative fertility rate in HIV-positive women in CD4 compartment *s* to that in HIV-positive women with CD4 counts of 500 cells/μl or higher (by definition, $B_1(1)$ and $B_1(2)$ are both 1). Based on a recent analysis of pregnancy incidence rates in HIV-positive women in the Western Cape province of South Africa [3], we set these ratios to 0.99 in the CD4 350-499 category, 0.90 in the CD4 200-349 category, and 0.66 in the CD4 <200 category. These rate ratios are consistent with CD4 effects observed in other African cohorts [63-65].

The $B_2(2)$ adjustment factor represents the relative fertility rate in HIV-diagnosed women compared to undiagnosed HIV-positive women (by definition, $B_2(0)$ and $B_2(1)$ are both 1). Based on recent Thembisa fits to paediatric HIV data sources [6], we set $B_2(2)$ to 0.939, consistent with previous studies showing that HIV diagnosis is associated with increases in condom use [71, 73, 78] and lower childbearing intentions [342-344].

The $B_3(1)$ adjustment factor represents the relative rate of fertility in women on ART when compared to women who are untreated (by definition, $B_3(0)$ is 1). Based on the previously-mentioned analysis of Western Cape data [3], the $B_3(1)$ parameter has been set to 1.35. (This is the result from a sensitivity analysis in which HIV-positive women were censored after their last visit or laboratory result, which is considered more reliable than the main analysis for the purpose of estimating differences between groups of HIV-positive women.) This higher rate of pregnancy incidence in women on ART, after controlling for recent CD4 count, is consistent with the findings of some studies [65, 345], although results from other studies have been inconsistent [346, 347].

The B_0 adjustment factor represents the relative rate of fertility in undiagnosed HIV-positive women in the early stages of HIV infection (CD4 ≥500 cells/μl), when compared to fertility in sexually experienced HIV-negative women of the same age. This parameter is difficult to estimate directly, as most studies do not report fertility rates in undiagnosed HIV-positive women, or do not include comparisons with HIV-negative women. However, one might expect B_0 to be greater than 1 if women who have recently acquired HIV are more sexually active and therefore more likely to become pregnant. In a sensitivity analysis of the Western Cape data, it was found that pregnancy incidence rates in women on ART with CD4 counts above 500 cells/μl were 1.42 (95% CI: 1.38-1.45) times those in HIV-negative women [3]. Substituting this and the other previously-assumed values into the equation for $\Gamma(s, v, a, d)$

gives 1.42 = B_0 $B_2(2)$ × 1.35. Equivalently, B_0 = (1.42/1.35)/$B_2(2)$ ≈ 1/$B_2(2)$. We use this approximation to determine B_0 from $B_2(2)$, i.e. we set B_0 to 1.065 (1/0.939).

For the purpose of calculating the HIV-negative fertility rate, $F(x,t)$, we define $N^i_{v,a,s,d}(x,t)$ to be the total number of women aged *x* with sexual experience indicator *i* (0 for virgins, 1 for sexually-experienced women), HIV testing history *v*, ART status *a*, CD4 stage *s*, and ART duration *d* years. The average fertility rate is then

$$\overline{F}(x,t) = \frac{F(x,t)\left[\sum_{v} N^1_{v,0,0,0}(x,t) + \sum_{v,a,s,d} N^1_{v,a,s,d}(x,t)\Gamma(s,v,a,d)\right]}{\sum_{i,v,a,s,d} N^i_{v,a,s,d}(x,t)} \qquad (6.3)$$

and this equation is then used to solve for $F(x,t)$, given the $\overline{F}(x,t)$ value. In the years that follow 2010, we have projected the HIV-negative fertility rates forward on the assumption of a steady decline in HIV-negative fertility, converging toward an ultimate set of fertility rates. These assumptions about declining future non-HIV fertility are the same as in the ASSA2008 'lite' model.

The assumed proportion of births that are male is 0.5039, again based on the ASSA2008 model.

6.3 Non-HIV mortality

The age-specific probabilities (q_x) of non-HIV/AIDS mortality for 1997-2010 were derived from the central mortality rates ($_nm_x$) for all-cause and HIV-specific mortality from the 2010 National Burden of Disease (NBD) study [348]. First m_0, $_4m_1$, $_5m_5$, … $_5m_{80}$, and m_{85+} were derived by subtracting the HIV/AIDS-specific rates from the all-cause rates. Next, because of the erratic nature of the rates at the older ages, the rates above age 65 were smoothed to follow the curve of the average rates by age over the period, scaled to the level of the rates in each year. Following this, Beers interpolation was applied to the rates from $_4m_1$ to m_{85+} to produce rates at individual ages from 2 to 87. These rates were then converted to probabilities of death for ages 2 to 80. Probabilities above age 80 were derived from extrapolated central mortality rates assuming that rates followed a Gompertz curve, increasing by 9% per year of age, to further reduce fluctuation over time and age. Finally q_0 was set equal to 1-exp(-0.983m_0) and q_1 was set equal to 1-exp(-3.985$_4m_1$)/[(1-q_2)(1-q_3)(1-q_4)], where m_0 and $_4m_1$ were the rates derived from the NBD estimates.

Rates for 2011 to 2015 were set as those projected using the CARe_3.2.xls model.

Probabilities of death for 1985 were set to those from the ASSA2008 model, and for 1986 to 1996, the probabilities of death were determined by linear interpolation between the estimates for 1985 and 1997. Beyond 2015, non-HIV/AIDS mortality rates are assumed to decline to a common set of ultimate rates at age-specific rates of decline. The ultimate rates and annual rates of decline are the same as assumed in the ASSA2008 models.

6.4 Migration

For each year from 1985 to 2015, we specify a number of net in-migrants (immigrants less emigrants) for each age and sex. The numbers of migrants were set in two stages. Initially the numbers of migrants by sex and single age for each year 1985 to 2000 were set equal those from the ASSA2008 models. The numbers for 2001-2010 were set as per those used to produce the alternative mid-year estimates [349]. These numbers were derived from the change in the numbers of people by place of birth (province or outside South Africa) between censuses, less an estimate of the number of South African-born emigrants as captured by censuses in the main countries of destination (UK, Australia, New Zealand, USA and Canada), scaled to match the total numbers recorded in the official mid-year estimates [350].

After this, these numbers were adjusted by an age-specific number (fixed over time) so that the projection of the population to the middle of 1996, 2001 and 2011 matched the census counts (approximately, for example, allowing for differences that might be expected due to errors in the census, such as undercounting of children or age exaggeration at the old ages). Although adjustment of migration of those born in the intercensal period was avoided as far as possible, there are probably some instances where the adjustment of migration compensated for errors in fertility. These adjustments were made at a provincial level, with national net immigration being the sum of the resultant provincial net in-migration.

Generally, these adjustments were determined by subtracting the projected numbers in five-year age groups before adjustment from the census count in these age groups. ${}_5M_{x\text{-}5}$ was set to $({}_5P^c_x - {}_5P_x)/10$, ${}_5M_0$ to $({}_5P^c_5 - {}_5P_5)/5$ and M_{85+} to ${}_5M_{80}$, where ${}_5P_x$ represents the number of people in the population aged between x and x+5, the superscript c represents the census count and ${}_5M_x$ represents the additional number of migrants aged between x and x+5 required for the adjustment. The age range requiring adjustment for each census was limited to that needed to correct for major deviations in one census from what would be expected given the other two, on the assumption that the estimates of migration reported by census questions are likely to be less accurate than the census counts.

The numbers at each age for 2011 to 2015 were set equal to those for 2010. Beyond 2015, the numbers at each age are assumed to trend asymptotically to zero at a rate of 4.5% per annum.

For each age, sex and year, we calculate a migration adjustment factor, which is one plus the number of net in-migrants divided by the number of individuals of the relevant age and sex at the end of the relevant projection year. This migration adjustment factor is applied multiplicatively to all sexual behaviour and HIV disease sub-strata within the relevant age-sex stratum. The implicit assumption that is made in applying this adjustment factor is that migrants (whether they are coming into South Africa or leaving South Africa) have the same sexual behaviour and HIV disease profile, on average, as the rest of the South African population. However, when the model is applied to each province, the multiplicative factors are calculated separately for the HIV-negative and HIV-positive populations, to reflect the effect of differences in HIV prevalence between provinces and the resulting differences in prevalence between migrants and their receiving populations (for more detail, see Appendix A of the provincial report [1]).

7. Statistical analysis

The model is calibrated to adult HIV prevalence data, ART and mortality data, using a Bayesian approach. The sections that follow describe the different steps in more detail. The model is also calibrated separately to paediatric HIV data, routine HIV testing data and key population prevalence data, as described in the appendices (see section 1 for an overview of the calibration process).

7.1 Prior distributions

The parameters that are allowed to vary in the calibration, and the corresponding prior distributions chosen to represent the uncertainty around these parameters, are summarized in Table 7.1. Most of these prior distributions have been referred to previously (see section references in last column), except in the case of the initial HIV prevalence in women in the high-risk group (this parameter 'seeds' the epidemic). Considering that the HIV prevalence in the first national antenatal clinic survey in 1990 was 0.76% and this grew by a multiple of 1.8 in each of the next two years [351], it is unlikely that HIV prevalence in women aged 15-49 in 1985 would have been more than 0.04% (0.0076×1.8^{-5}), since antenatal HIV prevalence tends to exceed prevalence in the general female population [352]. Since we assume that 25% of women are in the high-risk group, this suggests an upper limit of 0.16% on the initial HIV prevalence in the high risk group (0.0004/0.25). The initial HIV prevalence in 15-49 year old females in the high risk group has therefore been assigned a uniform (0, 0.002) prior. The initial ratio of male prevalence to female prevalence, as well as the initial age distribution of HIV, is set to be consistent with patterns of infection observed in the early stages of the epidemic in KwaZulu-Natal in 1991 [353].

Table 7.1: Prior distributions

	Prior distribution	Prior mean, std deviation	Ref.
Gamma density of relative rates of short-term partnership formation by age, in unmarried females			
Mean	Gamma (89.1, 2.30)	38.7, 4.1	2.2
Standard deviation	Gamma (47.5, 2.46)	19.3, 2.8	2.2
Sexual mixing parameter	Beta (11.50, 12.46)	0.53, 0.12	2.5
Bias in reported condom use at last sex	Uniform (0, 1)	0.50, 0.29	2.8
Reduction in unprotected sex after HIV diagnosis	Beta (27.9, 21.1)	0.18, 0.13	2.10
Average survival in absence of ART (years)	Gamma (144, 12)	12, 1	3.1
RR of HIV disease progression in women	Gamma (369, 384)	0.96, 0.05	3.1
Increase in HIV disease progression per 10-year increase in age	Gamma (9, 50)	0.18, 0.06	3.1
Reduction in mortality* per unit increase in rate of ART initiation (at CD4<200) over last 3 years	Gamma (4.59, 0.612)	7.5, 3.5	3.4
Female-to-male transmission probability in short-term/ non-spousal partnerships	Beta (7.05, 874)	0.008, 0.003	4.1
Male-to-female transmission probability in short-term/ non-spousal partnerships	Beta (5.68, 468)	0.012, 0.005	4.1
Odds of viral suppression relative to that in IeDEA-SA	Gamma (84.8, 100.6)	0.844, 0.092	4.3
Initial HIV prevalence in high-risk women, ages 15-49	Uniform (0, 0.002)	0.001, 0.00058	7.1

* On a natural log scale. IeDEA-SA = International epidemiology Databases to Evaluate AIDS, Southern Africa. RR = relative rate.

7. 2 Likelihood definition

The model is calibrated to two HIV prevalence data sources: antenatal clinic survey data and household survey data. In addition, the model is calibrated to recorded death data and antiretroviral metabolite data. The likelihood for all four data sources is simply the product of the likelihood calculated for each individual data source, as detailed below.

7.2.1 Likelihood definition for antenatal clinic survey data

The model is fitted to antenatal HIV prevalence data from national surveys that have been conducted from 1997 to 2015 (survey data collected prior to 1997 have not been included, as these early antenatal surveys were based on convenience samples and reported 95% confidence intervals did not include survey design effects). Although no antenatal survey was conducted in 2016, the survey was repeated in 2017, and these data have also been included in the model calibration. We include HIV prevalence estimates for 5 age groups (15-19, 20-24, 25-29, 30-34 and 35-39).

Suppose that $H_{x,t}(\boldsymbol{\varphi})$ is the model estimate of HIV prevalence in pregnant women aged x to $x + 4$, in year t, where the vector $\boldsymbol{\varphi}$ represents the values of the model input parameters. This is calculated from equation (6.3) as

$$H_{x,t}(\boldsymbol{\varphi}) = 1 - \frac{\sum_{j=x}^{x+4} F(j,t) \sum_{v} N^{1}_{v,0,0,0}(j,t)}{\sum_{j=x}^{x+4} \overline{F}(j,t) \sum_{i,v,a,s,d} N^{i}_{v,a,s,d}(j,t)}. \qquad (7.1)$$

This model estimate of HIV prevalence in pregnant women is multiplied by an adjustment factor for calibration purposes, to account for the strong association between recent HIV acquisition and incidence of pregnancy, since both are related to recent unprotected sex. This adjustment factor is defined as $\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}$, where $\lambda_{x,t}(\boldsymbol{\varphi})$ is the model estimate of the fraction of HIV infections in women aged x to $x + 4$ in year t that have been acquired in the last 12 months, and θ is a scaling factor that controls the strength of the association between incident pregnancy and incident HIV (when there is no association, the implicit assumption in previous versions of Thembisa, θ is 0).

The corresponding prevalence of HIV actually measured in the antenatal survey is represented by $y_{x,t}$. It is assumed that if $\boldsymbol{\varphi}$ is the true set of parameter values, then the difference between the logit-transformed model estimate and the logit-transformed observed prevalence is normally distributed. The mean of this normal distribution represents the extent of antenatal bias, which arises due to the exclusion of women receiving private antenatal care from the sample and other behavioural factors. The variance of the distribution is assumed to be composed of a ‘survey error’ term (representing the uncertainty around the survey estimate due to binomial variation and cluster variation in the survey) and a ‘model error’

term (representing the error that may arise due to the assumption that the antenatal bias is constant over time and constant with respect to age). More formally, it is assumed that

$$\log\left(\frac{y_{x,t}}{1-y_{x,t}}\right)=\log\left(\frac{H_{x,t}(\boldsymbol{\varphi})\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}}{1-H_{x,t}(\boldsymbol{\varphi})\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}}\right)+b+m_{x,t}+\varepsilon_{x,t}, \qquad (7.2)$$

where b is the antenatal bias parameter, $m_{x,t} \sim N\left(0,\sigma_m^2\right)$ and $\varepsilon_{x,t} \sim N\left(0,\sigma_{x,t}^2\right)$. The latter two terms represent the model error and the survey error respectively. The logit transformations ensure that the error terms are closer to normality and that the model error terms are roughly independent of the level of HIV prevalence. The assumption that the antenatal bias b is the same across all ages and years is not realistic, but we have found that including the $\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}$ term substantially reduces the extent of the variation in bias across age groups and over time, with the variation being minimized when θ is set to 0.04 (manuscript under review). We have therefore set θ to 0.04 for the purpose of this analysis. The model error term can be thought of as representing the residual error that remains due to the assumption of a constant antenatal bias, after including the $\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}$ adjustment.

For a given parameter combination $\boldsymbol{\varphi}$, the antenatal bias parameter is estimated using the formula

$$\hat{b}=\frac{1}{100}\sum_x\sum_t\left(\log\left(\frac{y_{x,t}}{1-y_{x,t}}\right)-\log\left(\frac{H_{x,t}(\boldsymbol{\varphi})\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}}{1-H_{x,t}(\boldsymbol{\varphi})\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}}\right)\right). \qquad (7.3)$$

(The 100 in the denominator corresponds to the total number of observations, i.e. 5 age-specific HIV prevalence measures for each of 20 surveys.) The $\sigma_{x,t}^2$ values are estimated from the 95% confidence intervals that have been published for the various survey estimates. Once these $\sigma_{x,t}^2$ values have been obtained, the σ_m^2 parameter is estimated using the formula

$$\hat{\sigma}_m^2=\frac{1}{100}\sum_x\sum_t\left(\log\left(\frac{y_{x,t}}{1-y_{x,t}}\right)-\log\left(\frac{H_{x,t}(\boldsymbol{\varphi})\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}}{1-H_{x,t}(\boldsymbol{\varphi})\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}}\right)-\hat{b}\right)^2-\sigma_{x,t}^2. \qquad (7.4)$$

The likelihood in respect of the antenatal data is then calculated based on the assumption that the error terms are normally distributed:

$$L(\mathbf{y}\mid\boldsymbol{\varphi})=\prod_x\prod_t\frac{\exp\left(-\frac{\left(\text{logit}\left(y_{x,t}\right)-\text{logit}\left(H_{x,t}(\boldsymbol{\varphi})\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}\right)-\hat{b}\right)^2}{2\left(\hat{\sigma}_m^2+\sigma_{x,t}^2\right)}\right)}{\sqrt{2\pi\left(\hat{\sigma}_m^2+\sigma_{x,t}^2\right)}}, \qquad (7.5)$$

where $\mathbf{y}$ represents the matrix of $y_{x,t}$ values, across age bands 15-19 to 35-39, and across calendar years 1997-2015 and 2017.

7.2.2 Likelihood definition for household survey prevalence data

The model is calibrated to HIV prevalence data from four nationally-representative household surveys conducted by the Human Sciences Research Council (HSRC) in 2005 [18], 2008 [19], 2012 [4] and 2017 [5], as well as HIV prevalence data from the 2016 Demographic and Health Survey (DHS) [126]. HIV prevalence levels in each survey are estimated by 5-year age group (from 15-19 up to 55-59) and by sex. The approach adopted in defining the likelihood function in respect of the HSRC and DHS HIV prevalence data is the same as that for the antenatal data, except that the bias term (b) and model error term (m) are both omitted, and there is no $\lambda_{x,t}(\boldsymbol{\varphi})^{\theta}$ adjustment factor. The omission of the bias term is consistent with the approach adopted in other uncertainty analyses of HIV data in developing countries [354, 355], in which it is assumed that household prevalence data provide an unbiased estimate of HIV prevalence in the general population. The model error term is omitted because it is not necessary if the survey estimates are truly unbiased, and because (as noted previously) the primary purpose of the model error term is to represent the error introduced by the assumption of a constant antenatal bias (which is not relevant when considering data from household surveys). Even if the household survey estimates were biased, the 95% confidence intervals around the household prevalence estimates are very wide, relative to the confidence intervals around the antenatal survey estimates, and the model error would therefore be small relative to the survey error.

7.2.3 Likelihood definition for antiretroviral metabolite data

We calculate a likelihood to represent the goodness of model fit to 2012 and 2017 household survey estimates of the proportions of HIV-positive adults who are on ART. In both household surveys, the proportion of HIV-positive adults on ART was estimated based on testing for antiretroviral metabolites (efavirenz, nevirapine, lopinavir and other less commonly used drugs, i.e. accounting for most first- and second-line ART regimens) [4, 5]. Although the survey also collected self-reported data on ART use, we have not used these data in calibration, in part to be consistent with the methods used in the HSRC survey reports, and in part because there were high levels of non-response to questions about ART use (Jeffrey Eaton, personal communication). Estimates of coverage were also obtained separately for men and women, so that the model was calibrated to a total of 4 data points (Table 7.2). We calculated the likelihood on the assumption that the difference between the survey estimate of ART coverage and the modelled ART coverage, on a logit scale, was normally distributed with zero mean and variance calculated from the 95% confidence interval around the survey estimate.

Table 7.2: Survey estimates of the proportion of HIV-positive adults (15+) on ART

Sex	2012	2017
Male	24.2% (19.3-29.2%)	58.0% (53.7-62.3%)
Female	34.2% (30.8-37.6%)	66.7% (64.2-69.1%)

A number of other sources provide data on ART coverage based on self-reported receipt of ART; for example, the DHIS provides data on the proportion of HIV-positive pregnant women who report having started ART prior to their current pregnancy. We have not included these data in the model calibration, as we do not consider self-reporting of ART

status to be reliable. For example, in the 2017 national antenatal survey, viral load testing was conducted in all HIV-positive mothers, and viral loads of less than 1000 RNA copies/ml were detected in 39% of women who reported not being on ART [356]; such a high rate of viral suppression does not seem plausible in HIV-positive women who are truly untreated. Several other South African studies have found a substantial prevalence of detectable antiretroviral metabolites in HIV-positive individuals who report being undiagnosed [124, 357-359], although one South African study found only minimal disagreement between self-reported ART coverage and the ART coverage based on self-report [360].

7.2.4 Likelihood definition for recorded death data

To calculate the likelihood in respect of the reported death data, we restrict this analysis to deaths occurring over the period from the start of 1997 to the end of 2016 [361]. Because cause of death information is seldom captured accurately, and reported AIDS deaths are likely to be only a fraction of the actual HIV-related deaths [362], we compare model estimates of all-cause mortality with reported levels of all-cause mortality. This comparison is only likely to be meaningful in those age groups in which a substantial proportion of deaths are HIV-related, and this analysis is therefore restricted to deaths occurring from ages 20 to 59. Mortality data are grouped in 5-year age bands for calibration purposes, and estimates are considered separately for males and females.

Suppose that $\Theta_{g,x,t}(\boldsymbol{\varphi})$ represents the model estimate of the number of deaths in individuals of sex *g*, between ages *x* and *x* + 4, in year *t*, where the vector $\boldsymbol{\varphi}$ represents the values of the model input parameters. Further suppose that $R_{g,x,t}$ represents the reported number of deaths in individuals of sex *g*, between ages *x* and *x* + 4, in year *t*. In order to specify a likelihood function for the reported death data, it must be assumed that a certain proportion of adult deaths, $\gamma_{g,x,t}$, is reported. It is assumed that if $\boldsymbol{\varphi}$ is the true set of parameter values, then the difference between the log-transformed model estimate of the number of reported deaths ($\Theta_{g,x,t}(\boldsymbol{\varphi})\gamma_{g,x,t}$) and the log-transformed actual number of reported deaths is normally distributed with zero mean. More formally, the likelihood is calculated on the assumption that

$$\log\left(R_{g,x,t}\right) = \log\left(\Theta_{g,x,t}(\boldsymbol{\varphi})\gamma_{g,x,t}\right) + \varepsilon_{g,x,t}, \qquad (7.6)$$

where $\varepsilon_{g,x,t} \sim N\left(0, \sigma_d^2\right)$. The parameter $\varepsilon_{g,x,t}$ can be regarded as comprising both a 'model error' and 'random binomial error' component, but because the population numbers are very large, the random binomial component of the error is relatively small on the log scale. It is therefore reasonable to assume that the variance of the error term is independent of the population size in the relevant sex and age group.

The $\gamma_{g,x,t}$ parameters have been estimated from a variety of sources. Over the period from October 1996 to October 2001, Dorrington *et al* [363] estimate that the fraction of adult deaths recorded was 84%, based on death distribution methods (i.e. based on comparing the recorded numbers of adult deaths to the changes in the population sizes in each age cohort over the inter-census period). The authors also estimate that the annual increase in the proportion of deaths recorded, over this 5-year period, was 1.7% in men and 2.1% in women,

based on an assumption of stable mortality rates at ages 65 and older (where AIDS would be expected to have relatively little impact on mortality). In the period after 2001, estimates of the completeness of adult death recording have been around 93%, based on similar methods [364-366]. Based on these estimates, we set initial completeness assumptions – independent of age and sex – that increase linearly from 80.2% in 1997 to 87.8% in 2001 (an increase of 1.9% per annum, with 84% completeness in 1999) and 93% in 2004, after which completeness is assumed to remain constant (Table 7.3). The assumption of constant completeness after 2004 is supported by an analysis of factors affecting the recording of deaths in ART patients, which showed no significant change in the completeness of vital registration over the 2004-2014 period [367].

In the final set of completeness assumptions, we use the completeness estimates by age and sex, as estimated in the analysis of factors affecting the recording of deaths in ART patients over the 2004-2014 period [367], and scale these down by the ratio of initial completeness assumptions to 0.93 in the period prior to 2004. The completeness assumptions are shown in Table 7.3.

Table 7.3: Completeness assumptions (fraction of deaths that are recorded)

Year	1997	1998	1999	2000	2001	2002	2003	2004+
Initial completeness assumptions								
	0.802	0.821	0.84	0.859	0.878	0.897	0.914	0.930
Final completeness assumptions								
Women aged								
20-24	0.798	0.817	0.836	0.855	0.874	0.892	0.909	0.925
25-29	0.809	0.828	0.847	0.866	0.886	0.905	0.922	0.938
30-34	0.817	0.836	0.855	0.875	0.894	0.913	0.931	0.947
35-39	0.823	0.842	0.862	0.881	0.901	0.920	0.937	0.954
40-44	0.827	0.847	0.866	0.886	0.905	0.925	0.943	0.959
45-49	0.831	0.850	0.870	0.890	0.909	0.929	0.947	0.963
50-54	0.834	0.853	0.873	0.893	0.913	0.932	0.950	0.967
55-59	0.836	0.856	0.876	0.895	0.915	0.935	0.953	0.969
Men aged								
20-24	0.756	0.774	0.792	0.810	0.828	0.846	0.862	0.877
25-29	0.772	0.791	0.809	0.827	0.845	0.864	0.880	0.896
30-34	0.789	0.807	0.826	0.845	0.863	0.882	0.899	0.914
35-39	0.802	0.821	0.840	0.859	0.878	0.897	0.914	0.930
40-44	0.813	0.832	0.852	0.871	0.890	0.909	0.927	0.943
45-49	0.821	0.841	0.860	0.880	0.899	0.918	0.936	0.952
50-54	0.827	0.847	0.866	0.886	0.906	0.925	0.943	0.959
55-59	0.832	0.851	0.871	0.891	0.910	0.930	0.948	0.964

The maximum likelihood estimate of the parameter σ_d^2 is calculated as

$$\hat{\sigma}_d^2 = \frac{1}{320}\sum_g \sum_x \sum_{t=1997}^{2016} \left[\log\left(R_{g,x,t}\right) - \log\left(\Theta_{g,x,t}(\boldsymbol{\varphi})\gamma_{g,x,t}\right)\right]^2, \qquad (7.7)$$

where 320 is the total number of observations (20 years, 8 age groups, for men and women separately). The likelihood in respect of the reported death data is then calculated based on the assumed normality of the error terms:

$$L(\mathbf{R} \mid \boldsymbol{\varphi}) = \prod_{g} \prod_{x} \prod_{t=1997}^{2016} \left(2\pi\hat{\sigma}_d^2\right)^{-0.5} \exp\left(- \frac{\left(\log\left(R_{g,x,t}\right) - \log\left(\Theta_{g,x,t}(\boldsymbol{\varphi})\gamma_{g,x,t}\right)\right)^2}{2\hat{\sigma}_d^2} \right), \quad (7.8)$$

where **R** represents the matrix of reported death data.

7.3 Posterior simulation

The posterior distribution was simulated numerically using Incremental Mixture Importance Sampling (IMIS) [368]. Following the recommendations of Raftery and Bao [368], an initial set of 10 000 parameter combinations was randomly drawn from the prior distributions in Table 7.1 and the likelihood was calculated for each. Importance sampling was then used to draw a second sample of 1 000 parameter combinations from the region of the parameter space with the highest likelihood values, and the procedure was repeated iteratively, updating the importance sampling distribution at each step to reflect the region of the parameter space with the highest likelihood values, until the algorithm converged on a posterior sample that was sufficiently heterogeneous. A posterior sample of 1 000 parameter combinations was drawn, and means and 95% confidence intervals were calculated from this sample.

7.4 Combining uncertainty from different analyses

In the presentations of the posterior model estimates in sections 8.1-8.3, posterior means and confidence intervals represent only the uncertainty in the 13 parameters listed in Table 7.1. However, for the purpose of representing uncertainty in other sections, and for the purpose of representing uncertainty in the Thembisa model estimates on the Thembisa website, we extend these ranges to reflect the sources of uncertainty described in Appendix B. For example, the 95% confidence interval around the published estimate of proportion of HIV-positive adults who are diagnosed should reflect not only the uncertainty about adult HIV transmission probabilities and HIV disease progression rates, but also the uncertainty around rates of HIV diagnosis by age and sex.

To achieve this, the 1000 parameter combinations obtained in section 7.3 are combined with the 1000 sets of parameter combinations generated in Appendices B to generate 1000 distinct combinations of 24 parameter values (13 corresponding to parameters in Table 7.1, and 11 in Table B7). For each combination of 24 parameters, the model is run. After the 1000 sets of results have been generated, means, 2.5 and 97.5 percentiles are calculated from the distributions of model results.

8. Results of model calibration

8.1 Comparison of prior and posterior distributions

Table 8.1 compares the prior and posterior means for the 13 parameters that are allowed to vary when fitting the model to the adult HIV data. For most of these parameters, the prior and posterior distributions overlap substantially, though the posterior 95% confidence intervals are substantially narrower, reflecting the increased precision due to the HIV prevalence data and mortality data. However, the extent of the reduction in unprotected sex after HIV diagnosis in the posterior analysis (a 56.4% reduction) is substantially greater than the prior mean of 18%. Consistent with previous analyses based on the STI-HIV Interaction model and the ASSA AIDS and Demographic model [15], we find that it is not possible to fit the model to observed HIV prevalence trends unless we assume substantial bias in the reporting of condom use at last sex (posterior mean of 0.89 for the condom bias parameter).

Table 8.1: Comparison of prior and posterior distributions

	Prior distribution (mean, 95% CI)	Posterior distribution (mean, 95% CI)
Gamma density of relative rates of short-term partnership formation by age, in unmarried females		
Mean	38.7 (31.1-47.1)	33.7 (33.4-34.3)
Standard deviation	19.3 (14.2-25.2)	17.8 (17.4-18.2)
Sexual mixing parameter	0.530 (0.295-0.758)	0.726 (0.689-0.765)
Bias in reported condom use at last sex	0.500 (0.025-0.975)	0.890 (0.842-0.918)
Reduction in unprotected sex after HIV diagnosis	0.180 (0.013-0.496)	0.564 (0.496-0.621)
Average survival in absence of ART (years)	12.00 (10.12-14.04)	12.26 (12.02-12.45)
RR of HIV disease progression in women	0.960 (0.864-1.060)	0.893 (0.880-0.904)
Increase in HIV disease progression per 10-year increase in age	0.180 (0.082-0.315)	0.134 (0.125-0.145)
Reduction in mortality* per unit increase in rate of ART initiation (at CD4<200) over last 3 years	7.50 (2.29-15.76)	5.83 (5.01-6.62)
Female-to-male transmission probability in short-term/ non-spousal partnerships	0.0080 (0.0032-0.0149)	0.0062 (0.0060-0.0064)
Male-to-female transmission probability in short-term/ non-spousal partnerships	0.0120 (0.0043-0.0236)	0.0200 (0.0193-0.0208)
Odds of viral suppression relative to that in IeDEA-SA	0.844 (0.674-1.033)	0.867 (0.835-0.903)
Initial HIV prevalence in high risk women, ages 15-49	0.100% (0.005-0.195%)	0.151% (0.142-0.159%)

* On a natural log scale.

8.2 Calibration to adult HIV prevalence data

Figure 8.1 shows the calibration of the model to the antenatal survey HIV prevalence data (although the data from the 1990-1996 surveys were not included in the likelihood definition, they are included here as a validation of the model). The posterior mean model estimates of antenatal HIV prevalence are generally consistent with the survey data, although the model slightly over-estimates HIV prevalence in pregnant women aged 20-24, particularly in the earlier stages of the HIV epidemic. The model also appears to slightly under-estimate HIV prevalence in pregnant women aged 30-34 over the 2004-2010 period.

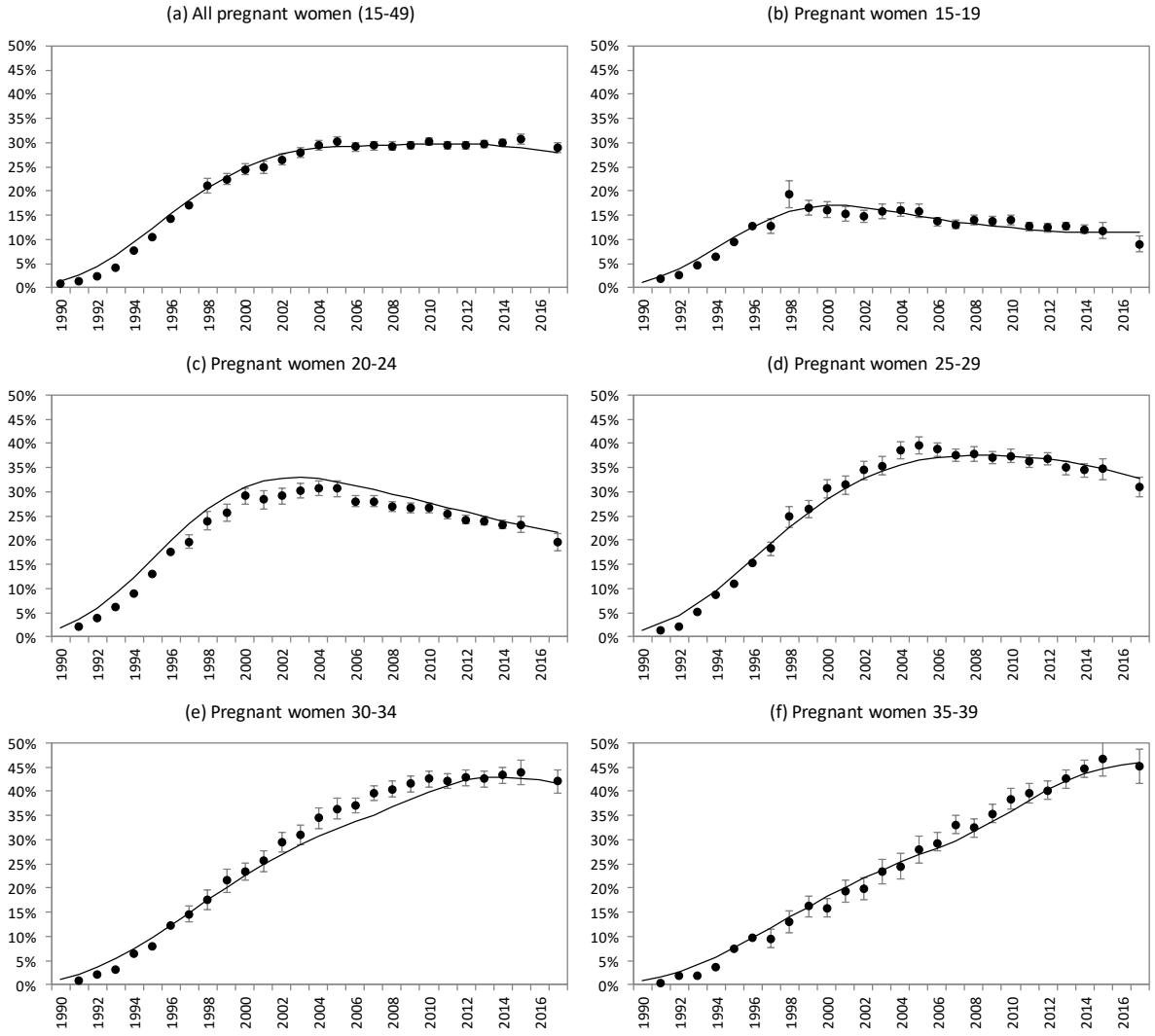


Figure 8.1: HIV prevalence levels in pregnant women attending public antenatal clinics
Dots represent HIV prevalence levels reported in surveys conducted from 1990-2015 and 2017 (the 1998 data were adjusted to correct an error in the provincial weights in that year). Solid lines represent the posterior mean model estimates of HIV prevalence in pregnant women, after adjusting for antenatal bias. Survey data in the pre-1997 period are included in the graphs even though they were not used in defining the likelihood function.

The model provides a reasonably good fit to the HSRC prevalence survey data (Figure 8.2). The model fits the 2017 survey data reasonably well, although the model tends to slightly under-estimate the levels of HIV prevalence in men.

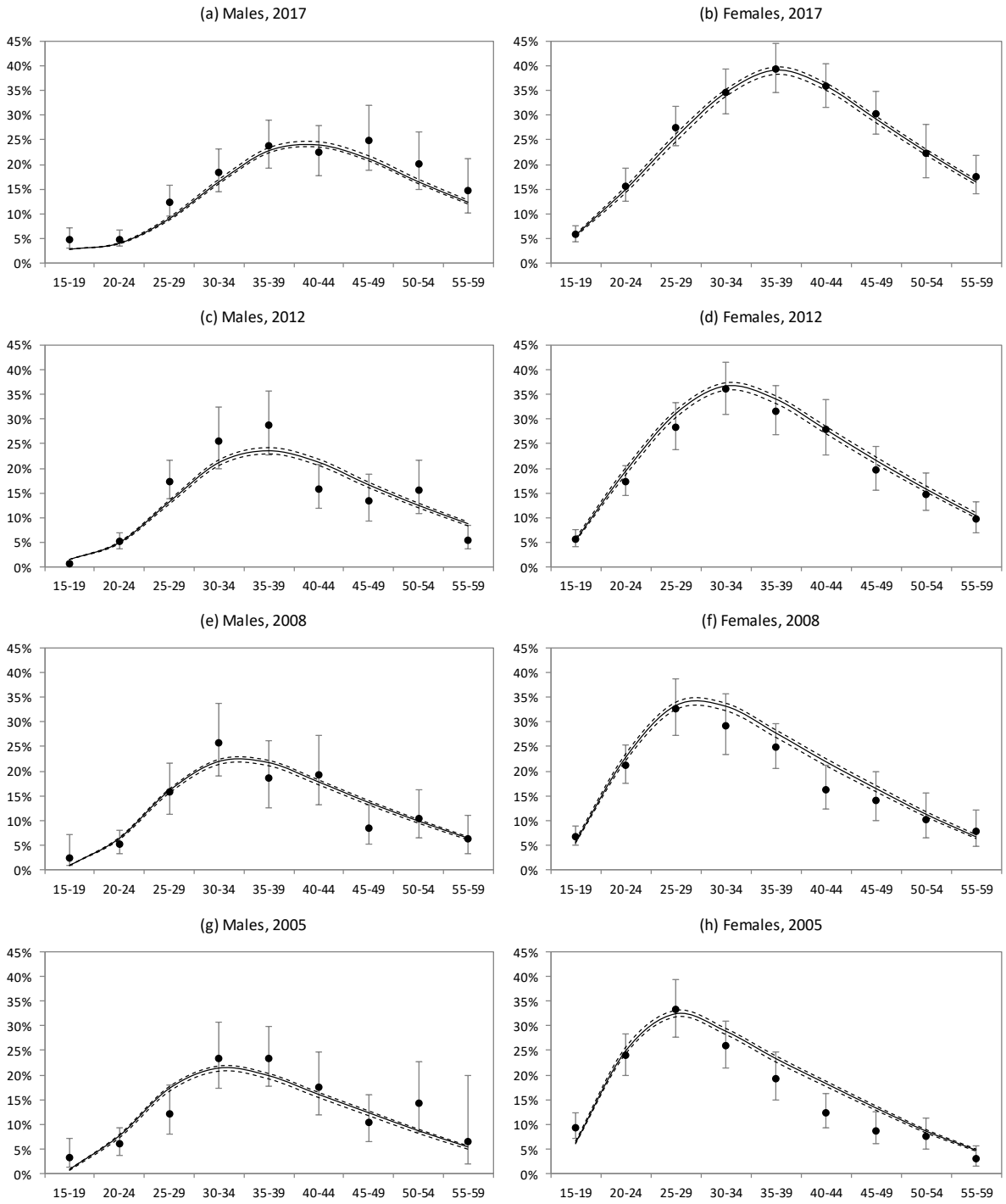


Figure 8.2: HIV prevalence levels in the general population

Dots represent HSRC survey prevalence estimates, together with 95% confidence intervals. Solid lines represent the posterior mean model estimates of HIV prevalence, and dashed lines represent the 95% confidence intervals around these estimates.

Figure 8.3 shows the model fit to the 2016 DHS data. Although the model estimates of HIV prevalence appear roughly consistent with the HIV prevalence data, the survey estimates have relatively wide confidence intervals around them, reflecting relatively small sample sizes.

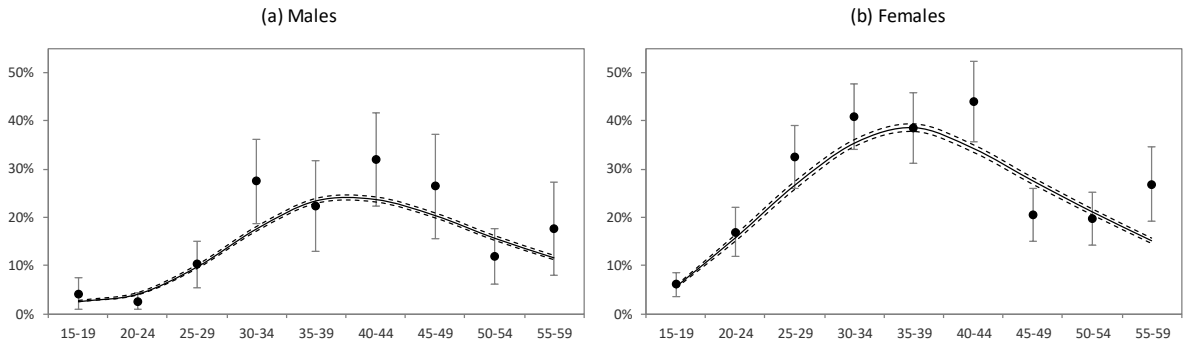


Figure 8.3: HIV prevalence levels in the general population in 2016
Dots represent DHS prevalence estimates, together with 95% confidence intervals. Solid lines represent the posterior mean model estimates of HIV prevalence, and dashed lines represent the 95% confidence intervals around these estimates.

Figure 8.4 shows that the model also matches the household survey prevalence trends over time, although the most recent HSRC and DHS survey results suggest a more substantial increase in HIV prevalence in recent years than the model suggests. A possible explanation for the unexpectedly high prevalence in the two most recent surveys is differential non-response by socio-economic status: in the 2016 DHS, for example, the fraction of household members who were interviewed and tested for HIV ranged between 39% in the highest wealth quintile and 69% in the lowest wealth quintile, and HIV prevalence was inversely associated with wealth quintile [126]. To the extent that this under-sampling of higher socio-economic status groups is not corrected for by the survey weights, HIV prevalence could be over-stated.

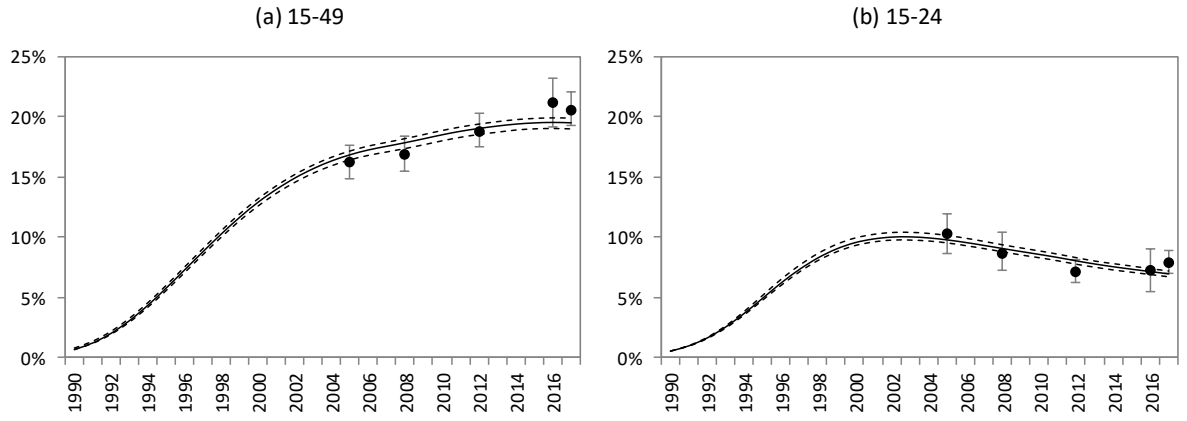


Figure 8.4: HIV prevalence trends in the general population
Dots represent HSRC and DHS survey prevalence estimates, together with 95% confidence intervals. Solid lines represent the posterior mean model estimates of HIV prevalence, and dashed lines represent the 95% confidence intervals around these estimates.

The model has also been calibrated to HIV prevalence data from HIV testing services (Appendix B), surveys of sex workers and MSM (Appendix C), and household survey estimates of prevalence in children (Appendix E). The model matches levels of HIV prevalence observed in individuals attending HIV testing services (Figure B2). The model also matches the key population HIV prevalence data roughly, but because none of the survey

estimates are nationally representative, there is substantial variation in survey estimates of prevalence, and consequently wide confidence intervals around the model estimates of prevalence in key populations (Figure C1). Model estimates of HIV prevalence in children are also consistent with survey data [6].

8.3 Calibration to antiretroviral metabolite data

Figure 8.5 shows the model fit to the survey estimates of the fraction of HIV-positive adults on ART (i.e. having detectable antiretroviral metabolites). The model appears to significantly under-estimate the ART coverage in 2017, suggesting that either (a) the model is under-estimating the number of patients on ART, (b) the model is over-estimating HIV prevalence, or (c) there is a non-response bias (with HIV-positive people who are untreated being less likely to agree to HIV testing than those on ART). Further investigation is required to identify the most likely explanation, although explanation (b) seems unlikely, given that any reduction in the model estimate of HIV prevalence in recent years would yield a substantially poorer fit to the 2016-17 HIV prevalence data (Figure 8.4a). Explanation (c) would be inconsistent with the finding that the model slightly *over*-estimates ART coverage in 2012 (particularly in men).

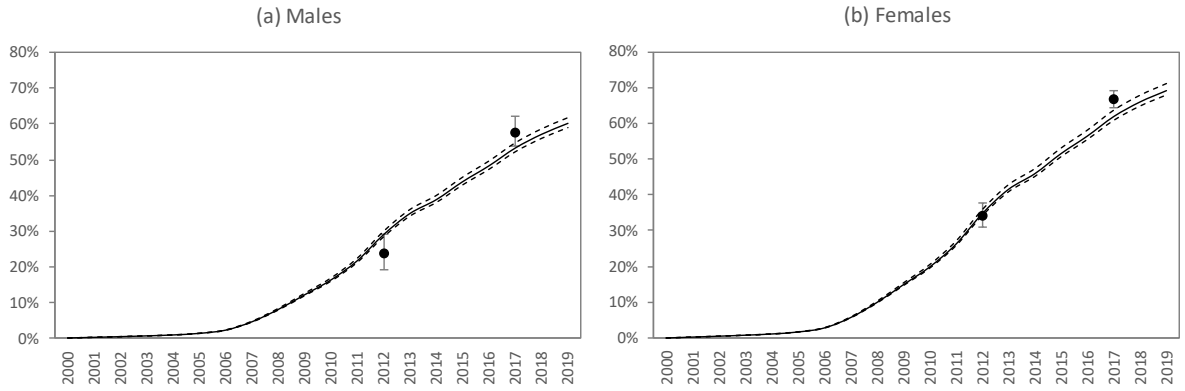


Figure 8.5: ART coverage in the adult population (15 years and older)
Dots represent HSRC survey prevalence estimates, together with 95% confidence intervals. Solid lines represent the posterior mean model estimates of ART coverage, and dashed lines represent the 95% confidence intervals around these estimates.

8.4 Calibration to adult mortality data

Figure 8.6 compares the model estimates of deaths over the 20-59 age range with the corresponding recorded numbers of deaths (after adjusting the latter for incomplete vital registration). Model estimates are in good agreement with the data in most years, although the model slightly over-estimates the numbers of recorded deaths in men in the most recent year for which data are available (2016), possibly as a result of late death registrations. Similar patterns are observed when age-specific comparisons are performed (Figure 8.7). The model does not fit the recorded numbers of male deaths in the 50-59 age group well, which could be an indication of problems with the non-HIV mortality assumptions.

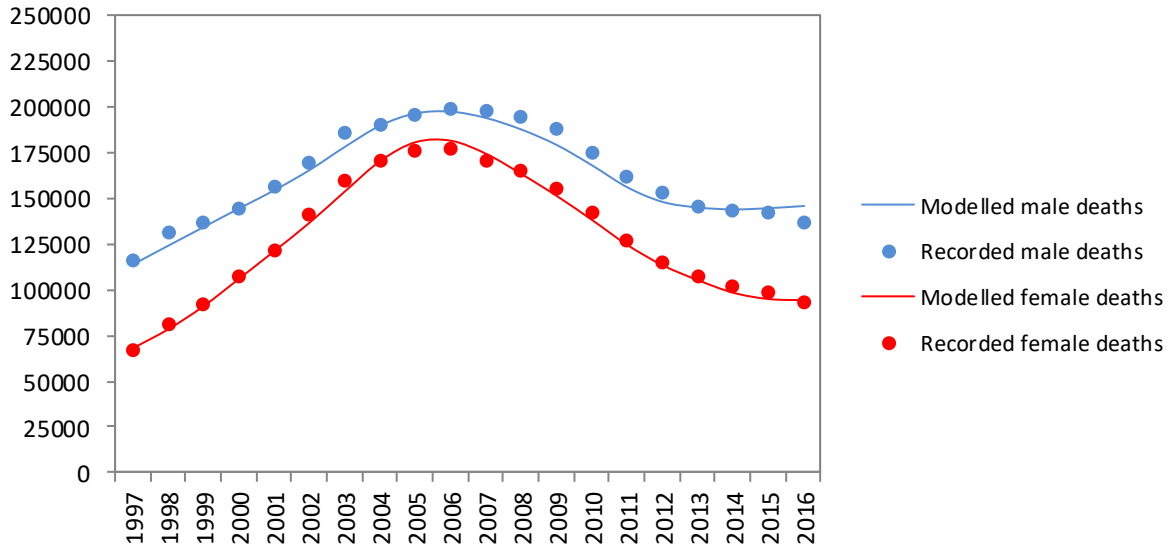


Figure 8.6: Numbers of deaths in adults aged 20-59

Dots represent recorded numbers of deaths, after adjusting for incomplete registration. Solid lines represent the posterior mean model estimates.

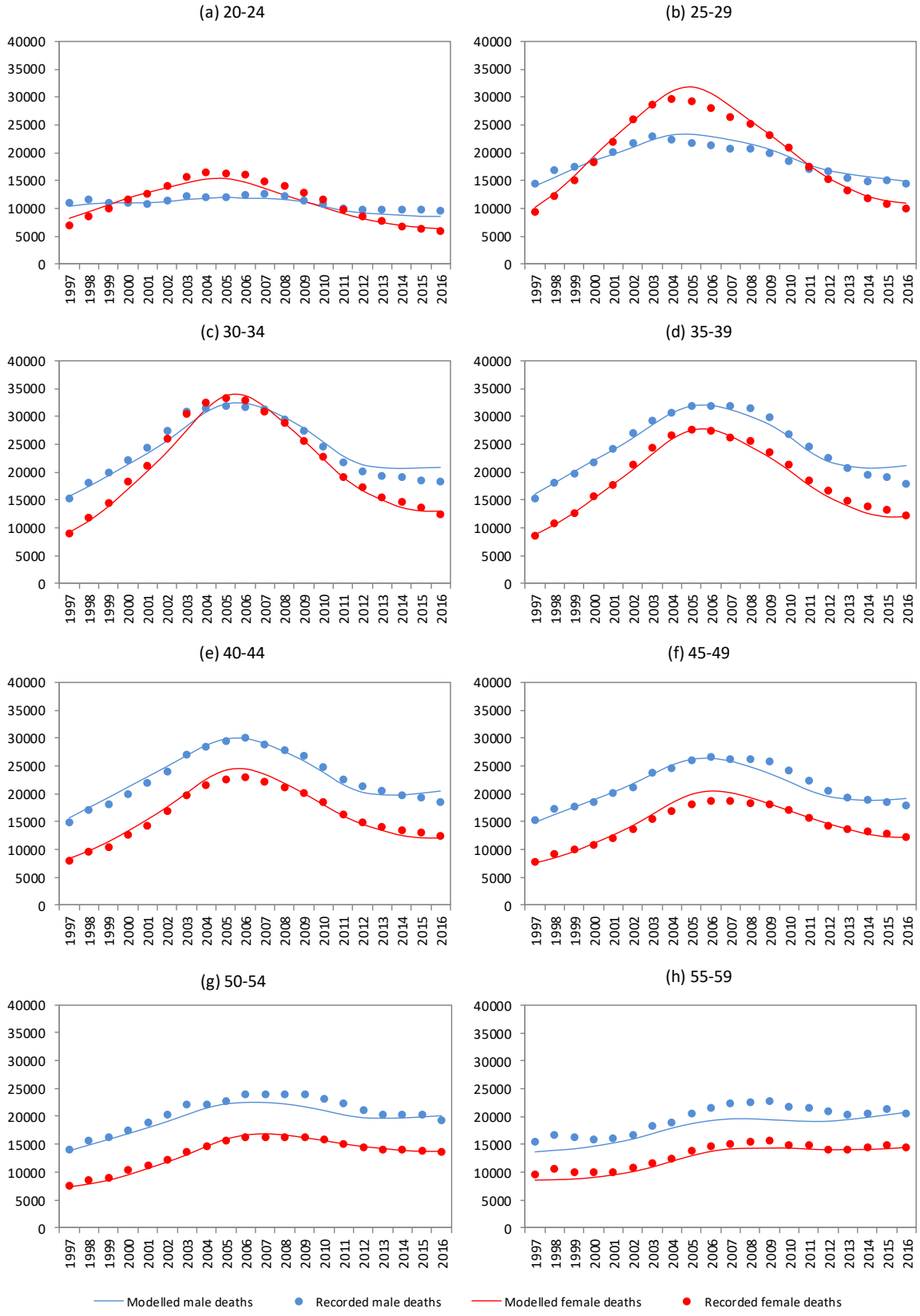


Figure 8.7: Numbers of deaths in adults, by five-year age group

Dots represent recorded numbers of deaths, after adjusting for incomplete registration. Solid lines represent the posterior mean model estimates.

8.5 Validation against HIV incidence estimates

Figure 8.8 compares the model estimates of HIV incidence in 15-49 year olds with estimates published by the HSRC [5, 369]. The first three estimates were obtained using a synthetic cohort approach, and model estimates are roughly consistent with these estimates, although the confidence intervals around these estimates are very wide. The two most recent estimates were obtained using a multi-assay testing algorithm based on the LAg avidity assay. Overall the model is very consistent with the empirical estimates of HIV incidence, even though these were not included in the model calibration.

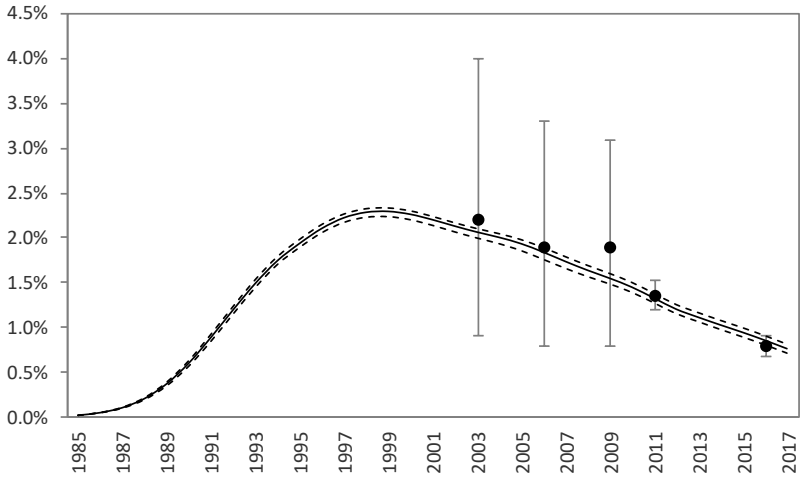


Figure 8.8: HIV incidence in 15-49 age group
Dots represent estimates derived directly from HSRC survey data (as Thembisa projection years run from mid-year to mid-year, the results from the 2012 and 2017 surveys are presented in the 2011-12 and 2016-17 projection years respectively). Solid lines represent the posterior mean model estimates and dashed lines represent 95% confidence intervals.

Figure 8.9 compares the model estimates of mother-to-child transmission rates from routine infant testing data and surveys with corresponding model estimates. Model estimates of perinatal mother-to-child transmission rates are reasonably close to routine data sources (Figure 8.9(a)), which include the District Health Information System (DHIS) [370] and the National Health Laboratory Service (NHLS) [325]. However, these estimates of perinatal transmission are under-estimates of the total perinatal transmission because they do not reflect transmission from mothers who are undiagnosed. There is a lack of data on postnatal transmission rates, although the SAPMTCTE study, which followed mothers who were diagnosed either antenatally or at their 6-week immunization visit, found that cumulative transmission (perinatal and postnatal) up to 18 months was 4.3% (95% CI: 3.8-5.0%) [371]. Our model estimates are consistent with this survey (Figure 8.9(b)), although the definition of postnatal transmission considered here is an under-estimate of all postnatal transmission, since some transmission occurs after 18 months, and substantial transmission occurs from mothers who are undiagnosed. The reported transmission rates have not been adjusted for possible false-positive PCR results, which may lead to reported transmission rates exaggerating true transmission rates [372]. Reported rates have also not been adjusted for likely false-negative PCR results in PMTCT-exposed infants [334], and lower rates of

screening in mothers who have not received ART antenatally [331, 333, 373], both of which may lead to reported transmission rates under-estimating true transmission rates.

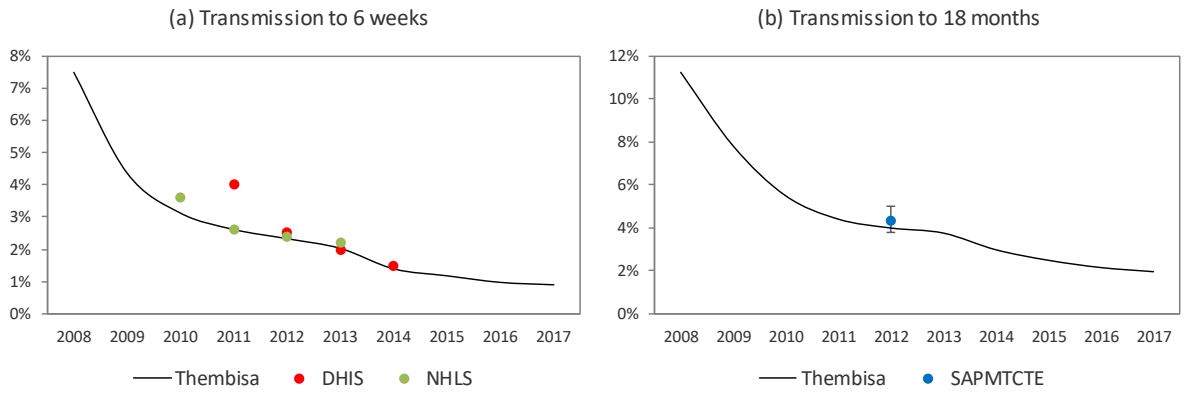


Figure 8.9: Mother-to-child transmission rates

Dots represent survey prevalence estimates. Solid lines represent the posterior mean model estimates of HIV prevalence. In panel (a), the denominator is all HIV-positive women who were diagnosed antenatally (excluding mothers who were not diagnosed), and in panel (b) the denominator is the number of HIV-positive mothers who were diagnosed either antenatally or at their 6-week immunization visit (again excluding mothers who were not diagnosed). Panel (a) represents only perinatal transmission, while panel (b) represents combined perinatal and postnatal transmission.

8.6 Validation against reported ART data

Figure 8.10 compares the model estimates of numbers of ART patients (after adjusting to exclude the independently-estimated numbers receiving ART in the private sector) with the reported numbers of patients receiving ART in the public sector. In the period up to 2009, when public sector statistics reflected mostly cumulative enrolment, model estimates of cumulative enrolment appear consistent with the reported cumulative enrolment. Thereafter public sector statistics were reported as 'current enrolment', although it is suspected that many clinics were still reporting cumulative enrolment. From 2012 onward, model estimates of current enrolment are more consistent with the reported 'current enrolment' data than the model estimates of cumulative enrolment, reflecting the gradual transition to reporting of current enrolment that started in late 2009. The only estimate of cumulative enrolment in recent years is the estimate of 4.2 million in March 2015, based on viral load data from the NHLS [374]. The model estimate of cumulative enrolment is roughly consistent with this estimate, suggesting that the model assumptions about rates of ART interruption and ART resumption (described in Appendix G) are plausible.

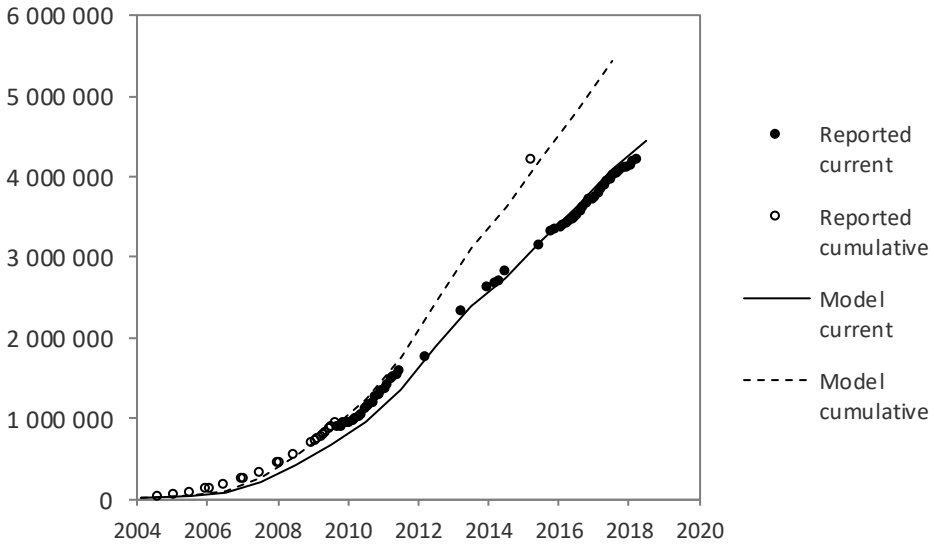


Figure 8.10: ART enrolment in the South African public sector
Dots represent reported numbers receiving ART in the public sector. Solid and dashed lines represent the posterior mean model estimates of numbers currently and cumulatively enrolled on the ART programme, respectively.

Figure 8.11 shows that the modelled age distribution of adults on ART compared against two data sources. The first is laboratory data on numbers of individuals with viral load tests done over the period from October 2018 to September 2019, obtained from the National Health Laboratory Service (NHLS). The second is data from the National Department of Health's nationally-implemented TB/HIV information system (THIS) in June 2018, excluding the Western Cape. (THIS is the definitive source of nationally-reported HIV, ART, and drug-sensitive TB performance data, specifically all child and adult 'total remaining on ART' figures.) Although the NHLS data suggest a slightly younger age distribution in female ART patients than estimated by the model (Figure 8.11a), the THIS data appear quite consistent with the modelled age distribution of female ART patients (Figure 8.11c). The NHLS and THIS data both suggest substantially more male ART patients in the 50+ age group than estimated by Thembisa, and fewer men on ART in the 35-39 age group (Figures 8.11b and 8.11d).

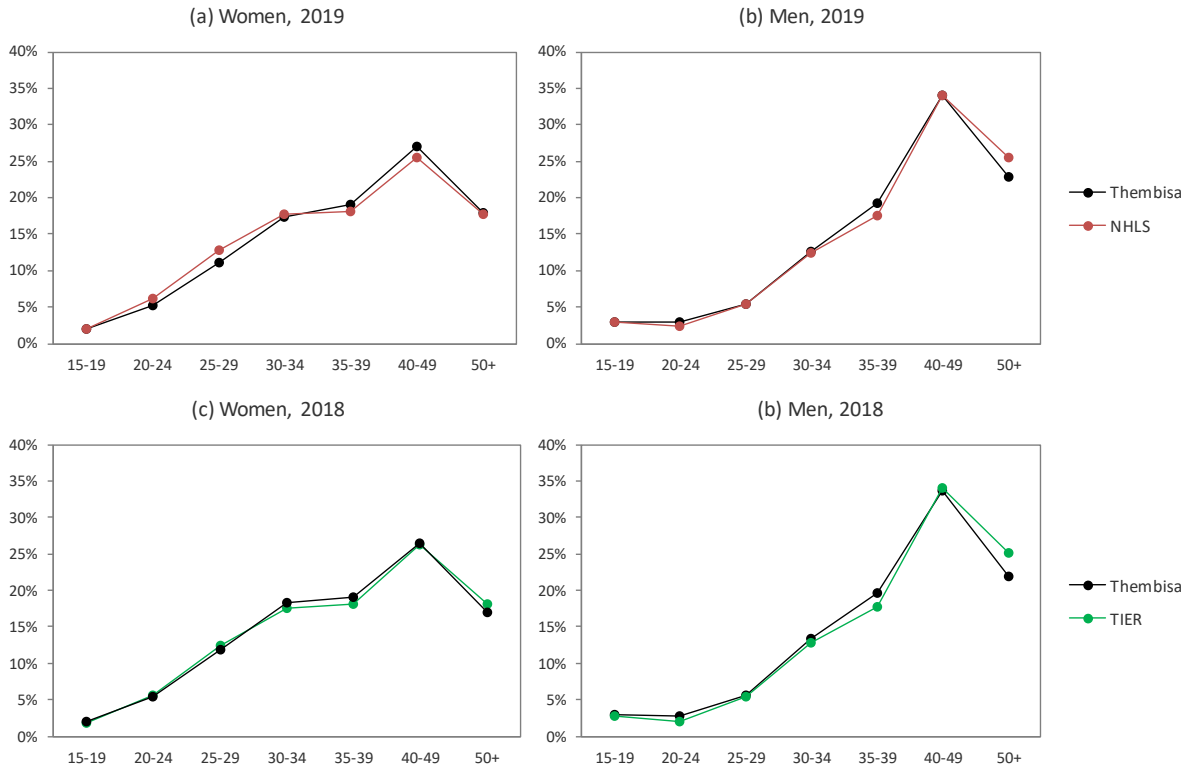


Figure 8.11: Age distribution of adult patients on ART

8.7 Comparison of Thembisa versions 4.2 and 4.3

Figure 8.12 compares a selection of model outputs from Thembisa versions 4.2 and 4.3. At most ages and in most years, HIV prevalence is estimated to be higher in Thembisa version 4.3 than in version 4.2 (Figure 8.12a), largely due to the change in the assumptions about antenatal bias (described in section 7.2.1), but also because of increases in client-to-sex worker transmission in the earlier stages of the HIV epidemic (described in Appendix C), which cause increases in HIV prevalence in the earlier stages of the HIV epidemic. Most of the increase in adult HIV prevalence is in women, with relatively little change in male estimates of HIV prevalence between versions 4.2 and 4.3 (Figure 8.12b-c). Version 4.3 produces a substantially higher estimate of HIV prevalence in children (Figure 8.12d) and mother-to-child transmission (Figure 8.12e), largely as a result of increased estimates of births to HIV-positive mothers (Figure 8.12f). The latter change is partly driven by the increased estimate of HIV prevalence in adult women, but also by revisions to assumptions about antenatal bias and revisions to assumptions about fertility rates in women on ART (in Thembisa version 4.3 we assume a significant increase in fertility after women start ART, in contrast to Thembisa version 4.2, which assumed only a modest increase in fertility mediated by CD4 recovery after ART initiation).

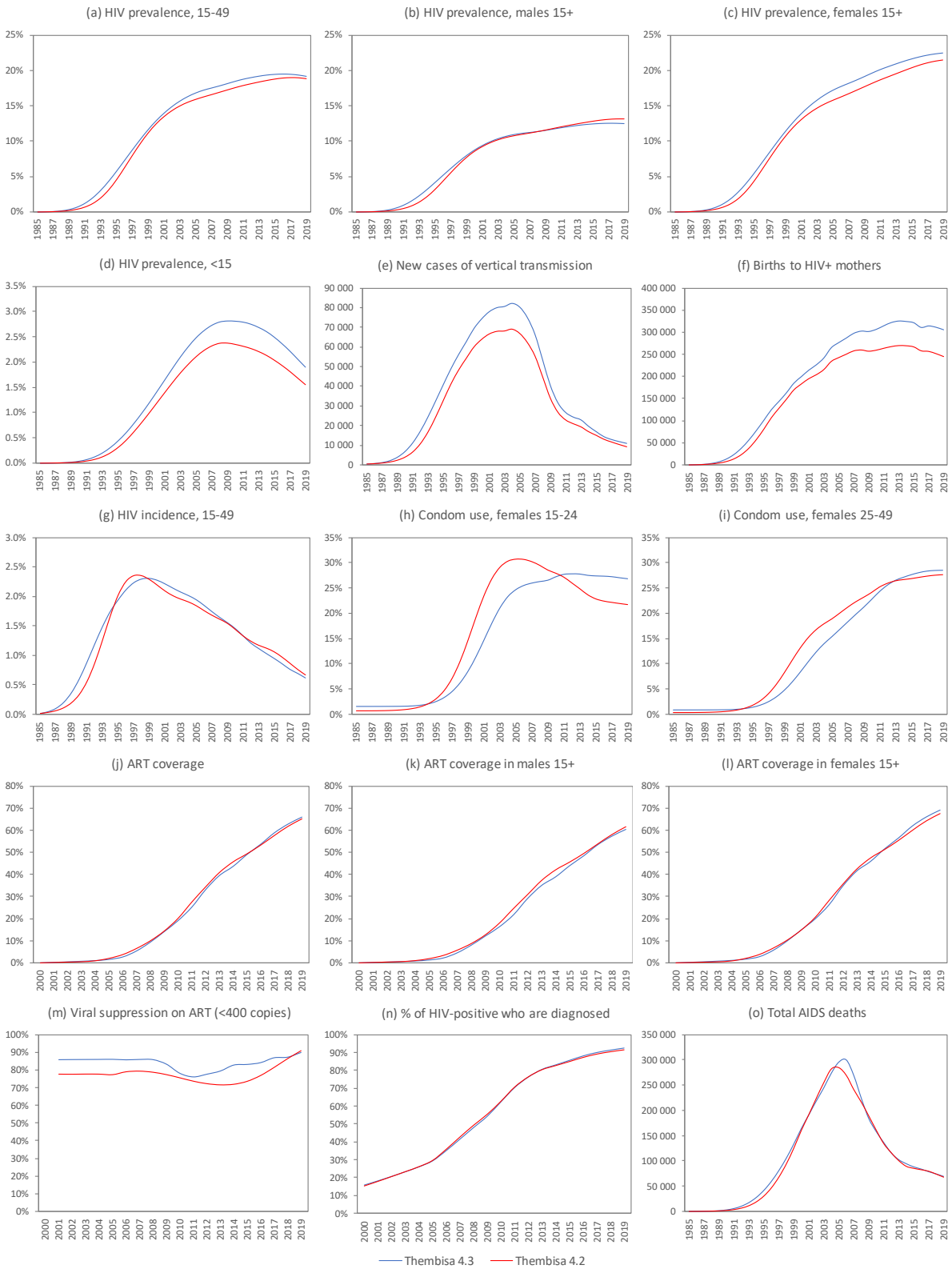


Figure 8.12: Key outputs of Thembisa versions 4.2 (red) and 4.3 (blue)

Estimates of adult HIV incidence in the early stages of the HIV epidemic are higher in version 4.3 than in version 4.2 (Figure 8.12g) because of the assumed increase in client-to-sex worker transmission in the early stages of the HIV epidemic. However, in the period after

1994, the estimates of adult HIV incidence in the two versions are roughly consistent. Estimates of trends in condom use are quite different in the two versions (Figure 8.12h-i), as the more formal statistical approach to estimating condom usage trends (described in section 2.8) yields a slightly later rise in condom use and negligible evidence of 'risk compensation' (which was assumed in Thembisa 4.2 to have led to substantial declines in condom use in youth). Overall estimates of ART coverage are not much changed when comparing versions 4.2 and 4.3 (Figure 8.12j-l), but the difference in ART coverage between men and women has increased. For example, version 4.2 estimated that in 2018 the difference in ART coverage between adult women and men was 6 percentage points (64.3% versus 58.1%), whereas the difference in version 4.3 is 9 percentage points (66.2% versus 57.2%).

Estimates of viral suppression are substantially higher in version 4.3 than in version 4.2, although the differences in the most recent years are minimal (Figure 8.12m). This is because we have adopted a more formal Bayesian approach in fitting the model of viral suppression, reflecting the uncertainty regarding the bias due to missing viral load data (see Appendix F). In contrast to version 4.2, we have not included data from the NHLS when estimating rates of viral suppression, due to concerns that these data were skewing the assumed relationship between the viral load testing coverage and viral suppression. Estimates of levels of HIV diagnosis (Figure 8.12n) and AIDS deaths (Figure 8.12o) are similar in the two versions of the model.

9. Discussion

Thembisa version 4.3 is a significant advance on Thembisa version 4.2, with several important structural changes that have led to revisions in model estimates. Perhaps most significantly, estimates of HIV prevalence in Thembisa 4.3 are higher than before, and more in line with the results of recent national household surveys. For example, the estimated HIV prevalence in the 15-49 age group in 2017 is 19.5% (95% CI: 19.0-19.9%), higher than the estimate of 19.0% (95% CI: 18.2-19.6%) in Thembisa version 4.2, but more in line with the estimated prevalence of 20.6% (95% CI: 19.2-22.0%) in the 2017 HSRC survey [5] and 21.2% (95% CI: 19.2-23.2%) in the 2016 DHS [126]. Although the new model estimates are clearly still lower than the survey estimates, they fall within the 95% confidence intervals around the survey estimates, which is reassuring.

There are a number of likely explanations for the increase in HIV prevalence. Firstly, the model assumptions about bias in antenatal survey data have been revised. Previously, it was assumed that the difference between the model estimate of HIV prevalence in pregnant women and the antenatal survey estimate was constant over time and across age groups; in the new model we assume that the extent of the difference diminishes as HIV incidence declines, in line with results of agent-based models [ref]. This means that the relative difference is greater at younger ages and in earlier years, and hence there is a shift in the modelled trend in HIV prevalence over time as well as a shift in the modelled age distribution of HIV infection.

A second explanation is that the model assumptions about rates of client-to-sex worker transmission have been revised to reflect higher transmission probabilities in the earlier stages of the HIV epidemic [208]. As transmission between sex workers and clients is important in the early spread of HIV, this model change has the effect of increasing HIV prevalence in the early stages of the HIV epidemic. This also leads to model estimates of HIV prevalence in sex workers that are more consistent with survey data. Although none of the survey estimates are nationally representative, the survey data suggest that levels of HIV prevalence in sex workers have been relatively stable (at around 60%) since the mid-1990s (Figure C1), and Thembisa 4.3 is more consistent with this trend.

Another key change to the model is the revisions to assumptions about fertility in HIV-positive women. In the previous version of Thembisa, fertility in HIV-positive women was assumed to reduce in proportion to their change in coital frequency, which in turn was assumed to depend only on CD4 count. However, our recent analysis of pregnancy incidence in the Western Cape suggests that much of the increase in pregnancy incidence that occurs after the initiation of ART is independent of CD4 recovery [3], and we have therefore allowed for a substantial increase in fertility after ART initiation. In addition, we allow for the possibility that women who have recently acquired HIV might be more likely to fall pregnant than HIV-negative women (by virtue of their higher levels of recent unprotected sexual activity), and that HIV-diagnosed women might be less likely to become pregnant than undiagnosed HIV-positive women (because HIV diagnosis is associated with reduced childbearing intentions in women who are untreated [342-344]). The net effect of these changes, together with the increases in HIV prevalence in women of reproductive age, is a substantial increase in estimated numbers of births to HIV-positive women. This in turn has

led to increased estimates of mother-to-child transmission and paediatric HIV prevalence. The new estimates of paediatric HIV prevalence are more consistent with the results of the 2017 HSRC survey; in the survey, paediatric HIV prevalence was 2.7% (95 % CI: 2.2-3.3%) [5], which compares to an estimate of 2.2% (95% CI: 2.2-2.3%) in Thembisa version 4.3 and 1.8% (95% CI: 1.7-1.9%) in version 4.2.

Despite the significant revisions to the model assumptions about antenatal bias and the effects of HIV on fertility, there is still room for improvement in the modelling of these dynamics. Although antenatal bias is related to the average recency of HIV infection, and this dynamic is now incorporated in the model calibration, there are a number of other sources of antenatal bias that could be more explicitly parameterized. These include the bias due to the surveys being conducted only in women attending public antenatal facilities (i.e. no representation of women using private facilities), the bias due to differences in breastfeeding durations between HIV-positive and HIV-negative women (important because lactational amenorrhoea is a significant determinant of fertility [375]), and the bias due to the lack of confirmatory testing in some of the surveys (i.e. HIV prevalence could be over-estimated due to false-positive test results). Accounting for these biases more explicitly would ensure both more reliable estimates of births to HIV-positive mothers, and better estimates of HIV prevalence trends over time.

Several changes have been made to the modelling of viral suppression after ART initiation. In the previous version of Thembisa we used a cubic function to represent the time trend in viral suppression. In Thembisa version 4.3 we instead assume that the trend in viral suppression levels is the same as in IeDEA-SA cohorts over the 2005-2018 period [189], but allowing for the level of viral suppression to be different from that in the IeDEA-SA cohorts. The extent of this difference is estimated through a formal Bayesian approach, incorporating TIER data on viral suppression. These results suggest higher levels of viral suppression than estimated in the previous version of Thembisa, which is mainly because in the previous version of Thembisa, the patients who were missing viral load measurements in TIER were assumed more likely to be unsuppressed than patients with recorded viral loads. Our revised analysis suggests relatively little difference in viral suppression between patients with and without recorded viral loads (Appendix F).

A limitation of Thembisa is that it does not take into account that there are significant age and sex differences in viral suppression in ART patients [376-378]. This has implications for modelling the effect of ART on HIV incidence, particularly as it is younger adults and males who tend to generate the most transmission in the absence of ART, and these are also the groups who are least likely to be virally suppressed after ART initiation. Although Thembisa currently produces estimates of viral suppression separately for men and women, these estimates understate the true extent of the sex difference in viral suppression. Another limitation of our model of viral suppression is that we do not currently link the assumptions about viral suppression to the assumptions about mortality after ART initiation, though some association would be expected. Further work is required to better characterize and model this association, using local data.

Another new development is that the model is calibrated to antiretroviral metabolite data. These data are important as they provide indirect information about HIV prevalence in a population. If the numbers of patients on ART is known with a high degree of precision, then the survey estimate of ART coverage tells us something about whether the survey is likely to

be biased towards over-estimating or under-estimating HIV prevalence in the population. For example, the 2017 HSRC survey estimated there were about 4.40 million people on ART in 2017 [5], which is higher than programme data (3.94 million in the public sector plus approximately 300 000 in private sector gives 4.24 million, as at mid-2017). Although this difference is not large, it suggests that the survey might have slightly over-sampled HIV-positive individuals, i.e. HIV prevalence might have been slightly over-estimated.

We have also validated Thembisa using new data on the age distribution of adults on ART, obtained from both the NHLS and THIS. Although Thembisa is roughly consistent with both data sources, consistency could be improved (particularly in the numbers of older men on ART), and including these data sources in future calibrations of the model may help to improve the model consistency with the NHLS and THIS data.

Another important change in Thembisa version 4.3 is that we have distinguished between first-line and second-line ART. This is important for the purpose of costing the ART programme, as second-line antiretroviral drugs are more expensive than first-line drugs. Rates of switching to second-line ART are assumed to depend on baseline CD4 count, and differ for adults and children. A limitation of the model is that we do not model the differences in mortality rates between patients on first-line and second-line treatment, nor do we model the effect of viral suppression on rates of switching to second line. There is also a lack of historic data on numbers of patients on second-line ART, which makes it difficult to predict the future demand for second-line ART with much confidence.

A key change to the behavioural assumptions is that the new model is based on a more formal statistical analysis of nationally representative data on condom use. Previously we assumed that there was significant 'risk compensation' in recent years, causing reductions in condom use, particularly among youth. This assumption was originally prompted by the publication of the 2012 HSRC survey, which suggested significant declines in condom use relative to the 2008 survey [4]. However, our more systematic analysis of all the national data suggests that there is little, if any, evidence of risk compensation.

Future revisions to Thembisa will need to consider the potential effect of HIV on sexual debut in vertically-infected adolescents – an increasingly important issue given the growing numbers of vertically-infected adolescents [310]. Although there is some data to suggest that HIV is associated with a delay in sexual debut [379, 380], data are sparse and there is little consistency across settings. Allowing for the effect of HIV on sexual debut may nevertheless be important both for predicting the extent to which vertically-acquired HIV contributes to HIV transmission amongst adolescents and young adults, and also in terms of understanding the bias in antenatal survey data in young women.

Future revisions to Thembisa will also need to consider the increasingly substantial role of self-testing in HIV diagnosis, and the implications this has for the calibration to routine testing data. It will also be important to model index testing (testing partners of newly diagnosed individuals) more explicitly, in part because self-testing kits are increasingly being targeted to partners of newly diagnosed individuals, and in part because index testing is one of the most cost-effective HIV testing strategies [381], which needs to be modelled more explicitly.

Acknowledgements

The updates to version 4.3 of the Thembisa model were funded by UNAIDS (in partnership with the South African Department of Health). Earlier funding for the development of the Thembisa model was provided by the South African National AIDS Council, the National Institutes of Health (grant 1R01AI094586-01) and the Hasso Plattner Foundation.

We are grateful to the following individuals who provided input and assisted us in obtaining data for the current version of Thembisa: Thapelo Seatlhodi, Tshepo Molapo, Riona Govender, Gregory Kuzmak, Natsai Shoko, Billia Luwaca, Jeffrey Eaton, Gesine Meyer-Rath, Ali Feizzadeh and Sarah Porter.

References

1. Johnson LF, Dorrington RE. Modelling the impact of HIV in South Africa's provinces: 2020 update. Centre for Infectious Disease Epidemiology and Research, University of Cape Town; 2020. Available: https://www.thembisa.org/
2. Johnson LF, Dorrington RE. Thembisa version 4.2: A model for evaluating the impact of HIV/AIDS in South Africa. University of Cape Town; 2019. Available: https://www.thembisa.org/
3. Johnson LF, Mutemaringa T, Heekes A, Boulle A. The effect of HIV and antiretroviral treatment on pregnancy rates in the Western Cape province of South Africa. *J Infect Dis* 2020; **221**:1953-1962.
4. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, *et al.* South African National HIV Prevalence, Incidence, and Behaviour Survey, 2012. Cape Town: Human Sciences Research Council; 2014. Available: http://www.hsrc.ac.za/en/research-outputs/view/6871. Accessed 16 April 2014
5. Simbayi LC, Zuma K, Zungu N, Moyo S, Marinda E, Jooste S, *et al.* South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2017. Cape Town: Human Sciences Research Council; 2019. Available: https://www.hsrcpress.ac.za/books/south-african-national-hiv-prevalence-incidence-behaviour-and-communication-survey-2017. Accessed 6 Nov 2019
6. Johnson LF, Patrick M, Stephen C, Patten G, Dorrington RE, Maskew M, *et al.* Steep declines in pediatric AIDS mortality in South Africa, despite poor progress towards pediatric diagnosis and treatment targets. *Pediatr Infect Dis J* 2020; [In press].
7. Dunkle KL, Jewkes RK, Murdock DW, Sikweyiya Y, Morrell R. Prevalence of consensual male-male sex and sexual violence, and associations with HIV in South Africa: a population-based cross-sectional study. *PLoS Med* 2013; **10**:e1001472.
8. Cloete A, Simbayi LC, Rehle T, Jooste S, Mabaso M, Townsend L, *et al.* The South African Marang Men's Project: HIV bio-behavioral surveys using respondent-driven sampling conducted among men who have sex with men in Cape Town, Durban and Johannesburg. Cape Town: Human Sciences Research Council; 2014. Available: http://www.hsrcpress.ac.za/product.php?productid=2328&cat=0&page=1&featured&freedownload=1. Accessed 20 Nov 2014
9. Lane T, Osmand T, Marr A, Shade SB, Dunkle K, Sandfort T, *et al.* The Mpumalanga Men's Study (MPMS): Results of a baseline biological and behavioral HIV surveillance survey in two MSM communities in South Africa. *PLoS One* 2014; **9**:e111063.
10. Lane T, Raymond HF, Dladla S, Rasethe J, Struthers H, McFarland W, *et al.* High HIV prevalence among men who have sex with men in Soweto, South Africa: results from the Soweto Men's Study. *AIDS Behav* 2011; **15**:626-634.
11. Vu L, Tun W, Sheehy M, Nel D. Levels and correlates of internalized homophobia among men who have sex with men in Pretoria, South Africa. *AIDS Behav* 2012; **16**:717-723.
12. Nattrass N, Maughan-Brown B, Seekings J, Whiteside A. Poverty, sexual behaviour, gender and HIV infection among young black men and women in Cape Town, South Africa. *Afr J AIDS Res* 2012; **11**:307-317.

13. Jewkes RK, Nduna M, Jama PN, Dunkle KL, Levin JB. Steadys, roll-ons and hit and runs: using indigenous typology to measure number of sexual partners [Abstract TuPpE2069]. *14th International AIDS Conference*. Barcelona, Spain; 2002.
14. Kenyon CR, Osbak K, Buyze J, Johnson S, van Lankveld J. Variations of sexual scripts relating to concurrency by race, class, and gender in South Africa. *J Sex Res* 2015; **52**:878-886.
15. Johnson LF, Hallett TB, Rehle TM, Dorrington RE. The effect of changes in condom usage and antiretroviral treatment coverage on HIV incidence in South Africa: a model-based analysis. *J Roy Soc Interface* 2012; **9**:1544-1554.
16. Williams B, Gilgen D, Campbell C, Taljaard D, MacPhail C. *The natural history of HIV / AIDS in South Africa: A biomedical and social survey in Carletonville*. Johannesburg: Council for Scientific and Industrial Research; 2000.
17. Hallman K. Socioeconomic Disadvantage and Unsafe Sexual Behaviors among Young Women and Men in South Africa. New York: Population Council, Policy Research Division; 2004.
18. Shisana O, Rehle T, Simbayi LC, Parker W, Zuma K, Bhana A, *et al.* South African National HIV Prevalence, HIV Incidence, Behaviours and Communication Survey, 2005. Cape Town: HSRC Press; 2005. Available: http://www.hsrcpress.ac.za. Accessed 1 Dec 2005
19. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Pillay-van Wyk V, *et al.* South African national HIV prevalence, incidence, behaviour and communication survey, 2008: A turning tide among teenagers? Cape Town: Human Sciences Research Council; 2009. Available: http://www.hsrcpress.ac.za. Accessed 9 June 2009
20. Kelly K. Communicating for action: A contextual evaluation of youth responses to HIV/AIDS. Department of Health; 2000. Available: http://www.cadre.org.za. Accessed 12 October 2006
21. Department of Health. South Africa Demographic and Health Survey 1998: Full Report. 1999.
22. Dunkle KL, Jewkes RK, Brown HC, Gray GE, McIntryre JA, Harlow SD. Transactional sex among women in Soweto, South Africa: prevalence, risk factors and association with HIV infection. *Soc Sci Med* 2004; **59**:1581-1592.
23. Mpofu E, Flisher AJ, Bility K, Onya H, Lombard C. Sexual partners in a rural South African setting. *AIDS Behav* 2006; **10**:399-404.
24. Pettifor AE, van der Straten A, Dunbar MS, Shiboski SC, Padian NS. Early age of first sex: a risk factor for HIV infection among women in Zimbabwe. *AIDS* 2004; **18**:1435-1442.
25. Dunkle KL, Jewkes R, Nduna M, Jama N, Levin J, Sikweyiya Y, *et al.* Transactional sex with casual and main partners among young South African men in the rural Eastern Cape: prevalence, predictors, and associations with gender-based violence. *Soc Sci Med* 2007; **65**:1235-1248.
26. Harrison A, Cleland J, Frohlich J. Young people's sexual partnerships in KwaZulu-Natal, South Africa: patterns, contextual influences, and HIV risk. *Stud Fam Plann* 2008; **39**:295-308.
27. Wand H, Ramjee G. The relationship between age of coital debut and HIV seroprevalence among women in Durban, South Africa: a cohort study. *BMJ Open* 2012; **2**:e000285.
28. Health and Development Africa. Impact Assessment of the Khomanani Campaign 2004 - 2006. 2007.

29. Johnson S, Kincaid L, Laurence S, Chikwava F, Delate R, Mahlasela L. Second National HIV Communication Survey 2009. Pretoria: Johns Hopkins Health and Education in South Africa; 2010. Available: http://jhhesa.org.za/docs/NCS_2009.pdf. Accessed 8 March 2011
30. Thurston IB, Dietrich J, Bogart LM, Otwombe KN, Sikkema KJ, Nkala B, *et al.* Correlates of sexual risk among sexual minority and heterosexual South African youths. *Am J Public Health* 2014; **104**:1265-1269.
31. Peltzer K. Sexual orientation and HIV risk among youth in South Africa: a brief report. *J Psychol Afr* 2014; **24**:241-245.
32. Mensch BS, Hewett PC, Erulkar AS. Reporting of sensitive behavior by adolescents: a methodological experiment in Kenya. *Demography* 2003; **40**:247-268.
33. Johnson LF, Dorrington RE, Bradshaw D, Pillay-Van Wyk V, Rehle TM. Sexual behaviour patterns in South Africa and their association with the spread of HIV: insights from a mathematical model. *Demographic Res* 2009; **21**:289-340.
34. Johnson LF, Dorrington RE. Modelling the impact of HIV in South Africa's provinces: 2019 update. University of Cape Town; 2019. Available: https://www.thembisa.org/
35. Arnold MP, Struthers H, McIntyre J, Lane T. Contextual correlates of per partner unprotected anal intercourse rates among MSM in Soweto, South Africa. *AIDS Behav* 2013; **17** (Suppl 1):S4-11.
36. Statistics South Africa. Marriages and divorces, 2004. Pretoria; 2006. Available: http://www.statssa.gov.za/publications/Report-03-07-01/Report-03-07-012004.pdf. Accessed 2 Aug 2008
37. Bah S. The improvement of marriages and divorces statistics in South Africa: Relevance, registration issues and challenges. Population Studies Centre, University of Western Ontario; 1999. Available: http://sociology.uwo.ca/popstudies/dp/dp99-2.pdf. Accessed 17 Oct 2006
38. Leclerc PM, Garenne M. Clients of commercial sex workers in Zambia: prevalence, frequency and risk factors. *Open Demography J* 2008; **1**:1-10.
39. Van der Ryst E, Joubert G, Steyn F, Heunis C, Le Roux J, Williamson C. HIV/AIDS-related knowledge, attitudes and practices among South African military recruits. *S Afr Med J* 2001; **91**:587-591.
40. Ramjee G, Gouws E. Prevalence of HIV among truck drivers visiting sex workers in KwaZulu-Natal, South Africa. *Sex Transm Dis* 2002; **29**:44-49.
41. Jochelson K, Mothibeli M, Leger JP. Human immunodeficiency virus and migrant labor in South Africa. *Int J Health Serv* 1991; **21**:157-173.
42. University of California San Francisco, Anova Health Institute, Wits Reproductive Health and HIV Research Institute. South African Health Monitoring Study (SAHMS), Final Report: The Integrated Biological and Behavioural Survey among Female Sex Workers, South Africa 2013-2014. San Francisco; 2015. Available: https://www.health-e.org.za/wp-content/uploads/2016/03/South-African-Health-Monitoring-Survey-An-Integrated-Biological-and-Behavioral-Survey-among-Female-Sex-Workers-South-Africa-2013-2014.pdf. Accessed 9 June 2016
43. Konstant TL, Rangasami J, Stacey MJ, Stewart ML, Nogoduka C. Estimating the number of sex workers in South Africa: rapid population size estimation. *AIDS Behav* 2015; **19** (Suppl 1):S3-15.
44. Varga CA. The condom conundrum: barriers to condom use among commercial sex workers in Durban, South Africa. *Afr J Reprod Health* 1997; **1**:74-88.

45. Abdool Karim QA, Abdool Karim SS, Soldan K, Zondi M. Reducing the risk of HIV infection among South African sex workers: socioeconomic and gender barriers. *Am J Public Health* 1995; **85**:1521-1525.
46. Ramjee G, Abdool Karim SS, Sturm AW. Sexually transmitted infections among sex workers in KwaZulu-Natal, South Africa. *Sex Transm Dis* 1998; **25**:346-349.
47. Dunkle KL, Beksinska ME, Rees VH, Ballard RC, Htun Y, Wilson ML. Risk factors for HIV infection among sex workers in Johannesburg, South Africa. *Int J STD AIDS* 2005; **16**:256-261.
48. Gould C, Fick N. *Selling Sex in Cape Town: Sex Work and Human Trafficking in a South African City*: Institute for Security Studies; 2008.
49. Peltzer K, Seoka P, Raphala S. Characteristics of female sex workers and their HIV/AIDS/STI knowledge, attitudes and behaviour in semi-urban areas in South Africa. *Curationis* 2004; March 2004:4-11.
50. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, *et al.* Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS One* 2008; **3**:e1954.
51. Delva W, Richter M, De Koker P, Chersich M, Temmerman M. Sex work during the 2010 FIFA World Cup: results from a three-wave cross-sectional survey. *PLoS One* 2011; **6**:e28363.
52. Rees H, Beksinska ME, Dickson-Tetteh K, Ballard RC, Htun Y. Commercial sex workers in Johannesburg: risk behaviour and HIV status. *S Afr J Sci* 2000; **96**:283-284.
53. Luseno WK, Wechsberg WM. Correlates of HIV testing among South African women with high sexual and substance-use risk behaviours. *AIDS Care* 2009; **21**:178-184.
54. Varga CA. Coping with HIV/AIDS in Durban's commercial sex industry. *AIDS Care* 2001; **13**:351-365.
55. Richter ML, Chersich M, Temmerman M, Luchters S. Characteristics, sexual behaviour and risk factors of female, male and transgender sex workers in South Africa. *S Afr Med J* 2013; **103**:246-251.
56. Garnett G, Anderson R. Sexually transmitted diseases and sexual behaviour: insights from mathematical models. *J Infect Dis* 1996; **174**:S150-S160.
57. Wingood GM, Reddy P, Lang DL, Saleh-Onoya D, Braxton N, Sifunda S, *et al.* Efficacy of SISTA South Africa on sexual behavior and relationship control among isiXhosa women in South Africa: results of a randomized-controlled trial. *J Acquir Immun Defic Syndr* 2013; **63** (Suppl 1):S59-65.
58. Harling G, Newell ML, Tanser F, Kawachi I, Subramanian S, Bärnighausen T. Do age-disparate relationships drive HIV incidence in young women? Evidence from a population cohort in rural KwaZulu-Natal, South Africa. *J Acquir Immun Defic Syndr* 2014; **66**:443-451.
59. Eaton LA, Pitpitan EV, Kalichman SC, Sikkema KJ, Skinner D, Watt MH, *et al.* Men who report recent male and female sex partners in Cape Town, South Africa: an understudied and underserved population. *Arch Sex Behav* 2013; **42**:1299-1308.
60. Sex Worker Education and Advocacy Taskforce. Beginning to build the picture: South African national survey of sex worker knowledge, experiences and behaviour. 2013.
61. Hankins C, Tran T, Lapointe N. Sexual behaviour and pregnancy outcome in HIV-infected women. *J Acquir Immun Defic Syndr* 1998; **18**:479-487.

62. Greenblatt RM, Bacchetti P, Barkan S, Augenbraun M, Silver S, Delapenha R, *et al.* Lower genital tract infections among HIV-infected and high-risk uninfected women: findings of the Women's Interagency HIV Study (WIHS). *Sex Transm Dis* 1999; **26**:143-151.
63. Sedgh G, Larsen U, Spiegelman D, Msamanga G, Fawzi WW. HIV-1 disease progression and fertility in Dar es Salaam, Tanzania. *J Acquir Immun Defic Syndr* 2005; **39**:439-445.
64. Loko MA, Toure S, Dakoury-Dogbo N, Gabillard D, Leroy V, Anglaret X. Decreasing incidence of pregnancy by decreasing CD4 cell count in HIV-infected women in Côte d'Ivoire: a 7-year cohort study. *AIDS* 2005; **19**:443-445.
65. Myer L, Carter RJ, Katyal M, Toro P, El-Sadr WM, Abrams EJ. Impact of antiretroviral therapy on incidence of pregnancy among HIV-infected women in Sub-Saharan Africa: a cohort study. *PLoS Med* 2010; **7**:e1000229.
66. McClelland RS, Hassan WM, Lavreys L, Richardson BA, Mandaliya K, Ndinya-Achola J, *et al.* HIV-1 acquisition and disease progression are associated with decreased high-risk sexual behaviour among Kenyan female sex workers. *AIDS* 2006; **20**:1969-1973.
67. McGrath N, Richter L, Newell ML. Sexual risk after HIV diagnosis: a comparison of pre-ART individuals with CD4>500 cells/µl and ART-eligible individuals in a HIV treatment and care programme in rural KwaZulu-Natal, South Africa. *J Int AIDS Soc* 2013; **16**:18048.
68. McClelland RS, Baeten JM, Richardson BA, Lavreys L, Emery S, Mandaliya K, *et al.* A comparison of genital HIV-1 shedding and sexual risk behavior among Kenyan women based on eligibility for initiation of HAART according to WHO guidelines. *J Acquir Immun Defic Syndr* 2006; **41**:611-615.
69. Weinhardt LS, Carey MP, Johnson BT, Bickham NL. Effects of HIV counseling and testing on sexual risk behavior: a meta-analytic review of published research, 1985-1997. *Am J Public Health* 1999; **89**:1397-1405.
70. Matovu JKB, Gray RH, Makumbi F, Wawer MJ, Serwadda D, Kigozi G, *et al.* Voluntary HIV counseling and testing acceptance, sexual risk behavior and HIV incidence in Rakai, Uganda. *AIDS* 2005; **19**:503-511.
71. Cremin I, Nyamukapa C, Sherr L, Hallett TB, Chawira G, Cauchemez S, *et al.* Patterns of self-reported behaviour change associated with receiving voluntary counselling and testing in a longitudinal study from Manicaland, Zimbabwe. *AIDS Behav* 2010; **14**:708-715.
72. Cawley C, Wringe A, Slaymaker E, Todd J, Michael D, Kumugola Y, *et al.* The impact of voluntary counselling and testing services on sexual behaviour change and HIV incidence: observations from a cohort study in rural Tanzania. *BMC Infect Dis* 2014; **14**:159.
73. Marlow HM, Maman S, Moodley D, Curtis S, McNaughton Reyes L. HIV status and postpartum contraceptive use in an antenatal population in Durban, South Africa. *Contraception* 2015; **91**:39-43.
74. Ngubane N, Patel D, Newell ML, Coovadia HM, Rollins N, Coutsoudis A, *et al.* Messages about dual contraception in areas of high HIV prevalence are not heeded. *S Afr Med J* 2008; **98**:209-212.
75. Morroni C, Myer L, Mlobeli R, Gutin S, Grimsrud A. Dual protection among South African women and men: perspectives from HIV care, family planning and sexually transmitted infection services. Women's Health Research Unit, University of Cape Town; 2007. Available: http://www.acquireproject.org/archive/files/6.0_integrate_fp-

lapms/6.2_resources/6.2.2_studies/south_africa_dual_protection_final_report.pdf. Accessed 10 July 2016
76. George G, Beckett S, Cawood C, Khanyile D, Govender K, Kharsany ABM. Impact of HIV testing and treatment services on risky sexual behaviour in the uMgungundlovu District, KwaZulu-Natal, South Africa: a cross-sectional study. *AIDS Res Ther* 2019; **16**:20.
77. Mwangi M, Bunnell R, Nyoka R, Gichangi A, Makokha E, Kim A, *et al.* Unsafe sex among HIV-infected adults in Kenya: results of a nationally representative survey. *J Acquir Immun Defic Syndr* 2011; **58**:80-88.
78. Voluntary HIV-1 Counselling and Testing Efficacy Study Group. Efficacy of voluntary HIV-1 counselling and testing in individuals and couples in Kenya, Tanzania and Trinidad: a randomised trial. *Lancet* 2000; **356**:103-112.
79. Müller O, Sarangbin S, Ruxrungtham K, Sittitrai W, Phanuphak P. Sexual risk behaviour reduction associated with voluntary HIV counselling and testing in HIV infected patients in Thailand. *AIDS Care* 1995; **7**:567-572.
80. Wamoyi J, Mbonye M, Seeley J, Birungi J, Jaffar S. Changes in sexual desires and behaviours of people living with HIV after initiation of ART: implications for HIV prevention and health promotion. *BMC Public Health* 2011; **11**:633.
81. McClelland RS, Graham SM, Richardson BA, Peshu N, Masese LN, Wanje GH, *et al.* Treatment with antiretroviral therapy is not associated with increased sexual risk behavior in Kenyan female sex workers. *AIDS* 2010; **24**:891-897.
82. Moatti JP, Prudhomme J, Traore DC, Juillet-Amari A, Akribi HA, Msellati P. Access to antiretroviral treatment and sexual behaviours of HIV-infected patients aware of their serostatus in Côte d'Ivoire. *AIDS* 2003; **17 (Suppl 3)**:S69-77.
83. Venkatesh KK, de Bruyn G, Lurie MN, Mohapi L, Pronyk P, Moshabela M, *et al.* Decreased sexual risk behavior in the era of HAART among HIV-infected urban and rural South Africans attending primary care clinics. *AIDS* 2010; **24**:2687-2696.
84. Doyle JS, Degenhardt L, Pedrana AE, McBryde ES, Guy RJ, Stoové MA, *et al.* Effects of HIV antiretroviral therapy on sexual and injecting risk-taking behavior: a systematic review and meta-analysis. *Clin Infect Dis* 2014; **59**:1483-1494.
85. Risher K, Rehle T, Simbayi L, Shisana O, Celentano DD. Antiretroviral treatment and sexual risk behavior in South Africa. *AIDS Behav* 2016; **20**:710-716.
86. Johnson LF, Dorrington RE, Moolla H. Progress towards the 2020 targets for HIV diagnosis and antiretroviral treatment in South Africa. *South Afr J HIV Med* 2017; **18**:a694.
87. Touloumi G, Pantazis N, Antoniou A, Stirnadel HA, Walker SA, Porter K. Highly active antiretroviral therapy interruption: predictors and virological and immunologic consequences. *J Acquir Immun Defic Syndr* 2006; **42**:554-561.
88. Hughes JP, Baeten JM, Lingappa JR, Magaret AS, Wald A, de Bruyn G, *et al.* Determinants of per-coital-act HIV-1 infectivity among African HIV-1-serodiscordant couples. *J Infect Dis* 2012; **205**:358-365.
89. Prins M, Meyer L, Hessol NA. Sex and the course of HIV infection in the pre- and highly active antiretroviral therapy eras. *AIDS* 2005; **19**:357-370.
90. Lessells RJ, Mutevedzi PC, Cooke GS, Newell ML. Retention in HIV care for individuals not yet eligible for antiretroviral therapy: rural KwaZulu-Natal, South Africa. *J Acquir Immun Defic Syndr* 2011; **56**:e79-86.
91. Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, Mwita W, *et al.* Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and

middle-income countries before highly active antiretroviral therapy. *AIDS* 2007; **21** (Suppl 6):S55-63.
92. Collaborative Group on AIDS Incubation and HIV Survival. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *Lancet* 2000; **355**:1131-1137.
93. Bhaskaran K, Hamouda O, Sannes M, Boufassa F, Johnson AM, Lambert PC, *et al.* Changes in the risk of death after HIV seroconversion compared with mortality in the general population. *JAMA* 2008; **300**:51-59.
94. Johnson LF, May MT, Dorrington RE, Cornell M, Boulle A, Egger M, *et al.* Estimating the impact of antiretroviral treatment on adult mortality trends in South Africa: a mathematical modelling study. *PLoS Med* 2017; **14**:e1002468.
95. Lodi S, Phillips A, Touloumi G, Geskus R, Meyer L, Thiébaut R, *et al.* Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count thresholds <200, <350, and <500 cells/mm^3: assessment of need following changes in treatment guidelines. *Clin Infect Dis* 2011; **53**:817-825.
96. Lodi S, Phillips A, Touloumi G, Pantazis N, Bucher HC, Babiker A, *et al.* CD4 decline in seroconverter and seroprevalent individuals in the precombination of antiretroviral therapy era. *AIDS* 2010; **24**:2697-2704.
97. Erikstrup C, Kallestrup P, Zinyama R, Gomo E, Mudenge B, Gerstoft J, *et al.* Predictors of mortality in a cohort of HIV-1-infected adults in rural Africa. *J Acquir Immun Defic Syndr* 2007; **44**:478-483.
98. Isingo R, Zaba B, Marston M, Ndege M, Mngara J, Mwita W, *et al.* Survival after HIV infection in the pre-antiretroviral therapy era in a rural Tanzanian cohort. *AIDS* 2007; **21** (Suppl 6):S5-S13.
99. Hogg RS, Strathdee SA, Craib KJ, O'Shaughnessy MV, Montaner JS, Schechter MT. Lower socioeconomic status and shorter survival following HIV infection. *Lancet* 1994; **344**:1120-1124.
100. Pezzotti P, Phillips AN, Dorrucci M, Lepri AC, Galai N, Vlahov D, *et al.* Category of exposure to HIV and age in the progression to AIDS: longitudinal study of 1199 people with known dates of seroconversion. *BMJ* 1996; **313**:583-586.
101. Auvert B, Males S, Puren A, Taljaard D, Carael M, Williams B. Can highly active antiretroviral therapy reduce the spread of HIV? A study in a township of South Africa. *J Acquir Immun Defic Syndr* 2004; **36**:613-621.
102. Rehle TM, Shisana O. Estimates of eligibility for antiretroviral treatment (ART) and projected ART impact on AIDS mortality among South African educators. *SAHARA J* 2005; **2**:304-310.
103. Connelly D, Veriava Y, Roberts S, Tsotetsi J, Jordan A, DeSilva E, *et al.* Prevalence of HIV infection and median CD4 counts among health care workers in South Africa. *S Afr Med J* 2007; **97**:115-120.
104. Kranzer K, Lawn SD, Johnson LF, Bekker LG, Wood R. Community viral load and CD4 count distribution among people living with HIV in a South African township: implications for treatment as prevention. *J Acquir Immun Defic Syndr* 2013; **63**:498-505.
105. Malaza A, Mossong J, Bärnighausen T, Viljoen J, Newell ML. Population-based CD4 counts in a rural Area in South Africa with high HIV prevalence and high antiretroviral treatment coverage. *PLoS One* 2013; **8**:e70126.
106. van Rooyen H, Barnabas RV, Baeten JM, Phakathi Z, Joseph P, Krows M, *et al.* High HIV testing uptake and linkage to care in a novel program of home-based HIV

counseling and testing with facilitated referral in KwaZulu-Natal, South Africa. *J Acquir Immun Defic Syndr* 2013; **64**:e1-8.
107. Huerga H, Maman D, Etard JF, Farhat JB, Bouhenia M. Mbongolwane and Eshowe HIV Impact in Population Survey. Epicentre and Medecins sans Frontieres; 2014.
108. Kharsany ABM, Cawood C, Khanyile D, Lewis L, Grobler A, Puren A, *et al.* Community-based HIV prevalence in KwaZulu-Natal, South Africa: results of a cross-sectional household survey. *Lancet HIV* 2018; **5**:e427-e437.
109. Johnson LF. Access to antiretroviral treatment in South Africa, 2004-2011. *South Afr J HIV Med* 2012; **13**:22-27.
110. Payne R, Muenchhoff M, Mann J, Roberts HE, Matthews P, Adland E, *et al.* Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proc Natl Acad Sci U S A* 2014; **111**:E5393-5400.
111. Blanquart F, Grabowski MK, Herbeck J, Nalugoda F, Serwadda D, Eller MA, *et al.* A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. *Elife* 2016; **5**:e20492.
112. Churchyard GJ, Mametja LD, Mvusi L, Ndjeka N, Hesseling AC, Reid A, *et al.* Tuberculosis control in South Africa: successes, challenges and recommendations. *S Afr Med J* 2014; **104**:244-248.
113. Herbeck JT, Müller V, Maust BS, Ledergerber B, Torti C, Di Giambenedetto S, *et al.* Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. *AIDS* 2012; **26**:193-205.
114. Pantazis N, Porter K, Costagliola D, De Luca A, Ghosn J, Guiguet M, *et al.* Temporal trends in prognostic markers of HIV-1 virulence and transmissibility: an observational cohort study. *Lancet HIV* 2014; **1**:e119-126.
115. Johnson LF, Dorrington RE. Thembisa version 4.1: A model for evaluating the impact of HIV/AIDS in South Africa. 2018.
116. Holmes CB, Wood R, Badri M, Zilber S, Wang B, Maartens G, *et al.* CD4 decline and incidence of opportunistic infections in Cape Town, South Africa: implications for prophylaxis and treatment. *J Acquir Immun Defic Syndr* 2006; **42**:464-469.
117. Anglaret X, Minga A, Gabillard D, Ouassa T, Messou E, Morris B, *et al.* AIDS and non-AIDS morbidity and mortality across the spectrum of CD4 cell counts in HIV-infected adults before starting antiretroviral therapy in Côte d'Ivoire. *Clin Infect Dis* 2012; **54**:714-723.
118. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, *et al.* Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* 2005; **191**:1403-1409.
119. Donnell D, Baeten JM, Kiarie J, Thomas KK, Stevens W, Cohen CR, *et al.* Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: a prospective cohort analysis. *Lancet* 2010; **375**:2092-2098.
120. Boily MC, Baggaley RF, Wang L, Masse B, White RG, Hayes RJ, *et al.* Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis* 2009; **9**:118-129.
121. Johnson LF, Mossong J, Dorrington RE, Schomaker M, Hoffmann CJ, Keiser O, *et al.* Life expectancies of South African adults starting antiretroviral treatment: Collaborative analysis of cohort studies. *PLoS Med* 2013; **10**:e1001418.
122. Maheu-Giroux M, Marsh K, Doyle CM, Godin A, Lanièce Delaunay C, Johnson LF, *et al.* National HIV testing and diagnosis coverage in sub-Saharan Africa: a new modeling tool for estimating the 'first 90' from program and survey data. *AIDS* 2019; **33** (Suppl 3):S255-269.

123. Johnson LF, Dorrington RE. Modelling the impact of HIV in South Africa's provinces: 2018 update. Centre for Infectious Disease Epidemiology and Research working paper; 2018. Available: https://www.thembisa.org/downloads
124. Johnson LF, Rehle TM, Jooste S, Bekker LG. Rates of HIV testing and diagnosis in South Africa, 2002-2012: successes and challenges. *AIDS* 2015; **29**:1401-1409.
125. Human Sciences Research Council. HIV Impact Assessment Summary. 2018. Available: http://serve.mg.co.za/content/documents/2018/07/17/7M1RBtUShKFJbN3NL1Wr_HSRC_HIV_Survey_Summary_2018.pdf. Accessed 18 July 2018
126. Department of Health, Statistics South Africa, South African Medical Research Council, ICF. South Africa Demographic and Health Survey 2016. Pretoria; 2019. Available: https://www.dhsprogram.com/pubs/pdf/FR337/FR337.pdf. Accessed 19 March 2019
127. Snow RC, Madalane M, Poulsen M. Are men testing? Sex differentials in HIV testing in Mpumalanga Province, South Africa. *AIDS Care* 2010; **22**:1060-1065.
128. Kincaid DL, Parker W, Johnson S, Schierhout G, Kelly K, Connolly C, *et al.* AIDS Communication Programmes, HIV Prevention, and Living with HIV and AIDS in South Africa, 2006. Pretoria: Johns Hopkins Health and Education in South Africa; 2008. Available: http://jhhesa.org/PDF/National%20Comm%20Report.pdf. Accessed 8 March 2011
129. Pettifor A, MacPhail C, Suchindran S, Delany-Moretlwe S. Factors associated with HIV testing among public sector clinic attendees in Johannesburg, South Africa. *AIDS Behav* 2010; **14**:913-921.
130. Dalal S, Lee CW, Farirai T, Schilsky A, Goldman T, Moore J, *et al.* Provider-initiated HIV testing and counseling: increased uptake in two public community health centers in South Africa and implications for scale-up. *PLoS One* 2011; **6**:e27293.
131. McCoy D, Besser M, Visser R, Doherty T. Interim findings on the national PMTCT pilot sites: lessons and recommendations. Durban: Health Systems Trust; 2002. Available: http://www.hst.org.za/publications/478. Accessed 9 April 2006
132. Ramkissoon A, Kleinschmidt I, Beksinska M, Smit J, Hlazo J, Mabude Z. National Baseline Assessment of Sexually Transmitted Infection and HIV Services in South African Public Sector Health Facilities. Durban: Reproductive Health Research Unit; 2004. Available: http://www.rhru.co.za. Accessed 13 February 2004
133. Reagon G, Irlam J, Levin J. The National Primary Health Care Facilities Survey 2003. Durban: Health Systems Trust; 2004. Available: http://www.hst.org.za/publications/617. Accessed 6 Aug 2010
134. Barron P, Day C, Loveday M, Monticelli F. The District Health Barometer Year 1: January-December 2004. Durban: Health Systems Trust; 2005. Available: http://www.hst.org.za/publications/689. Accessed 29 Dec 2008
135. World Health Organization. Global Tuberculosis Control 2009: Epidemiology, Strategy, Financing. Geneva; 2009. Available: http://www.who.int/tb/publications/global_report/2009/pdf/full_report.pdf. Accessed 5 March 2012
136. Barron P, Day C, Monticelli F, Vermaak K, Okorafor O, Moodley K, *et al.* District Health Barometer 2005/06. Health Systems Trust; 2006. Available: http://www.hst.org.za/publications/701. Accessed 15 March 2007
137. World Health Organization. Global Tuberculosis Control 2010. Geneva; 2010. Available: http://www.who.int/tb/publications/global_report/2010/en/index.html. Accessed 12 March 2012

138. Stinson K, Boulle A, Coetzee D, Abrams EJ, Myer L. Initiation of highly active antiretroviral therapy among pregnant women in Cape Town, South Africa. *Trop Med Int Health* 2010; **15**:825-832.
139. Barron P, Day C, Monticelli F. The Disrict Health Barometer - Year 2006/07. Health Systems Trust; 2008. Available: http://www.hst.org.za/publications/717. Accessed 22 Feb 2008
140. Day C, Barron P, Monticelli F, Sello E. District Health Barometer 2007/08. Health Systems Trust; 2009. Available: http://www.hst.org.za/publications/850. Accessed 10 July 2009
141. Kigozi NG, Heunis JC, Wouters E, van den Berg HS. Tuberculosis patients' reasons for, and suggestions to address non-uptake of HIV testing: a cross-sectional study in the Free State Province, South Africa. *BMC Health Serv Res* 2011; **11**:110.
142. Médecins Sans Frontières. Khayelitsha Annual Activity Report, 2008-2009. Cape Town; 2010. Available: http://www.msf.org.za/Docs/Khayelitsha/Khayelitsha_Report_2008-2009.pdf. Accessed 24 Feb 2010
143. Day C, Monticelli F, Barron P, Haynes R, Smith J, Sello E. District Health Barometer: Year 2008/09. Durban: Health Systems Trust; 2010. Available: http://www.hst.org.za/publications/864. Accessed 25 June 2010
144. Kranzer K, Zeinecker J, Ginsberg P, Orrell C, Kalawe NN, Lawn SD, *et al.* Linkage to HIV care and antiretroviral therapy in Cape Town, South Africa. *PLoS One* 2010; **5**:e13801.
145. Goga AE, Dinh TH, Jackson DJ. Evaluation of the Effectiveness of the National Prevention of Mother-to-Child Transmission (PMTCT) Programme Measured at Six Weeks Postpartum in South Africa, 2010. South African Medical Research Council, National Department of Health of South Africa and PEPFAR/US Centers for Disease Control and Prevention; 2012. Available: http://www.doh.gov.za/docs/reports/2012/pmtcteffectiveness.pdf. Accessed 12 June 2012
146. World Health Organization. Global Tuberculosis Control 2011. Geneva; 2011. Available: http://whqlibdoc.who.int/publications/2011/9789241564380_eng.pdf. Accessed 12 March 2012
147. Massyn N, Peer N, English R, Padarath A, Barron P, Day C. District Health Barometer 2015/16. Durban; 2016. Available: http://www.hst.org.za/publications/district-health-barometer-201516-0. Accessed 5 March 2017
148. Department of Health. Annual Health Statistics 2012. 2013. Available: http://www.hst.org.za/sites/default/files/AnnualHealthStatistics2012_Aug2013.pdf. Accessed 20 Sept 2016
149. Massyn N, Day C, Barron P, Haynes R, English R, Padarath A. District Health Barometer 2011/12. Durban: Health Systems Trust; 2013. Available: http://www.hst.org.za/publications/district-health-barometer-201112. Accessed 23 Oct 2013
150. Massyn N, Padarath A, Peer N, Day C. District Health Barometer 2016/17. Durban: Health Systems Trust; 2017. Available: http://www.hst.org.za/publications/Pages/District-Health-Barometer-201617.aspx. Accessed 3 April 2018

151. Department of Health. Operational plan for comprehensive HIV and AIDS care, management and treatment for South Africa. 2003. Available: www.info.gov.za/otherdocs/2003/aidsplan.pdf. Accessed 28 Nov 2003.
152. Department of Health. Clinical guidelines for the management of HIV and AIDS in adults and adolescents. 2010. Available: http://www.doh.gov.za/docs/facts-f.html. Accessed 30 July 2010.
153. Mureithi L, Van Schaik N, English R. Changes to South African antiretroviral treatment guidelines in 2012. *Kwik Skwiz*: Health Systems Trust; 2012. Available: http://www.hst.org.za/publications/changes-south-african-antiretroviral-treatment-guidelines-2012. Accessed 23 Dec 2012
154. Department of Health. The South African Antiretroviral Treatment Guidelines 2013. 2013. Available: http://www.kznhealth.gov.za/medicine/2013_art_guidelines.pdf. Accessed 29 April 2014
155. Department of Health. National consolidated guidelines for the prevention of mother-to-child transmission of HIV and the management of HIV in children, adolescents and adults. Pretoria; 2015. Available: http://www.health.gov.za/index.php/2014-03-17-09-09-38/policies-and-guidelines/category/230-2015p. Accessed 12 Aug 2015
156. Larson BA, Brennan A, McNamara L, Long L, Rosen S, Sanne I, *et al.* Early loss to follow up after enrolment in pre-ART care at a large public clinic in Johannesburg, South Africa. *Trop Med Int Health* 2010; **15** (Suppl 1):43-47.
157. Charalambous S, Innes C, Muirhead D, Kumaranayake L, Fielding K, Pemba L, *et al.* Evaluation of a workplace HIV treatment programme in South Africa. *AIDS* 2007; **21** (Suppl 3):S73-78.
158. Nachega JB, Hislop M, Dowdy DW, Lo M, Omer SB, Regensberg L, *et al.* Adherence to highly active antiretroviral therapy assessed by pharmacy claims predicts survival in HIV-infected South African adults. *J Acquir Immun Defic Syndr* 2006; **43**:78-84.
159. Department of Health. ART Programme Analysis: Reviewing the ART programme from April 2004 to March 2014. Pretoria; 2015.
160. Kufa-Chakezha T, De Gita G, Jusu Ballah N, Puren A, Takuva S, Carmona S, *et al.* Determinants of CD4 immune recovery among individuals on ART in South Africa. Johannesburg: National Institute for Communicable Diseases; 2016. Available: http://documents.worldbank.org/curated/en/851301474884707261/Determinants-of-CD4-immune-recovery-among-individuals-on-antiretroviral-therapy-in-South-Africa-a-national-analysis. Accessed 14 June 2017
161. Katz IT, Essien T, Marinda ET, Gray GE, Bangsberg DR, Martinson NA, *et al.* Antiretroviral therapy refusal among newly diagnosed HIV-infected adults. *AIDS* 2011; **25**:2177-2181.
162. Govindasamy D, van Schaik N, Kranzer K, Wood R, Mathews C, Bekker LG. Linkage to HIV care from a mobile testing unit in South Africa by different CD4 count strata. *J Acquir Immun Defic Syndr* 2011; **58**:344-352.
163. Katz IT, Dietrich J, Tshabalala G, Essien T, Rough K, Wright AA, *et al.* Understanding treatment refusal among adults presenting for HIV-testing in Soweto, South Africa: a qualitative study. *AIDS Behav* 2015; **19**:704-714.
164. Rosen S, Fox MP. Retention in HIV care between testing and treatment in sub-Saharan Africa: a systematic review. *PLoS Med* 2011; **8**:e1001056.
165. Stevens WS, Gous NM, MacLeod WB, Long LC, Variava E, Martinson NA, *et al.* Multidisciplinary point-of-care testing in South African primary health care clinics

accelerates HIV ART initiation but does not alter retention in care. *Journal of Acquired Immune Deficiency Syndrome* 2017; **76**:65-73.
166. Mujugira A, Celum C, Thomas KK, Farquhar C, Mugo N, Katabira E, *et al.* Delay of antiretroviral therapy initiation is common in East African HIV-infected individuals in serodiscordant partnerships. *J Acquir Immun Defic Syndr* 2014; **66**:436-442.
167. Bor J, Chiu C, Ahmed S, Katz I, Fox MP, Rosen S, *et al.* Failure to initiate HIV treatment in patients with high CD4 counts: evidence from demographic surveillance in rural South Africa. *Trop Med Int Health* 2018; **23**:206-220.
168. Johnson LF, McLeod HD. Steady growth in antiretroviral treatment provision by disease management and community treatment programmes. *S Afr Med J* 2007; **97**:358-359.
169. Awsumb H, Little K, Aylward P, Hasen N. Characterizing the South African ART market [Abstract WEAD0205]. *9th International AIDS Society Conference*. Paris, France; 2017.
170. Department of Health. Health Indicators Update: Antiretroviral Indicators. 2013. Available: http://www.health.gov.za/reports.php. Accessed 14 May 2014
171. HIV-CAUSAL Collaboration. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. *Ann Intern Med* 2011; **154**:509-515.
172. Mahiane SG, Legeai C, Taljaard D, Latouche A, Puren A, Peillon A, *et al.* Transmission probabilities of HIV and herpes simplex virus type 2, effect of male circumcision and interaction: a longitudinal study in a township of South Africa. *AIDS* 2009; **23**:377-383.
173. Baeten JM, Richardson BA, Lavreys L, Rakwar JP, Mandaliya K, Bwayo JJ, *et al.* Female-to-male infectivity of HIV-1 among circumcised and uncircumcised Kenyan men. *J Infect Dis* 2005; **191**:546-553.
174. Morison L, Weiss HA, Buvé A, Caraël M, Abega S-C, Kaona F, *et al.* Commercial sex and the spread of HIV in four cities in sub-Saharan Africa. *AIDS* 2001; **15**:S61-S69.
175. Ramjee G, Williams B, Gouws E, Van Dyck E, De Deken B, Abdool Karim S. The impact of incident and prevalent herpes simplex virus-2 infection on the incidence of HIV-1 infection among commercial sex workers in South Africa. *J Acquir Immun Defic Syndr* 2005; **39**:333-339.
176. Ramjee G, Weber AE, Morar NS. Recording sexual behavior: comparison of recall questionnaires with a coital diary. *Sex Transm Dis* 1999; **26**:374-380.
177. Pettifor AE, Hudgens MG, Levandowski BA, Rees HV, Cohen MS. Highly efficient HIV transmission to young women in South Africa. *AIDS* 2007; **21**:861-865.
178. Auvert B, Ballard R, Campbell C, Caraël M, Carton M, Fehler G, *et al.* HIV infection in a South African mining town is associated with herpes simplex virus-2 seropositivity and sexual behaviour. *AIDS* 2001; **15**:885-898.
179. Gray R, Wawer M, Brookmeyer R, Sewankambo N, Serwadda D, Wabwire-Mangen F, *et al.* Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet* 2001; **357**:1149-1153.
180. Allen S, Tice J, Van de Perre P, Serufilira A, Hudes E, Nsengumuremyi F, *et al.* Effect of serotesting with counselling on condom use and seroconversion among HIV discordant couples in Africa. *BMJ* 1992; **304**:1605-1609.

181. Jin F, Jansson J, Law M, Prestage GP, Zablotska I, Imrie JC, *et al.* Per-contact probability of HIV transmission in homosexual men in Sydney in the era of HAART. *AIDS* 2010; **24**:907-913.
182. Scott HM, Vittinghoff E, Irvin R, Sachdev D, Liu A, Gurwith M, *et al.* Age, race/ethnicity, and behavioral risk factors associated with per contact risk of HIV infection among men who have sex with men in the United States. *J Acquir Immun Defic Syndr* 2014; **65**:115-121.
183. De Gruttola V, Seage GR, Mayer KH, Horsburgh CR. Infectiousness of HIV between male homosexual partners. *J Clin Epidemiol* 1989; **42**:849-856.
184. Røttingen J, Cameron DW, Garnett GP. A systematic review of the epidemiological interactions between classic sexually transmitted diseases and HIV: how much is really known? *Sex Transm Dis* 2001; **28**:579-597.
185. Sexton J, Garnett G, Røttingen J. Metaanalysis and metaregression in interpreting study variability in the impact of sexually transmitted diseases on susceptibility to HIV infection. *Sex Transm Dis* 2005; **32**:351-357.
186. Johnson LF, Lewis DA. The effect of genital tract infections on HIV-1 shedding in the genital tract: a systematic review and meta-analysis. *Sex Transm Dis* 2008; **35**:946-959.
187. Johnson LF, Dorrington RE, Bradshaw D, Coetzee DJ. The role of sexually transmitted infections in the evolution of the South African HIV epidemic. *Trop Med Int Health* 2012; **17**:161-168.
188. Cornell M, Grimsrud A, Fairall L, Fox MP, van Cutsem G, Giddy J, *et al.* Temporal changes in programme outcomes among adult patients initiating antiretroviral therapy across South Africa, 2002-2007. *AIDS* 2010; **24**:2263-2270.
189. Pillay T. Determinants of HIV viral suppression in South Africans receiving antiretroviral treatment. Cape Town: University of Cape Town; 2019.
190. Kanters S, Vitoria M, Doherty M, Socias ME, Ford N, Forrest JI, *et al.* Comparative efficacy and safety of first-line antiretroviral therapy for the treatment of HIV infection: a systematic review and network meta-analysis. *Lancet HIV* 2016; **3**:e510-e520.
191. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc Natl Acad Sci U S A* 2007; **104**:17441-17446.
192. Quinn T, Wawer M, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. *N Engl J Med* 2000; **342**:921-929.
193. Fideli US, Allen SA, Musonda R, Trask S, Hahn BH, Weiss H, *et al.* Virologic and immunologic determinants of heterosexual transmission of human immunodeficiency virus type 1 in Africa. *AIDS Res Hum Retroviruses* 2001; **17**:901-910.
194. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, *et al.* Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 2011; **365**:493-505.
195. Anglemyer A, Rutherford GW, Horvath T, Baggaley RC, Egger M, Siegfried N. Antiretroviral therapy for prevention of HIV transmission in HIV-discordant couples. *Cochrane Database Syst Rev* 2013; **4**:CD009153.
196. Fox MP, Van Cutsem G, Giddy J, Maskew M, Keiser O, Prozesky H, *et al.* Rates and predictors of failure of first-line antiretroviral therapy and switch to second-line ART in South Africa. *J Acquir Immun Defic Syndr* 2012; **60**:428-437.

197. World Health Organization. Consolidated strategic information guidelines for HIV in the health sector. 2015. Available: http://apps.who.int/iris/bitstream/10665/164716/1/9789241508759_eng.pdf. Accessed 18 Aug 2016
198. UNAIDS. Global AIDS Response Progress Reporting 2016: Construction of core indicators for monitoring the 2011 United Nations Political Declaration on HIV and AIDS. Geneva; 2016. Available: https://aidsreportingtool.unaids.org/static/docs/GARPR_Guidelines_2016_EN.pdf. Accessed 18 Aug 2016
199. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission (Cochrane Review). In: *The Cochrane Library*. Chichester, UK: John Wiley & Sons; 2004.
200. Holmes KK, Levine R, Weaver M. Effectiveness of condoms in preventing sexually transmitted infections. *Bull WHO* 2004; **82**:454-461.
201. Meekers D, Van Rossem R. Explaining inconsistencies between data on condom use and condom sales. *BMC Health Serv Res* 2005; **5**:5.
202. Plourde PJ, Pepin J, Agoki E, Ronald AR, Ombette J, Tyndall M, *et al.* Human immunodeficiency virus type 1 seroconversion in women with genital ulcers. *J Infect Dis* 1994; **170**:313-317.
203. Myer L, Wright TC, Jr., Denny L, Kuhn L. Nested case-control study of cervical mucosal lesions, ectopy, and incident HIV infection among women in Cape Town, South Africa. *Sex Transm Dis* 2006; **33**:683-687.
204. Moss GB, Clemetson D, D'Costa L, Plummer FA, Ndinya-Achola JO, Reilly M, *et al.* Association of cervical ectopy with heterosexual transmission of human immunodeficiency virus: results of a study of couples in Nairobi, Kenya. *J Infect Dis* 1991; **164**:588-591.
205. Petrova MI, van den Broek M, Balzarini J, Vanderleyden J, Lebeer S. Vaginal microbiota and its role in HIV transmission and infection. *FEMS Microbiol Rev* 2013; **37**:762-792.
206. Mackelprang RD, Baeten JM, Donnell D, Celum C, Farquhar C, de Bruyn G, *et al.* Quantifying ongoing HIV-1 exposure in HIV-1-serodiscordant couples to identify individuals with potential host resistance to HIV-1. *J Infect Dis* 2012; **206**:1299-1308.
207. Carpenter LM, Kamali A, Ruberantwari A, Malamba SS, Whitworth JA. Rates of HIV-1 transmission within marriage in rural Uganda in relation to the HIV sero-status of the partners. *AIDS* 1999; **13**:1083-1089.
208. Kimani J, Kaul R, Nagelkerke NJ, Luo M, MacDonald KS, Ngugi E, *et al.* Reduced rates of HIV acquisition during unprotected sex by Kenyan female sex workers predating population declines in HIV prevalence. *AIDS* 2008; **22**:131-137.
209. Boulle A, Van Cutsem G, Hilderbrand K, Cragg C, Abrahams M, Mathee S, *et al.* Seven year experience of a primary care antiretroviral treatment programme in Khayelitsha, South Africa. *AIDS* 2010; **24**:563-572.
210. Nash D, Katyal M, Brinkhof MW, Keiser O, May M, Hughes R, *et al.* Long-term immunologic response to antiretroviral therapy in low-income countries: a collaborative analysis of prospective studies. *AIDS* 2008; **22**:2291-2302.
211. Lok JJ, Bosch RJ, Benson CA, Collier AC, Robbins GK, Shafer RW, *et al.* Long-term increase in CD4+ T-cell counts during combination antiretroviral therapy for HIV-1 infection. *AIDS* 2010; **24**:1867-1876.

212. Schiffer JT, Mayer BT, Fong Y, Swan DA, Wald A. Herpes simplex virus-2 transmission probability estimates based on quantity of viral shedding. *J Roy Soc Interface* 2014; **11**:20140160.
213. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS Med* 2005; **2**:e298.
214. Gray RH, Kigozi G, Serwadda D, Makumbi F, Watya S, Nalugoda F, *et al.* Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *Lancet* 2007; **369**:657-666.
215. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, *et al.* Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *Lancet* 2007; **369**:643-656.
216. Weiss HA, Halperin D, Bailey RC, Hayes RJ, Schmid G, Hankins CA. Male circumcision for HIV prevention: from evidence to action? *AIDS* 2008; **22**:567-574.
217. Weiss HA, Hankins CA, Dickson K. Male circumcision and risk of HIV infection in women: a systematic review and meta-analysis. *Lancet Infect Dis* 2009; **9**:669-677.
218. Millett GA, Flores SA, Marks G, Reed JB, Herbst JH. Circumcision status and risk of HIV and sexually transmitted infections among men who have sex with men: a meta-analysis. *JAMA* 2008; **300**:1674-1684.
219. Sánchez J, Sal y Rosas VG, Hughes JP, Baeten JM, Fuchs J, Buchbinder SP, *et al.* Male circumcision and risk of HIV acquisition among MSM. *AIDS* 2011; **25**:519-523.
220. Connolly C, Simbayi LC, Shanmugam R, Nqeketo A. Male circumcision and its relationship to HIV infection in South Africa: results of a national survey in 2002. *S Afr Med J* 2008; **98**:789-794.
221. Human Sciences Research Council. South African national HIV prevalence, behavioural risks and mass media household survey 2002. 2002. Available: http://www.hsrcpress.ac.za. Accessed 18 Feb 2009
222. Department of Health. South Africa Demographic and Health Survey 2003: Preliminary Report. Pretoria; 2004. Available: http://www.doh.gov.za/docs/reports/2003/sadhs2003/part2.pdf. Accessed 6 Jan 2012
223. Lissouba P, Taljaard D, Rech D, Dermaux-Msimang V, Legeai C, Lewis D, *et al.* Adult male circumcision as an intervention against HIV: an operational study of uptake in a South African community (ANRS 12126). *BMC Infect Dis* 2011; **11**:253.
224. Lagarde E, Taljaard D, Puren A, Rain-Taljaard R, Auvert B. Acceptability of male circumcision as a tool for preventing HIV infection in a highly infected community in South Africa. *AIDS* 2003; **17**:89-95.
225. Weiss HA, Plummer ML, Changalucha J, Mshana G, Shigongo ZS, Todd J, *et al.* Circumcision among adolescent boys in rural northwestern Tanzania. *Trop Med Int Health* 2008; **13**:1054-1061.
226. Lavreys L, Rakwar JP, Thompson ML, Jackson DJ, Mandaliya K, Chohan BH, *et al.* Effect of circumcision on incidence of human immunodeficiency virus type 1 and other sexually transmitted diseases: a prospective cohort study of trucking company employees in Kenya. *J Infect Dis* 1999; **180**:330-336.
227. Thomas AG, Tran BR, Cranston M, Brown MC, Kumar R, Tlelai M. Voluntary medical male circumcision: a cross-sectional study comparing circumcision self-report and physical examination findings in Lesotho. *PLoS One* 2011; **6**:e27561.
228. Urassa M, Todd J, Boerma JT, Hayes R, Isingo R. Male circumcision and susceptibility to HIV infection among men in Tanzania. *AIDS* 1997; **11**:73-79.

229. Marshall E, Rain-Taljaard R, Tsepe M, Monkwe C, Hlatswayo F, Tshabalala S, *et al.* Sequential cross-sectional surveys in Orange Farm, a township of South Africa, revealed a constant low voluntary medical male circumcision uptake among adults despite demand creation campaigns and high acceptability. *PLoS One* 2016; **11**:e0158675.
230. Kikaya V, Skolnik L, Garcia MC, Nkonyana J, Curran K, Ashengo TA. Voluntary medical male circumcision programs can address low HIV testing and counseling usage and ART enrollment among young men: lessons from Lesotho. *PLoS One* 2014; **9**:e83614.
231. Phili R, Abdool Karim Q, Tlou B. Experiences in the implementation of provider-initiated counselling and testing and linkage to HIV services at urban public sector health facilities in KwaZulu-Natal. *Southern African Journal of Infectious Diseases* 2015; **30**:77-81.
232. World Health Organization. Global HIV/AIDS response: Epidemic update and health sector progress towards Universal Access: Progress Report 2011. 2011. Available: http://www.who.int/hiv/pub/progress_report2011/en/index.html. Accessed 28 Sept 2012
233. Department of Health. Annual Performance Plan: 2014/15 - 2016/17. 2014. Available: http://www.health.gov.za/docs/strategic/2013/app201415.pdf. Accessed 20 Aug 2014
234. Department of Health. Annual Report 2013-2014. 2014. Available: http://www.health.gov.za/annualreports.php. Accessed 25 Jan 2015
235. Department of Health. Annual Report 2014/15. Pretoria; 2015. Available: http://www.health.gov.za/index.php/2014-03-17-09-09-38/2014-03-17-09-24-31/category/239-ar2015. Accessed 1 Jan 2016
236. Department of Health. Annual Report 2015/16. Pretoria; 2016. Available: http://www.health.gov.za/index.php/2014-03-17-09-09-38/2014-03-17-09-24-31. Accessed 6 Feb 2017
237. Department of Health. Annual Report 2016/17. 2017. Available: http://www.health.gov.za/index.php/2014-03-17-09-09-38/2014-03-17-09-24-31. Accessed 11 April 2018
238. Department of Health. Annual Report 2017/18. 2018. Available: http://www.health.gov.za/index.php/2014-03-17-09-09-38/2014-03-17-09-24-31. Accessed 29 Nov 2018
239. Department of Health. Annual Report 2018/19. 2019. Available: http://www.health.gov.za/index.php/2014-03-17-09-09-38/2014-03-17-09-24-31#. Accessed 16 March 2020
240. Fonner VA, Dalglish SL, Kennedy CE, Baggaley R, O'Reilly KR, Koechlin FM, *et al.* Effectiveness and safety of oral HIV preexposure prophylaxis for all populations. *AIDS* 2016; **30**:1973-1983.
241. Molina JM, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, *et al.* On-demand preexposure prophylaxis in men at high risk for HIV-1 Infection. *N Engl J Med* 2015; **373**:2237-2246.
242. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, *et al.* Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* 2016; **387**:53-60.

243. Bekker LG, Roux S, Sebastien E, Yola N, Amico KR, Hughes JP, *et al.* Daily and non-daily pre-exposure prophylaxis in African women (HPTN 067/ADAPT Cape Town Trial): a randomised, open-label, phase 2 trial. *Lancet HIV* 2018; **5**:e68-78.
244. Hanscom B, Janes HE, Guarino PD, Huang Y, Brown ER, Chen YQ, *et al.* Preventing HIV-1 infection in women using oral preexposure prophylaxis: a meta-analysis of current evidence. *Journal of Acquired Immune Deficiency Syndrome* 2016; **73**:606-608.
245. Cottrell ML, Yang KH, Prince HM, Sykes C, White N, Malone S, *et al.* A translational pharmacology approach to predicting outcomes of preexposure prophylaxis against HIV in men and women Using tenofovir disoproxil fumarate with or without emtricitabine. *J Infect Dis* 2016; **214**:55-64.
246. Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM, *et al.* Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *N Engl J Med* 2012; **367**:423-434.
247. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, *et al.* Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N Engl J Med* 2012; **367**:399-410.
248. Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S, *et al.* Preexposure prophylaxis for HIV infection among African women. *N Engl J Med* 2012; **367**:411-422.
249. Mugwanya KK, Donnell D, Celum C, Thomas KK, Ndase P, Mugo N, *et al.* Sexual behaviour of heterosexual men and women receiving antiretroviral pre-exposure prophylaxis for HIV prevention: a longitudinal analysis. *Lancet Infect Dis* 2013; **13**:1021-1028.
250. Traeger MW, Schroeder SE, Wright EJ, Hellard ME, Cornelisse VJ, Doyle JS, *et al.* Effects of pre-exposure prophylaxis for the prevention of HIV infection on sexual risk Behavior in men who have sex with men: A systematic review and meta-analysis. *Clin Infect Dis* 2018; **67**:676-686.
251. Liu A, Cohen S, Follansbee S, Cohan D, Weber S, Sachdev D, *et al.* Early experiences implementing pre-exposure prophylaxis (PrEP) for HIV prevention in San Francisco. *PLoS Med* 2014; **11**:e1001613.
252. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, *et al.* Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. *Lancet Infect Dis* 2014; **14**:820-829.
253. Joseph Davey DL, Bekker LG, Gomba Y, Coates T, Myer L, Johnson LF. Modelling the potential impact of providing preexposure prophylaxis in pregnant and breastfeeding women in South Africa. *AIDS* 2019; **33**:1391-1395.
254. Kinuthia J, Pintye J, Mugwanya K, Serede M, Sila J, Abuna F, *et al.* High PrEP uptake among Kenyan pregnant women offered PrEP during antenatal care [Abstract 1047]. *Conference on Retroviruses and Opportunistic Infections*. Boston, USA; 2018.
255. Johnson LF, Stinson K, Newell ML, Bland RM, Moultrie H, Davies MA, *et al.* The contribution of maternal HIV seroconversion during late pregnancy and breastfeeding to mother-to-child transmission of HIV. *J Acquir Immun Defic Syndr* 2012; **59**:417-425.
256. Goga AE, Dinh TH, Jackson DJ, Lombard C, Delaney KP, Puren A, *et al.* First population-level effectiveness evaluation of a national programme to prevent HIV transmission from mother to child, South Africa. *J Epidemiol Community Health* 2015; **69**:240-248.

257. Draper B, Abdullah F. A review of the prevention of mother-to-child transmission programme of the Western Cape provincial government, 2003 - 2004. *S Afr Med J* 2008; **98**:431-434.
258. Johnson LF. A model of paediatric HIV in South Africa. Cape Town: Centre for Infectious Disease Epidemiology and Research, University of Cape Town; 2010. Available: http://webdav.uct.ac.za/depts/epi/publications/documents/Paediatric_HIV_modelling5.pdf. Accessed 23 Feb 2011
259. Tovo PA, Palomba E, Gabiano C, Galli L, de Martino M. Human immunodeficiency virus type 1 (HIV-1) seroconversion during pregnancy does not increase the risk of perinatal transmission. *Br J Obstet Gynaecol* 1991; **98**:940-942.
260. Rudin C, Lauper U, Biedermann K. HIV seroconversion during pregnancy [Abstract W.C.3247]. *7th International AIDS Conference*. Florence, Italy; 1991.
261. Hague RA, Mok JY, Johnstone FD, MacCallum L, Yap PL, Burns SM, *et al.* Maternal factors in HIV transmission. *Int J STD AIDS* 1993; **4**:142-146.
262. Nielsen-Saines K, Melo M, Varella I, Fonseca R, Lira R, Turella ML, *et al.* Primary HIV-1 infection during pregnancy: high rate of HIV-1 MTCT in a cohort of patients in southern Brazil. *Retrovirology* 2008; **5** (Suppl 1):O1.
263. Roongpisuthipong A, Siriwasin W, Simonds RJ, Sangtaweesin V, Vanprapar N, Wasi C, *et al.* HIV seroconversion during pregnancy and risk for mother-to-infant transmission. *J Acquir Immun Defic Syndr* 2001; **26**:348-351.
264. Birkhead GS, Pulver WP, Warren BL, Hackel S, Rodríguez D, Smith L. Acquiring human immunodeficiency virus during pregnancy and mother-to-child transmission in New York: 2002-2006. *Obstet Gynecol* 2010; **115**:1247-1255.
265. Leroy V, Sakarovitch C, Cortina-Borja M, McIntyre J, Coovadia H, Dabis F, *et al.* Is there a difference in the efficacy of peripartum antiretroviral regimens in reducing mother-to-child transmission of HIV in Africa? *AIDS* 2005; **19**:1865-1875.
266. Sperling RS, Shapiro DE, Coombs RW, Todd JA, Herman SA, McSherry GD, *et al.* Maternal viral load, zidovudine treatment, and the risk of transmission of human immunodeficiency virus type 1 from mother to infant. *N Engl J Med* 1996; **335**:1621-1629.
267. Dabis F, Bequet L, Ekouevi DK, Viho I, Rouet F, Horo A, *et al.* Field efficacy of zidovudine, lamivudine and single-dose nevirapine to prevent peripartum HIV transmission. *AIDS* 2005; **19**:309-318.
268. Lallemant M, Jourdain G, Le Coeur S, Mary JY, Ngo-Giang-Huong N, Koetsawang S, *et al.* Single-dose perinatal nevirapine plus standard zidovudine to prevent mother-to-child transmission of HIV-1 in Thailand. *N Engl J Med* 2004; **351**:217-228.
269. Hoffman RM, Black V, Technau K, van der Merwe KJ, Currier J, Coovadia A, *et al.* Effects of highly active antiretroviral therapy duration and regimen on risk for mother-to-child transmission of HIV in Johannesburg, South Africa. *J Acquir Immun Defic Syndr* 2010; **54**:35-41.
270. Bera E, Jwacu N, Pauls F, Mancotywa T, Ngcelwane N, Hlati Y. Risk factors for perinatal HIV-1 transmission in pregnant women requiring lifelong antiretroviral therapy: A longitudinal study at a tertiary hospital in South Africa. *S Afr J Obstet Gynaecol* 2010; **16**:6-13.
271. Homsy J, Bunnell R, Moore D, King R, Malamba S, Nakityo R, *et al.* Reproductive intentions and outcomes among women on antiretroviral therapy in rural Uganda: a prospective cohort study. *PLoS One* 2009; **4**:e4149.

272. Dryden-Peterson S, Jayeoba O, Hughes MD, Jibril H, Keapoletswe K, Tlale J, *et al.* Highly active antiretroviral therapy versus zidovudine for prevention of mother-to-child transmission in a programmatic setting, Botswana. *J Acquir Immun Defic Syndr* 2011; **58**:353-357.
273. Kim MH, Ahmed S, Preidis GA, Abrams EJ, Hosseinipour MC, Giordano TP, *et al.* Low rates of mother-to-child HIV transmission in a routine programmatic setting in Lilongwe, Malawi. *PLoS One* 2013; **8**:e64979.
274. Gibb DM, Kizito H, Russell EC, Chidziva E, Zalwango E, Nalumenya R, *et al.* Pregnancy and infant outcomes among HIV-infected women taking long-term ART with and without tenofovir in the DART trial. *PLoS Med* 2012; **9**:e1001217.
275. Breastfeeding and HIV International Transmission Study Group. Late postnatal transmission of HIV-1 in breast-fed children: an individual patient data meta-analysis. *J Infect Dis* 2004; **189**:2154-2166.
276. Coovadia HM, Rollins NC, Bland RM, Little K, Coutsoudis A, Bennish ML, *et al.* Mother-to-child transmission of HIV-1 infection during exclusive breastfeeding in the first 6 months of life: an intervention cohort study. *Lancet* 2007; **369**:1107-1116.
277. Becquet R, Bland R, Leroy V, Rollins NC, Ekouevi DK, Coutsoudis A, *et al.* Duration, pattern of breastfeeding and postnatal transmission of HIV: pooled analysis of individual data from West and South African cohorts. *PLoS One* 2009; **4**:e7397.
278. Kumwenda NI, Hoover DR, Mofenson LM, Thigpen MC, Kafulafula G, Li Q, *et al.* Extended antiretroviral prophylaxis to reduce breast-milk HIV-1 transmission. *N Engl J Med* 2008; **359**:119-129.
279. Chasela CS, Hudgens MG, Jamieson DJ, Kayira D, Hosseinipour MC, Kourtis AP, *et al.* Maternal or infant antiretroviral drugs to reduce HIV-1 transmission. *N Engl J Med* 2010; **362**:2271-2281.
280. Six Week Extended-dose Nevirapine Study Team. Extended-dose nevirapine to 6 weeks of age for infants to prevent HIV transmission via breastfeeding in Ethiopia, India, and Uganda: an analysis of three randomised controlled trials. *Lancet* 2008; **372**:300-313.
281. Peltier CA, Ndayisaba GF, Lepage P, van Griensven J, Leroy V, Pharm CO, *et al.* Breastfeeding with maternal antiretroviral therapy or formula feeding to prevent HIV postnatal mother-to-child transmission in Rwanda. *AIDS* 2009; **23**:2415-2423.
282. Tonwe-Gold B, Ekouevi DK, Viho I, Amani-Bosse C, Toure S, Coffie PA, *et al.* Antiretroviral treatment and prevention of peripartum and postnatal HIV transmission in West Africa: evaluation of a two-tiered approach. *PLoS Med* 2007; **4**:e257.
283. Palombi L, Marazzi MC, Voetberg A, Magid NA. Treatment acceleration program and the experience of the DREAM program in prevention of mother-to-child transmission of HIV. *AIDS* 2007; **21 (Suppl 4)**:S65-71.
284. Kilewo C, Karlsson K, Ngarina M, Massawe A, Lyamuya E, Swai A, *et al.* Prevention of mother-to-child transmission of HIV-1 through breastfeeding by treating mothers with triple antiretroviral therapy in Dar es Salaam, Tanzania: the Mitra Plus study. *J Acquir Immun Defic Syndr* 2009; **52**:406-416.
285. Shapiro RL, Hughes MD, Ogwu A, Kitch D, Lockman S, Moffat C, *et al.* Antiretroviral regimens in pregnancy and breast-feeding in Botswana. *N Engl J Med* 2010; **362**:2282-2294.
286. Kesho Bora Study Group. Triple antiretroviral compared with zidovudine and single-dose nevirapine prophylaxis during pregnancy and breastfeeding for prevention of mother-to-child transmission of HIV-1 (Kesho Bora study): a randomised controlled trial. *Lancet Infect Dis* 2011; **11**:171-180.

287. Giuliano M, Andreotti M, Liotta G, Jere H, Sagno JB, Maulidi M, *et al.* Maternal antiretroviral therapy for the prevention of mother-to-child transmission of HIV in Malawi: maternal and infant outcomes two years after delivery. *PLoS One* 2013; **8**:e68950.
288. Thomas TK, Masaba R, Borkowf CB, Ndivo R, Zeh C, Misore A, *et al.* Triple-antiretroviral prophylaxis to prevent mother-to-child HIV transmission through breastfeeding - the Kisumu Breastfeeding Study, Kenya: a clinical trial. *PLoS Med* 2011; **8**:e1001015.
289. Ngoma MS, Misir A, Mutale W, Rampakakis E, Sampalis JS, Elong A, *et al.* Efficacy of WHO recommendation for continued breastfeeding and maternal cART for prevention of perinatal and postnatal HIV transmission in Zambia. *J Int AIDS Soc* 2015; **18**:19352.
290. Cohan D, Natureeba P, Koss CA, Plenty A, Luwedde F, Mwesigwa J, *et al.* Efficacy and safety of lopinavir/ritonavir versus efavirenz-based antiretroviral therapy in HIV-infected pregnant Ugandan women. *AIDS* 2015; **29**:183-191.
291. Coovadia HM, Brown ER, Fowler MG, Chipato T, Moodley D, Manji K, *et al.* Efficacy and safety of an extended nevirapine regimen in infant children of breastfeeding mothers with HIV-1 infection for prevention of postnatal HIV-1 transmission (HPTN 046): a randomised, double-blind, placebo-controlled trial. *Lancet* 2012; **379**:221-228.
292. Fitzgerald FC, Bekker LG, Kaplan R, Myer L, Lawn SD, Wood R. Mother-to-child transmission of HIV in a community-based antiretroviral clinic in South Africa. *S Afr Med J* 2010; **100**:827-831.
293. Van Schalkwyk M, Andersson MI, Zeier MD, La Grange M, Taljaard JJ, Theron GB. The impact of revised PMTCT guidelines: a view from a public sector ARV clinic in Cape Town, South Africa. *J Acquir Immun Defic Syndr* 2013; **63**:234-238.
294. Geddes R, Giddy J, Butler LM, Van Wyk E, Crankshaw T, Esterhuizen TM, *et al.* Dual and triple therapy to prevent mother-to-child transmission of HIV in a resource-limited setting - lessons from a South African programme. *S Afr Med J* 2011; **101**:651-654.
295. Kim AA, Hallett T, Stover J, Gouws E, Musinguzi J, Mureithi PK, *et al.* Estimating HIV incidence among adults in Kenya and Uganda: a systematic comparison of multiple methods. *PLoS One* 2011; **6**:e17535.
296. Department of Health. Clinical Guidelines: PMTCT (Prevention of Mother-to-Child Transmission). 2010. Available: http://www.rhru.co.za/Resources/Documents/2010%20PMTCT%20Guidelines.pdf. Accessed 7 June 2010
297. Fatti G, Shaikh N, Eley B, Jackson D, Grimwood A. Adolescent and young pregnant women at increased risk of mother-to-child transmission of HIV and poorer maternal and infant health outcomes: a cohort study at public health facilities in the Nelson Mandela Bay Metropolitan district, Eastern Cape, South Africa. *S Afr Med J* 2014; **104**:874-880.
298. Stinson K, Jennings K, Myer L. Integration of antiretroviral therapy services into antenatal care increases treatment initiation during pregnancy: a cohort study. *PLoS One* 2013; **8**:e63328.
299. Moodley D, Esterhuizen TM, Pather T, Chetty V, Ngaleka L. High HIV incidence during pregnancy: compelling reason for repeat HIV testing. *AIDS* 2009; **23**:1255-1259.

300. Jackson DJ, Chopra M, Doherty TM, Colvin MS, Levin JB, Willumsen JF, *et al.* Operational effectiveness and 36 week HIV-free survival in the South African programme to prevent mother-to-child transmission of HIV-1. *AIDS* 2007; **21**:509-516.
301. Lindbäck S, Thorstensson R, Karlsson A, von Sydow M, Flamholc L, Blaxhult A, *et al.* Diagnosis of primary HIV-1 infection and duration of follow-up after HIV exposure. *AIDS* 2000; **14**:2333-2339.
302. Orie EF, Songca PP, Moodley J. An audit of PMTCT services at a regional hospital in South Africa. *S Afr Fam Pract* 2009; **51**:492-495.
303. Grimwood A, Fatti G, Mothibi E, Eley B, Jackson D. Progress of preventing mother-to-child transmission of HIV at primary healthcare facilities and district hospitals in three South African provinces. *S Afr Med J* 2012; **102**:81-83.
304. Bland RM, Rollins NC, Coutsoudis A, Coovadia HM. Breastfeeding practices in an area of high HIV prevalence in rural South Africa. *Acta Paediatr* 2002; **91**:704-711.
305. Coutsoudis A, Pillay K, Kuhn L, Spooner E, Tsai WY, Coovadia HM. Method of feeding and transmission of HIV-1 from mothers to children by 15 months of age: prospective cohort study from Durban, South Africa. *AIDS* 2001; **15**:379-387.
306. Goga AE, Van Wyk B, Doherty T, Colvin M, Jackson DJ, Chopra M. Operational effectiveness of guidelines on complete breast-feeding cessation to reduce mother-to-child transmission of HIV: results from a prospective observational cohort study at routine prevention of mother-to-child transmission sites, South Africa. *J Acquir Immun Defic Syndr* 2009; **50**:521-528.
307. Doherty T, Chopra M, Jackson D, Goga A, Colvin M, Persson LA. Effectiveness of the WHO/UNICEF guidelines on infant feeding for HIV-positive women: results from a prospective cohort study in South Africa. *AIDS* 2007; **21**:1791-1797.
308. National Breastfeeding Consultative Group. The Tshwane declaration of support for breastfeeding in South Africa. *S Afr J Clin Nutr* 2011; **24**:214.
309. Goga AE, Jackson DJ, Singh M, Lombard C. Early (4-8 weeks postpartum) population-level effectiveness of WHO PMTCT option A, South Africa, 2012-2013. South African Medical Research Council and National Department of Health of South Africa; 2015.
310. Johnson LF, Davies MA, Moultrie H, Sherman GG, Bland RM, Rehle TM, *et al.* The effect of early initiation of antiretroviral treatment in infants on pediatric AIDS mortality in South Africa: a model-based analysis. *Pediatr Infect Dis J* 2012; **31**:474-480.
311. World Health Organization. Antiretroviral therapy for HIV infection in infants and children: towards universal access. Recommendations for a public health approach, 2006. Geneva; 2007. Available: http://www.who.int/hiv/pub/guidelines/art/en/ Accessed 26 Jan 2009
312. Marston M, Becquet R, Zaba B, Moulton LH, Gray G, Coovadia H, *et al.* Net survival of perinatally and postnatally HIV-infected children: a pooled analysis of individual data from sub-Saharan Africa. *Int J Epidemiol* 2011; **40**:385-396.
313. Charlebois ED, Ruel TD, Gasasira AF, Achan J, Kateera F, Akello C, *et al.* Short-term risk of HIV disease progression and death in Ugandan children not eligible for antiretroviral therapy. *J Acquir Immun Defic Syndr* 2010; **55**:330-335.
314. Violari A, Cotton MF, Gibb DM, Babiker AG, Steyn J, Madhi SA, *et al.* Early antiretroviral therapy and mortality among HIV-infected infants. *N Engl J Med* 2008; **359**:2233-2244.

315. Mphatswe W, Blanckenberg N, Tudor-Williams G, Prendergast A, Thobakgale C, Mkhwanazi N, *et al.* High frequency of rapid immunological progression in African infants infected in the era of perinatal HIV prophylaxis. *AIDS* 2007; **21**:1253-1261.
316. Diaz C, Hanson C, Cooper ER, Read JS, Watson J, Mendez HA, *et al.* Disease progression in a cohort of infants with vertically acquired HIV infection observed from birth: the Women and Infants Transmission Study (WITS). *J Acquir Immun Defic Syndr* 1998; **18**:221-228.
317. Blanche S, Newell ML, Mayaux MJ, Dunn DT, Teglas JP, Rouzioux C, *et al.* Morbidity and mortality in European children vertically infected by HIV-1. *J Acquir Immun Defic Syndr* 1997; **14**:442-450.
318. Marinda E, Humphrey JH, Iliff PJ, Mutasa K, Nathoo KJ, Piwoz EG, *et al.* Child mortality according to maternal and infant HIV status in Zimbabwe. *Pediatr Infect Dis J* 2007; **26**:519-526.
319. Fox MP, Brooks D, Kuhn L, Aldrovandi G, Sinkala M, Kankasa C, *et al.* Reduced mortality associated with breast-feeding-acquired HIV infection and breast-feeding among HIV-infected children in Zambia. *J Acquir Immun Defic Syndr* 2008; **48**:90-96.
320. Sutcliffe CG, Scott S, Mugala N, Ndhlovu Z, Monze M, Quinn TC, *et al.* Survival from 9 months of age among HIV-infected and uninfected Zambian children prior to the availability of antiretroviral therapy. *Clin Infect Dis* 2008; **47**:837-844.
321. Newell ML, Coovadia H, Cortina-Borja M, Rollins N, Gaillard P, Dabis F. Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: a pooled analysis. *Lancet* 2004; **364**:1236-1243.
322. Cross Continents Collaboration for Kids. Markers for predicting mortality in untreated HIV-infected children in resource-limited settings: a meta-analysis. *AIDS* 2008; **22**:97-105.
323. Hussey GD, Reijnhart RM, Sebens AM, Burgess J, Schaaf S, Potgieter S. Survival of children in Cape Town known to be vertically infected with HIV-1. *S Afr Med J* 1998; **88**:554-558.
324. Davies MA, Keiser O, Technau K, Eley B, Rabie H, Van Cutsem G, *et al.* Outcomes of the South African national antiretroviral treatment programme for children: the IeDEA Southern Africa collaboration. *S Afr Med J* 2009; **99**:730-737.
325. Massyn N, Day C, Peer N, Padarath A, Barron P, English R. District Health Barometer 2013/14. Durban; 2014.
326. Fadnes LT, Jackson D, Engebretsen IM, Zembe W, Sanders D, Sommerfelt H, *et al.* Vaccination coverage and timeliness in three South African areas: a prospective study. *BMC Public Health* 2011; **11**:404.
327. Fonn S, Sartorius B, Levin J, Likibi ML. Immunisation coverage estimates by cluster sampling survey of children (aged 12-23 months) in Gauteng province, 2003. *South Afr J Epidemiol Infect* 2006; **21**:164-169.
328. Rollins N, Mzolo S, Moodley T, Esterhuizen T, van Rooyen H. Universal HIV testing of infants at immunization clinics: an acceptable and feasible approach for early infant diagnosis in high HIV prevalence settings. *AIDS* 2009; **23**:1851-1857.
329. Smith SJ, Nimmo C, Fredlund V, Moodley P. Early infant diagnosis of HIV and fast initiation of anti-retroviral therapy in a rural African setting: how well are we doing? *Paediatr Int Child Health* 2014; **34**:203-207.
330. Hsiao NY, Stinson K, Myer L. Linkage of HIV-infected infants from diagnosis to antiretroviral therapy services across the Western Cape, South Africa. *PLoS One* 2013; **8**:e55308.

331. Kalk E, Kroon M, Boulle A, Osler M, Euvrard J, Stinson K, *et al.* Neonatal and infant diagnostic HIV-PCR uptake and associations during three sequential policy periods in Cape Town, South Africa: a longitudinal analysis. *J Int AIDS Soc* 2018; **21**:e25212.
332. Moyo F, Mazanderani A, Barron P, Bhardwaj S, Goga AE, Pillay Y, *et al.* Introduction of routine HIV birth testing in the South African national consolidated guidelines. *Pediatr Infect Dis J* 2018; **37**:559-563.
333. Dunning L, Kroon M, Fourie L, Ciaranello A, Myer L. Impact of birth HIV-PCR testing on the uptake of follow-up early infant diagnosis services in Cape Town, South Africa. *Pediatr Infect Dis J* 2017; **36**:1159-1164.
334. Lilian RR, Johnson LF, Moolla H, Sherman GG. A mathematical model evaluating the timing of early diagnostic testing in HIV-exposed infants in South Africa. *Journal of Acquired Immune Deficiency Syndrome* 2014; **67**:341-348.
335. Department of Health. Guidelines for the management of HIV in children. 2nd ed; 2010. Available: http://www.doh.gov.za/docs/index.html. Accessed 6 March 2011
336. World Health Organization. Report of the WHO Technical Reference Group, Paediatric HIV/ART Care Guideline Group Meeting, WHO Headquarters, Geneva, Switzerland, 10-11 April 2008. 2008. Available: http://www.who.int/hiv/pub/paediatric/WHO_Paediatric_ART_guideline_rev_mreport_2008.pdf. Accessed 6 Jan 2009
337. Lilian RR, Kalk E, Technau KG, Sherman GG. Birth diagnosis of HIV infection in infants to reduce infant mortality and monitor for elimination of mother-to-child transmission. *Pediatr Infect Dis J* 2013; **32**:1080-1085.
338. Department of Health. Policy and Guidelines for the Implementation of the PMTCT programme. 2008. Available: http://www.doh.gov.za/docs/policy/2008/pmtct.pdf. Accessed 6 Jan 2012
339. Actuarial Society of South Africa. ASSA2008 AIDS and Demographic Model. 2011. Available: http://aids.actuarialsociety.org.za. Accessed 5 April 2011
340. Dorrington RE, Moultrie TA. The age distribution of younger people in South Africa: What can it tell us about recent fertility trends? *7th African Population Conference*. Johannesburg, South Africa; 2015.
341. Sungkanuparph S, Kiertiburanakul S, Apisarnthanarak A, Malathum K, Watcharananan S, Sathapatayavongs B. Rapid CD4 decline after interruption of non-nucleoside reverse transcriptase inhibitor-based antiretroviral therapy in a resource-limited setting. *AIDS Res Ther* 2007; **4**:26.
342. Hoffman IF, Martinson FE, Powers KA, Chilongozi DA, Msiska ED, Kachipapa EI, *et al.* The year-long effect of HIV-positive test results on pregnancy intentions, contraceptive use, and pregnancy incidence among Malawian women. *Journal of Acquired Immune Deficiency Syndrome* 2008; **47**:477-483.
343. Kaida A, Laher F, Strathdee SA, Janssen PA, Money D, Hogg RS, *et al.* Childbearing intentions of HIV-positive women of reproductive age in Soweto, South Africa: the influence of expanding access to HAART in an HIV hyperendemic setting. *Am J Public Health* 2011; **101**:350-358.
344. Taulo F, Berry M, Tsui A, Makanani B, Kafulafula G, Li Q, *et al.* Fertility intentions of HIV-1 infected and uninfected women in Malawi: a longitudinal study. *AIDS Behav* 2009; **13** (Suppl 1):20-27.
345. Mbita G, Renju J, Lija G, Conserve DF, Todd J. Effect of antiretroviral therapy on fertility rate among women living with HIV in Tabora, Tanzania: An historical cohort study. *PLoS One* 2019; **14**:e0222173.

346. Patel RC, Onono M, Gandhi M, Blat C, Hagey J, Shade SB, *et al.* Pregnancy rates in HIV-positive women using contraceptives and efavirenz-based or nevirapine-based antiretroviral therapy in Kenya: a retrospective cohort study. *Lancet HIV* 2015; **2**:e474-482.
347. Elul B, Wools-Kaloustian KK, Wu Y, Musick BS, Nuwagaba-Biribonwoha H, Nash D, *et al.* Untangling the relationship between antiretroviral therapy use and incident pregnancy: a marginal structural model analysis using data from 47,313 HIV-positive women in East Africa. *Journal of Acquired Immune Deficiency Syndrome* 2016; **72**:324-332.
348. Msemburi W, Pillay-van Wyk V, Dorrington RE, Neethling I, Nannan N, Groenewald P, *et al.* Second national burden of disease study for South Africa: Cause-of-death profile for South Africa, 1997-2010. Cape Town: South African Medical Research Council; 2014.
349. Dorrington RE. Alternative South African mid-year estimates, 2013. Centre for Actuarial Research; 2013. Available: http://www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/Mono13.pdf. Accessed 19 Nov 2013
350. Statistics South Africa. Mid-year population estimates: 2013. Pretoria; 2013. Available: http://www.statssa.gov.za/publications/P0302/P03022013.pdf. Accessed 9 July 2015
351. Küstner H, Swanevelder J, van Middelkoop A. National HIV surveillance - South Africa, 1990 - 1992. *S Afr Med J* 1994; **84**:195-199.
352. Montana LS, Mishra V, Hong R. Comparison of HIV prevalence estimates from antenatal care surveillance and population-based surveys in sub-Saharan Africa. *Sex Transm Infect* 2008; **84 (Suppl 1)**:i78-i84.
353. Williams BG, Gouws E, Colvin M, Sitas F, Ramjee G, Abdool Karim SS. Patterns of infection: using age prevalence data to understand the epidemic of HIV in South Africa. *S Afr J Sci* 2000; **96**:305-312.
354. Morgan M, Walker N, Gouws E, Stanecki KA, Stover J. Improved plausibility bounds about the 2005 HIV and AIDS estimates. *Sex Transm Infect* 2006; **82** (suppl 3):iii71-77.
355. Alkema L, Raftery AE, Brown T. Bayesian melding for estimating uncertainty in national HIV prevalence estimates. *Sex Transm Infect* 2008; **84** (Suppl 1):i11-16.
356. Woldesenbet SA, Kufa T, Barron P, Chirombo BC, Cheyip M, Ayalew K, *et al.* Viral suppression and factors associated with failure to achieve viral suppression among pregnant women in South Africa: a national cross-sectional survey. *AIDS* 2019; [In press].
357. Grobler A, Cawood C, Khanyile D, Puren A, Kharsany ABM. Progress of UNAIDS 90-90-90 targets in a district in KwaZulu-Natal, South Africa, with high HIV burden, in the HIPSS study: a household-based complex multilevel community survey. *Lancet HIV* 2017; **4**:e505-e513.
358. Rohr JK, Xavier Gomez-Olive F, Rosenberg M, Manne-Goehler J, Geldsetzer P, Wagner RG, *et al.* Performance of self-reported HIV status in determining true HIV status among older adults in rural South Africa: a validation study. *J Int AIDS Soc* 2017; **20**:21691.
359. Sandfort TGM, Dominguez K, Kayange N, Ogendo A, Panchia R, Chen YQ, *et al.* HIV testing and the HIV care continuum among sub-Saharan African men who have sex with men and transgender women screened for participation in HPTN 075. *PLoS One* 2019; **14**:e0217501.

360. Huerga H, Shiferie F, Grebe E, Giuliani R, Farhat JB, Van Cutsem G, *et al.* A comparison of self-report and antiretroviral detection to inform estimates of antiretroviral therapy coverage, viral load suppression and HIV incidence in Kwazulu-Natal, South Africa. *BMC Infect Dis* 2017; **17**:653.
361. Statistics South Africa. Mortality and causes of death in South Africa, 2016: Findings from death notification. 2018. Available: http://www.statssa.gov.za/publications/P03093/P030932016.pdf. Accessed 18 Dec 2018
362. Groenewald P, Nannan N, Bourne D, Laubscher R, Bradshaw D. Identifying deaths from AIDS in South Africa. *AIDS* 2005; **19**:193-201.
363. Dorrington RE, Moultrie TA, Timæus IM. Estimation of mortality using the South African Census 2001 data. Centre for Actuarial Research; 2004. Available: http://www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/Mono11.pdf. Accessed 3 Dec 2015
364. Statistics South Africa. Mortality and causes of death in South Africa, 2011: Findings from death notification. Pretoria; 2014. Available: http://beta2.statssa.gov.za/publications/P03093/P030932011.pdf. Accessed 15 April 2014
365. Dorrington RE, Bradshaw D. Maternal mortality in South Africa: lessons from a case study in the use of deaths reported by households in censuses and surveys. *J Pop Res* 2011; **28**:49-73.
366. Dorrington R, Bradshaw D, Laubscher R. Rapid Mortality Surveillance Report 2012. Cape Town; 2014. Available: http://www.mrc.ac.za/bod/RapidMortalitySurveillanceReport2012.pdf. Accessed 21 April 2014
367. Johnson LF, Dorrington RE, Laubscher R, Hoffmann CJ, Wood R, Fox MP, *et al.* A comparison of death recording by health centres and civil registration in South Africans receiving antiretroviral treatment. *J Int AIDS Soc* 2015; **18**:20628.
368. Raftery AE, Bao L. Estimating and projecting trends in HIV/AIDS generalized epidemics using Incremental Mixture Importance Sampling. *Biometrics* 2010; **66**:1162-1173.
369. Rehle T, Johnson L, Hallett T, Mahy M, Kim A, Odido H, *et al.* A comparison of South African national HIV incidence estimates: a critical appraisal of different methods. *PLoS One* 2015; **10**:e0133255.
370. Massyn N, Peer N, Padarath A, Barron P, Day C. District Health Barometer 2014/15. Durban; 2015. Available: http://www.hst.org.za/sites/default/files/Complete_DHB_2014_15_linked.pdf. Accessed 5 March 2017
371. Goga A, Jackson D, Lombard C, Ramokolo V, Ngandu N, Sherman G, *et al.* Highest risk of mother to child transmission of HIV or death in the first 6 months postpartum: results from 18 month follow-up of an HIV-exposed national cohort, South Africa [Abstract TUAE0106]. *21st International AIDS Conference*. Durban, South Africa; 2016.
372. Feucht UD, Forsyth B, Kruger M. False-positive HIV DNA PCR testing of infants: implications in a changing epidemic. *S Afr Med J* 2012; **102**:149-152.
373. van Lettow M, Landes M, van Oosterhout JJ, Schouten E, Phiri H, Nkhoma E, *et al.* Prevention of mother-to-child transmission of HIV: a cross-sectional study in Malawi. *Bull WHO* 2018; **96**:256-265.

374. Fox M, Bor J, MacLeod W, Maskew M, Brennan A, Stevens W, *et al.* Is retention on ART underestimated due to patient transfers? Estimating system-wide retention using a national labs database in South Africa. *International AIDS Conference*. Durban, South Africa; 2016.
375. Bongaarts J. A framework for analyzing the proximate determinants of fertility. *Pop Dev Rev* 1978; **4**:105-132.
376. Hermans LE, Carmona S, Nijhuis M, Tempelman HA, Richman DD, Moorhouse M, *et al.* Virological suppression and clinical management in response to viremia in South African HIV treatment program: A multicenter cohort study. *PLoS Med* 2020; **17**:e1003037.
377. Joseph Davey D, Abrahams Z, Feinberg M, Prins M, Serrao C, Medeossi B, *et al.* Factors associated with recent unsuppressed viral load in HIV-1-infected patients in care on first-line antiretroviral therapy in South Africa. *Int J STD AIDS* 2018; **29**:603-610.
378. Green D, Tordoff DM, Kharono B, Akullian A, Bershteyn A, Morrison M, *et al.* Evidence of sociodemographic heterogeneity across the HIV treatment cascade and progress towards 90-90-90 in sub-Saharan Africa – a systematic review and meta-analysis. *J Int AIDS Soc* 2020; **23**:e25470.
379. Brittain K, Myer L, Phillips N, Cluver LD, Zar HJ, Stein DJ, *et al.* Behavioural health risks during early adolescence among perinatally HIV-infected South African adolescents and same-age, HIV-uninfected peers. *AIDS Care* 2019; **31**:131-140.
380. Judd A, Foster C, Thompson LC, Sturgeon K, Le Prevost M, Jungmann E, *et al.* Sexual health of young people with perinatal HIV and HIV negative young people in England. *PLoS One* 2018; **13**:e0205597.
381. Johnson LF, van Rensburg C, Govathson C, Meyer-Rath G. Optimal HIV testing strategies for South Africa: a model-based evaluation of population-level impact and cost-effectiveness. *Sci Rep* 2019; **9**:12621.
382. Sandfort TG, Lane T, Dolezal C, Reddy V. Gender expression and risk of HIV infection among black South African men who have sex with men. *AIDS Behav* 2015; **19**:2270-2279.
383. Council for Medical Schemes. Annual Report 2016/2017. 2017. Available: https://www.medicalschemes.com/Publications.aspx. Accessed 8 Feb 2018
384. Mkwanazi NB, Patel D, Newell ML, Rollins NC, Coutsoudis A, Coovadia HM, *et al.* Rapid testing may not improve uptake of HIV testing and same day results in a rural South African community: a cohort study of 12,000 women. *PLoS One* 2008; **3**:e3501.
385. Moodley D, Moodley P, Ndabandaba T, Esterhuizen T. Reliability of HIV rapid tests is user dependent. *S Afr Med J* 2008; **98**:707-709.
386. Mwisongo A, Peltzer K, Mohlabane N, Tutshana B. The quality of rapid HIV testing in South Africa: an assessment of testers' compliance. *Afr Health Sci* 2016; **16**:646-654.
387. Bock P, Phiri C, Piwowar-Manning E, Kosloff B, Mandla N, Young A, *et al.* Understanding low sensitivity of community-based HIV rapid testing: experiences from the HPTN 071 (PopART) trial in Zambia and South Africa. *J Int AIDS Soc* 2017; **20**:21780.
388. Jackson D, Naik R, Tabana H, Pillay M, Madurai S, Zembe W, *et al.* Quality of home-based rapid HIV testing by community lay counsellors in a rural district of South Africa. *J Int AIDS Soc* 2013; **16**:18744.

389. Wolpaw BJ, Mathews C, Chopra M, Hardie D, de Azevedo V, Jennings K, *et al.* The failure of routine rapid HIV testing: a case study of improving low sensitivity in the field. *BMC Health Serv Res* 2010; **10**:73.
390. Kufa T, Kharsany AB, Cawood C, Khanyile D, Lewis L, Grobler A, *et al.* Misdiagnosis of HIV infection during a South African community-based survey: implications for rapid HIV testing. *J Int AIDS Soc* 2017; **20** (Suppl 6):21753.
391. Bassett IV, Chetty S, Giddy J, Reddy S, Bishop K, Lu Z, *et al.* Screening for acute HIV infection in South Africa: finding acute and chronic disease. *HIV Med* 2011; **12**:46-53.
392. Gray RH, Makumbi F, Serwadda D, Lutalo T, Nalugoda F, Opendi P, *et al.* Limitations of rapid HIV-1 tests during screening for trials in Uganda: diagnostic test accuracy study. *BMJ* 2007; **335**:188.
393. Shanks L, Klarkowski D, O'Brien DP. False positive HIV diagnoses in resource limited settings: operational lessons learned for HIV programmes. *PLoS One* 2013; **8**:e59906.
394. Bruzzone B, Bisio F, Ventura A, Nigro N, Miguel LM, Mayinda Mboungou FA, *et al.* HIV serological screening in a population of pregnant women in the Republic of Congo: suitability of different assays. *Trop Med Int Health* 2008; **13**:900-903.
395. Urassa W, Nozohoor S, Jaffer S, Karama K, Mhalu F, Biberfeld G. Evaluation of an alternative confirmatory strategy for the diagnosis of HIV infection in Dar Es Salaam, Tanzania, based on simple rapid assays. *J Virol Methods* 2002; **100**:115-120.
396. Wilkinson D. HIV infection among pregnant women in the South African private medical sector. *AIDS* 1999; **13**:1783.
397. Johnson LF, Alkema L, Dorrington RE. A Bayesian approach to uncertainty analysis of sexually transmitted infection models. *Sex Transm Infect* 2010; **86**:169-174.
398. Peltzer K, Matseke G, Mzolo T, Majaja M. Determinants of knowledge of HIV status in South Africa: results from a population-based HIV survey. *BMC Public Health* 2009; **9**:174.
399. Tan WS, Chow EP, Fairley CK, Chen MY, Bradshaw CS, Read TR. Sensitivity of HIV rapid tests compared with fourth-generation enzyme immunoassays or HIV RNA tests. *AIDS* 2016; **30**:1951-1960.
400. Ng'ang'a A, Waruiru W, Ngare C, Ssempijja V, Gachuki T, Njoroge I, *et al.* The status of HIV testing and counseling in Kenya: Results from a nationally representative population-based survey. *J Acquir Immun Defic Syndr* 2014; **66** (Suppl 1):S27-36.
401. Reniers G, Eaton J. Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys. *AIDS* 2009; **23**:621-629.
402. Floyd S, Molesworth A, Dube A, Crampin AC, Houben R, Chihana M, *et al.* Underestimation of HIV prevalence in surveys when some people already know their status, and ways to reduce the bias. *AIDS* 2013; **27**:233-242.
403. Chanda MM, Ortblad KF, Mwale M, Chongo S, Kanchele C, Kamungoma N, *et al.* HIV self-testing among female sex workers in Zambia: A cluster randomized controlled trial. *PLoS Med* 2017; **14**:e1002442.
404. Gilbert PB, McKeague IW, Eisen G, Mullins C, Guéye-NDiaye A, Mboup S, *et al.* Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal. *Stat Med* 2003; **22**:573-593.
405. Chen L, Jha P, Stirling B, Sgaier SK, Daid T, Kaul R, *et al.* Sexual Risk Factors for HIV Infection in Early and Advanced HIV Epidemics in Sub-Saharan Africa: Systematic Overview of 68 Epidemiological Studies. *PLoS One* 2007; **2**:e1001.

406. Johnson LF, Dorrington RE, Bradshaw D, Coetzee DJ. The effect of syndromic management interventions on the prevalence of sexually transmitted infections in South Africa. *Sex Reprod Healthc* 2011; **2**:13-20.
407. Nagelkerke N, Abu-Raddad LJ, Awad SF, Black V, Williams B. A signature for biological heterogeneity in susceptibility to HIV infection? *Infectious Disease Modelling* 2018; **3**:139-144.
408. Schwartz S, Lambert A, Phaswana-Mafuya N, Kose Z, McIngana M, Holland C, *et al.* Engagement in the HIV care cascade and barriers to antiretroviral therapy uptake among female sex workers in Port Elizabeth, South Africa: findings from a respondent-driven sampling study. *Sex Transm Infect* 2017; **93**:290-296.
409. Poliah V, Paruk S. Depression, anxiety symptoms and substance use amongst sex workers attending a non-governmental organisation in KwaZulu-Natal, South Africa. *S Afr Fam Pract* 2017; **59**:116-122.
410. Slabbert M, Venter F, Gay C, Roelofsen C, Lalla-Edward S, Rees H. Sexual and reproductive health outcomes among female sex workers in Johannesburg and Pretoria, South Africa: Recommendations for public health programmes. *BMC Public Health* 2017; **17** (Suppl 3):442.
411. Leggett T. Drugs, sex work, and HIV in three South Africa cities. *Society in Transition* 2001; **32**:101-109.
412. Ndhlovu L, Searle C, Van Dam J, Mzaidume Y, Rasego B, Moema S. Reducing the transmission of HIV and sexually transmitted infections in a mining community: Findings from the Carletonville Mothusimpilo intervention project: 1998 to 2001. Washington, DC: Population Council; 2005. Available: http://www.popcouncil.org/pdfs/horizons/crltnvll.pdf. Accessed 1 Dec 2006
413. Greener R, Rambally L, Lafort Y, Beksinksa M, Drace M, Smit J. Condom use and access amongst female sex workers in eThekwini District, KwaZulu-Natal [Abstract MOPE266]. *20th International AIDS Conference*. Melbourne, Australia; 2014.
414. Black V, Williams BG, Maseko V, Radebe F, Rees HV, Lewis DA. Field evaluation of Standard Diagnostics' Bioline HIV/Syphilis Duo test among female sex workers in Johannesburg, South Africa. *Sex Transm Infect* 2016; **92**:495-498.
415. Johnson LF, Coetzee DJ, Dorrington RE. Sentinel surveillance of sexually transmitted infections in South Africa: a review. *Sex Transm Infect* 2005; **81**:287-293.
416. Johnson LF, Mulongeni P, Marr A, Lane T. Age bias in survey sampling and implications for estimating HIV prevalence in men who have sex with men: insights from mathematical modelling. *Epidemiol Infect* 2018; **146**:1036-1042.
417. Rispel LC, Metcalf CA, Cloete A, Reddy V, Lombard C. HIV prevalence and risk practices among men who have sex with men in two South African cities. *J Acquir Immun Defic Syndr* 2011; **57**:69-76.
418. Tucker A, Liht J, de Swardt G, Jobson G, Rebe K, McIntyre J, *et al.* An exploration into the role of depression and self-efficacy on township men who have sex with men's ability to engage in safer sexual practices. *AIDS Care* 2013; **25**:1227-1235.
419. Kufa T, Lane T, Manyuchi A, Singh B, Isdahl Z, Osmand T, *et al.* The accuracy of HIV rapid testing in integrated bio-behavioral surveys of men who have sex with men across 5 Provinces in South Africa. *Medicine* 2017; **96**:e7391.
420. Fearon E, Tenza S, Moodley K, Mokoena C, Smith AD, Bourne A, *et al.* A population-based assessment of HIV prevalence and engagement in the HIV care cascade amongst gay, bisexual and other men who have sex with men and transgender women (MSM/TG) in Johannesburg. 2018.

421. Schoub BD, Lyons SF, McGillivray GM, Smith AN, Johnson S, Fisher EL. Absence of HIV infection in prostitutes and women attending sexually-transmitted disease clinics in South Africa. *Trans R Soc Trop Med Hyg* 1987; **81**:874-875.
422. Dorrington R, Bradshaw D, Laubscher R, Nannan N. Rapid Mortality Surveillance Report 2016. Cape Town: South African Medical Researc Council; 2018. Available: http://www.mrc.ac.za/sites/default/files/files/2018-02-22/RapidMortalitySurveillanceReport2016.pdf. Accessed 11 Dec 2018
423. Darikwa TB, Dorrington R. The level and trends of child mortality in South Africa, 1996-2006. *African Population Studies* 2011; **25**:159-172.
424. Joubert J, Rao C, Bradshaw D, Dorrington RE, Vos T, Lopez AD. Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa. *Glob Health Action* 2012; **5**:19263.
425. Kerber KJ, Lawn JE, Johnson LF, Mahy M, Dorrington RE, Phillips H, *et al.* South African child deaths 1990-2011: have HIV services reversed the trend enough to meet Millennium Development Goal 4? *AIDS* 2013; **27**:2637-2648.
426. Maskew M, Bor J, MacLeod W, Carmona S, Sherman GG, Fox MP. Adolescent HIV treatment in South Africa's national HIV programme: a retrospective cohort study. *Lancet HIV* 2019; **6**:e760-768.
427. Bamford L. Leading causes of death in children. In: *Saving Children 2012-2013: An eighth survey of child healthcare in South Africa*. Edited by Stephen CR. Pretoria: Tshepesa Press; 2016. pp. 15-36.
428. Harper K. An overview of Child PIP data 2012-2013. In: *Saving Children 2012-2013: An eighth survey of child healthcare in South Africa*. Edited by Stephen CR. Pretoria: Tshepesa Press; 2016. pp. 1-13.
429. Price J, Willcox M, Kabudula CW, Herbst K, Hinton L, Kahn K, *et al.* Care pathways during a child's final illness in rural South Africa: Findings from a social autopsy study. *PLoS One* 2019; **14**:e0224284.
430. Jiamsakul A, Kariminia A, Althoff KN, Cesar C, Cortes CP, Davies MA, *et al.* HIV viral load suppression in adults and children receiving antiretroviral therapy - results from the IeDEA collaboration. *Journal of Acquired Immune Deficiency Syndrome* 2017; **76**:319-329.
431. George S, McGrath N, Oni T. The association between a detectable HIV viral load and non-communicable diseases comorbidity in HIV positive adults on antiretroviral therapy in Western Cape, South Africa. *BMC Infect Dis* 2019; **19**:348.
432. Fatti G, Grimwood A, Nachega JB, Nelson JA, LaSorda K, Zyl GV, *et al.* Better virological outcomes amongst people living with HIV initiating early antiretroviral treatment (CD4 counts >/= 500 cells/microL) in the HPTN 071 (PopART) trial in South Africa. *Clin Infect Dis* 2020; **70**:395-403.
433. Edet A, Akinsola H, Bessong PO. Virologic and immunologic responses of patients on highly active antiretroviral therapy in a rural community health centre in Limpopo, South Africa: A retrospective study. *South Afr J HIV Med* 2019; **20**:818.
434. Euvrard J, Schulz T, Hilderbrand K, Bosland M, Osler M, Boulle A, *et al.* How accurately do routinely reported HIV viral load suppression proportions reflect progress towards the 90-90-90 target in the population on antiretroviral treatment in Khayelitsha, South Africa? *S Afr Med J* 2019; **109**:174-177.
435. Sikazwe I, Eshun-Wilson I, Sikombe K, Czaicki N, Somwe P, Mody A, *et al.* Retention and viral suppression in a cohort of HIV patients on antiretroviral therapy in Zambia: Regionally representative estimates using a multistage-sampling-based approach. *PLoS Med* 2019; **16**:e1002811.

436. Hennessey C. From Lost to Found: the Silent Transfer of Patients on Antiretroviral Therapy in Khayelitsha, South Africa: University of Cape Town; 2015.
437. Zadrozny S, Weir S, Edwards J, Herce M. Applying new methods to estimate viral suppression: the "last 90". Chapel Hill, USA: MEASURE Evaluation, University of North Carolina; 2018.
438. Smith AFM, Gelfand AE. Bayesian statistics without tears - a sampling resampling perspective. *Am Stat* 1992; **46**:84-88.
439. Lilian RR, Mutasa B, Railton J, Mongwe W, McIntyre JA, Struthers HE, *et al.* A 10-year cohort analysis of routine paediatric ART data in a rural South African setting. *Epidemiol Infect* 2017; **145**:170-180.
440. Cornell M, Lessells R, Fox MP, Garone DB, Giddy J, Fenner L, *et al.* Mortality among adults transferred and lost to follow-up from antiretroviral therapy programmes in South Africa: a multicenter cohort study. *J Acquir Immun Defic Syndr* 2014; **67**:e67-75.
441. Kaplan SR, Oosthuizen C, Stinson K, Little F, Euvrard J, Schomaker M, *et al.* Contemporary disengagement from antiretroviral therapy in Khayelitsha, South Africa: A cohort study. *PLoS Med* 2017; **14**:e1002407.
442. Kranzer K, Lewis JJ, Ford N, Zeinecker J, Orrell C, Lawn SD, *et al.* Treatment interruption in a primary care antiretroviral therapy program in South Africa: cohort analysis of trends and risk factors. *J Acquir Immun Defic Syndr* 2010; **55**:e17-23.
443. Clouse K, Vermund SH, Maskew M, Lurie MN, MacLeod W, Malete G, *et al.* Mobility and clinic switching among postpartum women considered lost to HIV care in South Africa. *Journal of Acquired Immune Deficiency Syndrome* 2017; **74**:383-389.
444. Wilkinson LS, Skordis-Worrall J, Ajose O, Ford N. Self-transfer and mortality amongst adults lost to follow-up in ART programmes in low- and middle-income countries: systematic review and meta-analysis. *Trop Med Int Health* 2015; **20**:365-379.
445. Brinkhof MW, Pujades-Rodriguez M, Egger M. Mortality of patients lost to follow-up in antiretroviral treatment programmes in resource-limited settings: systematic review and meta-analysis. *PLoS One* 2009; **4**:e5790.
446. Haas AD, Zaniewski E, Anderegg N, Ford N, Fox MP, Vinikoor M, *et al.* Retention and mortality on antiretroviral therapy in sub-Saharan Africa: collaborative analyses of HIV treatment programmes. *J Int AIDS Soc* 2018; **21**:e25084.
447. Keiser O, Chi BH, Gsponer T, Boulle A, Orrell C, Phiri S, *et al.* Outcomes of antiretroviral treatment in programmes with and without routine viral load monitoring in Southern Africa. *AIDS* 2011; **25**:1761-1769.
448. Haas AD, Keiser O, Balestre E, Brown S, Bissagnene E, Chimbetete C, *et al.* Monitoring and switching of first-line antiretroviral therapy in adult treatment cohorts in sub-Saharan Africa: collaborative analysis. *Lancet HIV* 2015; **2**:e271-278.
449. Madec Y, Leroy S, Rey-Cuille MA, Huber F, Calmy A. Persistent difficulties in switching to second-line ART in sub-saharan Africa - a systematic review and meta-analysis. *PLoS One* 2013; **8**:e82724.
450. Ssempijja V, Nakigozi G, Chang L, Gray R, Wawer M, Ndyanabo A, *et al.* Rates of switching to second-line antiretroviral therapy and impact of delayed switching on immunologic, virologic, and mortality outcomes among HIV-infected adults with virologic failure in Rakai, Uganda. *BMC Infect Dis* 2017; **17**:582.
451. Davies MA, Moultrie H, Eley B, Rabie H, Van Cutsem G, Giddy J, *et al.* Virologic failure and second-line antiretroviral therapy in children in South Africa - the IeDEA

Southern Africa collaboration. *Journal of Acquired Immune Deficiency Syndrome* 2011; **56**:270-278.
452. Collins IJ, Goodall R, Smith C, Wools-Kaloustian K. Incidence of switch to second-line antiretroviral therapy and associated factors in children with HIV: an international cohort collaboration *Lancet HIV* 2019; **6**:e105-115.
453. Wools-Kaloustian K, Marete I, Ayaya S, Sohn AH, Van Nguyen L, Li S, *et al.* Time to first-line ART failure and time to second-line ART switch in the IeDEA Pediatric Cohort. *Journal of Acquired Immune Deficiency Syndrome* 2018; **78**:221-230.

Appendix A: Mathematical approach to modelling sexual behaviour

This appendix provides further mathematical detail regarding the modelling of sexual behaviour. Sections A.1-A.3 describe the calculations performed to ensure that male rates of partnership formation are consistent with female rates of partnership formation. Section A.4 explains the method for calculating female rates of movement into and out of commercial sex. Finally, section A.5 explains the approach to modelling divorce and widowhood. In all sections, the symbol $N_{g,i,l,j}(x,t)$ represents the number of sexually active individuals aged *x* in year *t*, who are of sex *g* and risk group *i*, in relationship category *l* (0 for heterosexual unmarried, 1 for heterosexual married/cohabiting, 2 for female sex workers and 3 for MSM) with a partner in risk group *j* (the *j* subscript is omitted in the case of unmarried individuals, i.e. for *l* = 0, 2 or 3). Within this group we define $X_{g,i,l,j}(x,a,s,v,d)$ to be the proportion who are in HIV stage *s* (representing CD4 category in untreated infection), with ART status *a* (0 if untreated), HIV testing history *v* and ART duration *d*.

A.1 Non-spousal heterosexual relationships

Suppose that $\Phi_{g,i}(x,t)$ is the total number of non-spousal relationships formed by individuals of sex *g* and age *x*, in risk group *i*, during year *t*. For high-risk women this is calculated as

$$\Phi_{2,1}(x,t) = N_{2,1,0}(x,t)\,c_{2,1,0}(x) + \left(N_{2,1,1,1}(x,t) + N_{2,1,1,2}(x,t)\right)c_{2,1,1}(x),$$

where $c_{g,i,l}(x)$ is the annual rate of non-marital partnership formation in individuals aged *x*, of sex *g* and marital status *l*, who are in risk group *i* (1 for high risk, 2 for low risk). For low-risk women the number of new partnerships is just

$$\Phi_{2,2}(x,t) = N_{2,2,0}(x,t)\,c_{2,2,0}(x),$$

since married women in the low risk group are assumed not to have extramarital partners. The total number of new heterosexual non-spousal partnerships involving men of age *y* is then calculated as

$$\Phi_{1,1}(y,t) + \Phi_{1,2}(y,t) = \sum_{x=10}^{90}\left(\Phi_{2,1}(x,t) + \Phi_{2,2}(x,t)\right)f_{2,0}(y \mid x),$$

where $f_{g,l}(y \mid x)$ is the probability that for an individual of sex *g* and age *x*, in a relationship of type *l*, the partner's age is *y* (as defined in section 2.6). The rate at which unmarried men in the high-risk group form new heterosexual partnerships in year *t* is then calculated by observing that

$$\Phi_{1,1}(y,t)+\Phi_{1,2}(y,t)=\left(N_{1,1,0}(y,t)+N_{1,1,3}(y,t)\left(1-\Omega(y)\right)\right)c_{1,1,0}(y,t)+$$
$$+\left(N_{1,2,0}(y,t)+N_{1,2,3}(y,t)\left(1-\Omega(y)\right)\right)c_{1,1,0}(y,t)L_1$$
$$+\left(N_{1,1,1,1}(y,t)+N_{1,1,1,2}(y,t)\right)c_{1,1,0}(y,t)R_1$$

where L_1 and R_1 are the relative rates of non-spousal partnership formation in unmarried low-risk men and married high-risk men respectively (expressed as multiples of the rate in unmarried high-risk men) and $\Omega(y)$ is the fraction of partners who are men, among MSM aged y. From this we calculate

$$c_{1,1,0}(y,t)=\frac{\Phi_{1,1}(y,t)+\Phi_{1,2}(y,t)}{N_{1,1,0}(y,t)+N_{1,2,0}(y,t)L_1+\left(N_{1,1,3}(y,t)+N_{1,2,3}(y,t)L\right)\left(1-\Omega(y)\right)+\left(N_{1,1,1,1}(y,t)+N_{1,1,1,2}(y,t)\right)R_1}.$$

It is worth noting in passing that the rates at which men form non-spousal relationships are a function of t, while the rates at which women form non-spousal relationships are assumed to be independent of t. This is because male sexual activity is assumed to change over time in response to demographic changes (relative numbers of males and females at different ages and numbers of married and unmarried individuals at different ages). In reality, both male and female sexual behaviour patterns would change and male behaviour would not be dictated entirely by female 'demand' for sexual partners, but in the interests of mathematical simplicity, we fix the female sexual behaviour parameters.

For a man who is aged y, starting a new non-spousal relationship in year t, the probability that his female partner is between the ages of x and $x + 1$ is

$$f_{1,0}(x\mid y,t)=\frac{\left(\Phi_{2,1}(x,t)+\Phi_{2,2}(x,t)\right)f_{2,0}(y\mid x)}{\Phi_{1,1}(y,t)+\Phi_{1,2}(y,t)}.$$

A.2 Mixing between risk groups in non-spousal heterosexual relationships

The total number of non-spousal heterosexual relationships formed by men in the high-risk group in year t is

$$\Phi_{1,1}(.,t)=\sum_{y=10}^{90}\left\{N_{1,1,0}(y,t)+N_{1,1,3}(y,t)\left(1-\Omega(y)\right)+\left(N_{1,1,1,1}(y,t)+N_{1,1,1,2}(y,t)\right)R_1\right\}c_{1,1,0}(y,t)$$

and the total number of non-spousal heterosexual relationships formed by low-risk men is

$$\Phi_{1,2}(.,t)=\sum_{y=10}^{90}\left(N_{1,2,0}(y,t)+N_{1,2,3}(y,t)\left(1-\Omega(y)\right)\right)c_{1,1,0}(y,t)L_1.$$

The total numbers of non-spousal heterosexual relationships formed by women in the high-risk and low-risk groups ($\Phi_{2,1}(.,t)$) and $\Phi_{2,2}(.,t)$) respectively) are similarly defined. For women who are in risk group i in year t, the probability that their non-spousal partner is in risk group j is

$$\rho_{2,i,0}(j,t) = (1-\varepsilon) \times I(i=j) + \varepsilon \times \frac{\Phi_{1,j}(.,t)}{\Phi_{1,1}(.,t) + \Phi_{1,2}(.,t)},$$

where ε is the assortativeness parameter described in section 2.5, and $I(i=j)$ is an indicator function (taking on value 1 when $i = j$ and value 0 when $i \neq j$). For men who are in risk group j in year t, the probability that their female non-spousal partner is in risk group i is calculated as

$$\rho_{1,j,0}(i,t) = \frac{\Phi_{2,i}(.,t)\rho_{2,i,0}(j,t)}{\Phi_{2,1}(.,t)\rho_{2,1,0}(j,t) + \Phi_{2,2}(.,t)\rho_{2,2,0}(j,t)}.$$

A.3 Partner age and risk group preferences in spousal relationships

We calculate the proportion of married men, aged y in year t, whose partners are aged x as:

$$f_{1,1}(x \mid y,t) = \frac{\left(N_{2,1,1,1}(x,t) + N_{2,1,1,2}(x,t) + N_{2,2,1,1}(x,t) + N_{2,2,1,2}(x,t)\right) f_{2,1}(y \mid x)}{\sum_{v=15}^{90} \left(N_{2,1,1,1}(v,t) + N_{2,1,1,2}(v,t) + N_{2,2,1,1}(v,t) + N_{2,2,1,2}(v,t)\right) f_{2,1}(y \mid v)}.$$

It is worth noting here that y represents the *current* partner age, not the age of partners in newly-formed spousal relationships, since there is an implicit allowance for differential rates of survival at different ages in the calculation of $f_{2,1}(y \mid x)$.

The number of men in risk group i who enter spousal relationships in year t is calculated as

$$D_{1,i}(t) = \sum_{y=15}^{90} \left(N_{g,i,0}(y,t) + N_{g,i,3}(y,t)\right) m_{g,i}(y,t),$$

where $m_{g,i}(y,t)$ is the annual probability of forming a new spousal relationship at age y. A similar formula is used to calculate the number of women who enter spousal relationships, except that the MSM term ($N_{g,i,3}(y,t)$) is omitted. For women who are in risk group i, entering into a spousal relationship in year t, the probability that their new partner is in risk group j is

$$\rho_{2,i,1}(j,t) = (1-\varepsilon) \times I(i=j) + \varepsilon \times \frac{D_{1,j}(t)}{D_{1,1}(t) + D_{1,2}(t)}.$$

For men in risk group j who are entering spousal relationships in year t, the probability that their new partner is in risk group i is calculated as

$$\rho_{1,j,1}(i,t) = \frac{D_{2,i}(t)\rho_{2,i,1}(j,t)}{D_{2,1}(t)\rho_{2,1,1}(j,t) + D_{2,2}(t)\rho_{2,2,1}(j,t)}.$$

A.4 Female rates of entry into and exit from sex work

At the end of each month the model updates female movements into and out of sex work based on assumed rates of retirement from sex work and based on male demand for sex work. The total male demand for sex workers at time *t* is calculated as

$$E(t) = \frac{1}{C} \sum_{x,l,j} \sum_{a,s,v,d} N_{1,1,l,j}(x,t) X_{1,1,l,j}(x,a,s,v,d) w_l(x) Y(a,s,d)$$

where $w_l(x)$ is the rate at which HIV-negative men visit sex workers (as defined in section 2.4), $Y(a,s,d)$ is the adjustment made to the coital frequencies of HIV-positive individuals (as defined in section 4.6), and *C* is the assumed average annual number of clients per sex worker. MSM are assumed to have no contact with female sex workers. As explained in section 2.4, there is assumed to be a constant sex worker age distribution, with $\phi(x)$ representing the fraction of sex workers who are aged *x* years. The required number of sex workers aged *x* at time *t* is therefore $E(t)\phi(x)$.

Suppose that $\tau(a,s,d)$ represents the monthly probability of retirement from commercial sex in sex workers who are in HIV stage *s*, with ART status *a* and ART duration *d* years. Then at age *x*, the total number of sex workers retiring from sex work in month *t* is

$$N_{2,1,2}(x,t-1) \sum_{a,s,v,d} X_{2,1,2}(x,a,s,v,d) \tau(a,s,d).$$

(It is worth noting that although the symbol $N_{2,1,2}(x,t-1)$ represents the number of sex workers at time $(t-1)$, the calculation is actually performed *after* HIV disease progression and AIDS mortality in month *t* have been updated.) In order to meet the male demand for sex workers, the number of women aged *x* who need to enter sex work during month *t* is

$$\Delta_c(x,t) = E(t-1)\phi(x) - N_{2,1,2}(x,t-1)\left(1 - \sum_{a,s,v,d} X_{2,1,2}(x,a,s,v,d)\tau(a,s,d)\right).$$

Women enter into sex work from the unmarried high-risk group, but it is assumed that women in the advanced stages of HIV disease are less likely to enter sex work than women who are HIV-negative or asymptomatic. The symbol $W(a,s,d)$ represents the relative probability of entry into commercial sex (compared to HIV-negative women) for women who are in HIV stage *s*, with ART status *a* and ART duration *d* years. For sexually experienced HIV-negative women in the high-risk unmarried group, who are aged *x* at time $(t-1)$, the probability of entry into sex work in month *t* is

$$\frac{\Delta_c(x,t)}{N_{2,1,1}(x,t-1) \sum_{a,s,v,d} X_{2,1,1}(x,a,s,v,d) W(a,s,d)}.$$

For HIV-positive women, the probability of entry into sex worker is obtained by multiplying the above expression by the relevant $W(a,s,d)$ factor.

The variables $\tau(a,s,d)$ and $W(a,s,d)$, discussed in section 2.4, are a function only of current CD4 count in untreated individuals (*s*), but for treated individuals the variable *s* represents the baseline CD4 category. In treated individuals the $\tau(a,s,d)$ and $W(a,s,d)$ variables are therefore calculated based on the expected distribution of current CD4 counts in individuals who started ART in CD4 category *s*, *d* years previously. This expected CD4 distribution is defined in Table 4.3.

A.5 Divorce and widowhood

Divorce and widowhood are calculated on an annual basis. Consider a married individual of age *x* and sex *g*, in risk group *i*, with married partner in risk group *j*. The probability that the relationship does not terminate in the current year is calculated as the product of three probabilities:

a) the probability that the partner does not die from AIDS;
b) the probability that the partner does not die from non-AIDS causes; and
c) the probability that the relationship does not end through divorce.

Considering the first probability, we define $q^{A}_{g,j,i}(y,t)$ to be the probability of AIDS death during the course of year *t*, for a married individual of age *x* and risk group *j*, who is alive at the start of year *t*. The average probability that the partner does not die from AIDS during year *t* is

$$1-\sum_{y=15}^{90} f_{g,1}(y \mid x,t)\, q^{A}_{3-g,j,i}(y,t)\,,$$

where (3 – *g*) is the sex opposite to *g*. Similarly, we define $q^{N}_{g}(y,t)$ to be the probability of death due to a non-AIDS cause during the course of year *t*, for a married individual of age *x* and sex *g*, who is alive at the start of year *t*. The average probability that the partner does not die from non-AIDS causes during year *t* is then

$$1-\sum_{y=15}^{90} f_{g,1}(y \mid x,t)\, q^{N}_{3-g}(y,t).$$

Finally, we define $\delta_{g}(x)$ to be the annual rate at which married individuals of age *x* and sex *g* divorce, so that the probability that the relationship does not end in divorce is $\exp\left(-\delta_{g}(x)\right)$. Combining these three expressions, the probability that an individual of age *x*, sex *g* and risk group *i*, who is married to a partner of risk group *j* at the start of year *t*, returns to the single state in the course of year *t* is

$$1-\left(1-\sum_{y=15}^{90} f_{g,1}(y \mid x,t)\, q^{A}_{3-g,j,i}(y,t)\right)\left(1-\sum_{y=15}^{90} f_{g,1}(y \mid x,t)\, q^{N}_{3-g}(y,t)\right)\exp\left(-\delta_{g}(x)\right).$$

A.6 Partner age preferences in MSM

Few studies report on age mixing patterns in MSM relationships in the South African setting. Arnold *et al* [35] found that in 758 male-male sexual relationships in Soweto, the average partner age difference was small (0.25 years) but there was high variation in partner age differences (standard deviation of 5.8 years). Based on what is known about the age distribution of sexually active MSM in South Africa, it is possible to use this information to determine how patterns of age mixing vary in relation to age. If $N(x)$ is the age distribution of sexually active MSM and $f_{1,3}(y \mid x)$ represents the proportion of male partners aged y for an MSM aged x, then for a random sample of MSM, the expected proportion of their partners who are aged y is

$$\int N(x) f_{1,3}(y \mid x) dx.$$

If the sample of sexually active MSM is truly representative, then we would expect that this proportion should be the same as $N(y)$. We would also expect that

$$\int N(x) \int f_{1,3}(y \mid x)(x-y)^2 \, dy dx = 5.8^2,$$

if the estimated standard deviation of 5.8 years [35] is correct. These constraints allow us to determine the likely patterns of sexual mixing. It is assumed that $N(x)$ is a gamma distribution, with a mean of 25 years and a standard deviation of 7 years [35, 382], with an age offset of 10 years to prevent implausible levels of sexual activity in very young boys. The $f_{1,3}(y \mid x)$ distribution is also assumed to be of gamma form, with mean of $\mu(x) = \max(x-10, x + A(25-x))$ and variance of B^2 (again, with an offset of 10 years to prevent sexual activity at young ages). The two free parameters, A and B, have been set to 0.45 and 5.0 years respectively, to yield a variance of partner age differences equal to 5.8^2, as well as a distribution of $\int N(x) f_{1,3}(y \mid x) dx$ values roughly consistent with the distribution of $N(y)$ values (Figure A1).

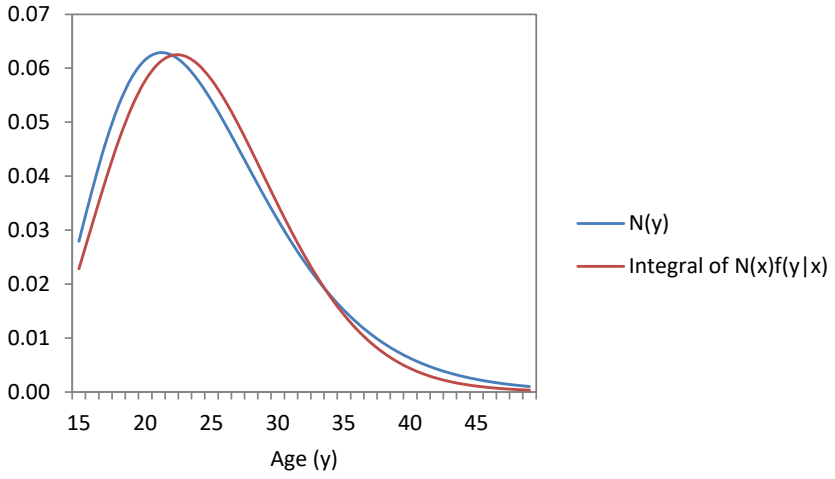


Figure A1: Age distribution of sexual activity in South African MSM

Appendix B: Calibration to HIV testing data in adults

A multi-parameter evidence synthesis approach has been used to estimate rates of HIV testing and diagnosis in South Africa. This involves triangulating self-reported levels of testing, total numbers of HIV tests and HIV prevalence levels in individuals seeking HIV testing, within a Bayesian framework. The sections that follow describe the data sources (sections B.1-B.4), the mathematical structure of the model (sections B.4-B.6), the Bayesian procedure (sections B.7 and B.8) and the comparison of the model estimates with the data (sections B.9 and B.10).

B.1 Numbers of HIV tests performed

Our approach to estimating the total number of HIV tests performed in South Africa is to aggregate estimates from four different sources: the public health sector, private medical schemes, insurance companies (tests performed on individuals applying for life insurance) and other private providers (tests performed by employee wellness programmes and workforce programmes that are independent of medical schemes). Of these four sources, the most significant are the public sector and the insurance industry, both of which have supplied data for several years. However, data are missing or incomplete for some years, and data from other sources are very limited. Our approach is therefore to estimate total numbers of HIV tests in South Africa for each of the years for which public sector data are available, then to use linear interpolation and extrapolation to estimate the totals for the other years. The data sources up to the 2011-12 year have been described previously in the supplementary material of a previous paper on adult HIV testing [124].

A significant limitation of the historic data is that it is often unclear whether the reported numbers of tests are for adults or for adults and children combined. Almost all HIV testing performed in the first 18 months of life relies on PCR testing, not antibody testing, and it is assumed that this PCR testing is not included in the reported totals. In the case of the insurance industry data, it would be rare to take out insurance on a child's life, and it can therefore be assumed that all of the insurance industry testing is done in adults. Tests performed as part of employee wellness programmes are also done almost exclusively in adults. In the case of the public sector data, reporting is inconsistent and sometimes contradictory. Some government reports indicate that the quoted numbers of tests are for 15-49 year olds [233, 234, 236], but some state that the reported totals are for ages 5 and older (Tshepo Molapo, personal communication), and most do not specify the age range for the reported statistics. Only in the four most recent reporting years (2015-16 to 2018-19) is a detailed age breakdown given. In the years prior to this, we assume that all reported numbers of HIV tests relate to the age group 15 and older. This assumption is intended as a compromise between the bias that would arise if the reported tests were for adults and children combined and the bias that would arise if the reported totals related only to 15-49 year olds.

Table B1 shows the assumed annual numbers of HIV tests performed in adults in each year. Assumptions for the period up to 2011/12 are the same as presented previously [124]. More recent public sector statistics were obtained from the Department of Health (Thapelo Seatlhodi and Tshepo Molapo, personal communication), while medical scheme estimates

were obtained by scaling up previous estimates for 2011 in proportion to the size of the medical scheme beneficiary population [383]. Due to lack of recent data on HIV testing by the insurance industry and other private sector organizations, annual testing numbers in these sectors are assumed to have remained unchanged. Although this assumption is unrealistic, the private sector contributes relatively little to total testing volumes in recent years (7%), and any bias introduced by these assumptions is therefore unlikely to distort the overall estimates substantially.

Table B1: Assumed numbers of HIV tests performed in South African adults (in thousands)

Year	Public sector	Medical schemes	Insurance industry	Other private	Total
2002-03	691000	285897	432268	75798	1484963
2003-04	821238	285897	432268	75798	1615201
2004-05	951476	285897	432268	75798	1745439
2005-06	1376582	290102	432268	75798	2174750
2006-07	1610755	298510	432268	75798	2417331
2007-08	1923430	315328	432268	75798	2746824
2008-09	2591441	327941	432268	75798	3427448
2009-10	6770000	336350	432268	75798	7614416
2010-11	9523400	344758	432268	75798	10376224
2011-12	8772000	353167	432268	75798	9633233
2012-13	8978177	361576	432268	75798	9847819
2013-14	7334942	365780	432268	75798	8208788
2014-15	8636033	369985	432268	75798	9514084
2015-16	11324134	369985	432268	75798	12202185
2016-17	12465313	374189	432268	75798	13347568
2017-18	11902403	374189	432268	75798	12784658
2018-19	12714196	374189	432268	75798	13596451

In the period prior to 2002, there is almost no data to guide assumptions about annual numbers of HIV tests performed. Given the lack of data prior to 2002, our approach is to assume that annual numbers of HIV tests in adults increased linearly from zero in 1990 to 1 484 963 in 2002.

B.2 HIV prevalence in individuals tested for HIV

Although numbers of HIV tests performed in South Africa are routinely reported, the reporting of the fraction of individuals testing positive has been sporadic. Methods used to estimate the HIV prevalence in individuals testing for HIV have been slightly modified from the method described previously [124] to take into account uncertainty regarding the sensitivity and specificity of the rapid testing, which has been recommended in South Africa since 2003 [384]. In South Africa and other African settings there has been particular concern regarding the accuracy of HIV rapid testing performed by lay health workers [385, 386], and it is therefore appropriate to rely on local estimates of sensitivity and specificity of rapid tests when performed in field settings. Table B2 summarizes the results of African studies that have evaluated the sensitivity and/or specificity of a rapid diagnostic algorithm (i.e. an initial test followed by a confirmatory test in the case of individuals who test positive). The assumed average sensitivity is 93.6% and the assumed average specificity is 99.7%, corresponding to the medians of the estimates in Table B2. In order to model the uncertainty around the

reported HIV prevalence levels, it is also necessary to specify a standard deviation around the sensitivity and specificity, which we set to 3.5% and 0.2% respectively.

Table B2: Sensitivity and specificity of rapid testing in South Africa

Study	Location	Years	Sensitivity	n	Specificity	n
Bock *et al* [387]	Western Cape, SA	2014	51.1%	90	100.0%	1496
		2015	74.4%	117	100.0%	2276
		2016	85.3%	34	100.0%	1375
Jackson *et al* [388]	KZN, SA	2009-11	98.0%	491	99.6%	3505
Wolpaw *et al* [389]	Cape Town, SA	2008-09	91.3%	150	-	
Kufa *et al* [390]	KZN, SA	2015-16	91.1%	326	99.9%	3382
Bassett *et al* [391]	Durban, SA	2007	98.5%	1314	-	
Gray *et al* [392]	Uganda	2003-04	97.7%	170	90.4%	1347
Shanks *et al* [393]	DRC	-	-		98.7%	2568
Bruzzone *et al* [394]	Congo	2005-06	100.0%	200	99.7%	3414
Urassa *et al* [395]	Tanzania	-	95.8%	215	99.7%	1646
Median			93.6%		99.7%	

DRC = Democratic Republic of Congo; KZN = KwaZulu-Natal; SA = South Africa.

Suppose that $\rho(t)$ is the true HIV prevalence among people who test for HIV in year t. The true prevalence is unknown for two reasons: (a) imperfect test sensitivity and specificity, and (b) uncertainty about HIV prevalence among individuals testing for HIV in the private sector. If it is assumed that J is the ratio of HIV prevalence among individuals tested in the private sector to that among individuals tested in the public sector then

$$\rho(t) = Z(t)\big(JR(t) + \big(1 - R(t)\big)\big), \qquad \text{(B1)}$$

where $Z(t)$ is the HIV prevalence among individuals tested in the public sector and $R(t)$ is the fraction of individuals receiving HIV testing in year t who test through the private sector (estimated from the data in Table B1). We can estimate J using data from 2010-11, the year for which we have the most complete HIV testing data. In this year, the HIV prevalence among individuals testing for HIV was 16.21% in the public sector, 1.59% in insurance applicants and 9.51% in workplace programmes. The HIV prevalence among individuals testing for HIV through medical schemes is unknown, but is generally considered to be lower than in the public sector [396]. If HIV prevalence in medical scheme testers is assumed to be half of that in public sector testers, we estimate J = 0.304, but if we assume the HIV prevalence in medical scheme testers to be 0 or the same as that in public sector testers (likely lower and upper bounds on the true prevalence), the estimated values of J are 0.102 and 0.506 respectively. Based on these calculations, we set E[J] to 0.304 and Var[J] = 0.117^2, on the assumption that J is uniformly distributed on the interval (0.102, 0.506). It follows from equation (B1) that Var[$\rho(t)$] = Var[J] $(R(t)Z(t))^2$.

Now suppose that $\theta(t)$ is the HIV prevalence that we might observe after the sensitivity and specificity of the HIV testing algorithm are taken into account:

$$\theta(t) = \rho(t)\big(Se + Sp - 1\big) + 1 - Sp\,, \qquad \text{(B2)}$$

where *Se* and *Sp* are the sensitivity and specificity respectively. We have previously shown [397] that the mean and variance of this expression are

$$\mathrm{E}[\theta(t)] = \mathrm{E}[\rho(t)]\left(\mathrm{E}[Se] + \mathrm{E}[Sp] - 1\right) + 1 - \mathrm{E}[Sp] \tag{B3}$$

and

$$\begin{aligned}\mathrm{Var}[\theta(t)] = {} & \mathrm{Var}[Se]\left(\mathrm{Var}[\rho(t)] + \mathrm{E}[\rho(t)]^2\right) \\ & + \mathrm{Var}[Sp]\left(\mathrm{Var}[\rho(t)] + (1 - \mathrm{E}[\rho(t)])^2\right) \\ & + \left(\mathrm{E}[Se] + \mathrm{E}[Se] - 1\right)^2 \mathrm{Var}[\rho(t)]\end{aligned} \tag{B4}$$

respectively. Note that $\mathrm{E}[\rho(t)] = Z(t)(\mathrm{E}[J]R(t) + (1 - R(t)))$. The only quantity in these equations for which we have not previously calculated values is $Z(t)$), which we take to be the reported HIV prevalence in public sector testers after adjusting for the expected sensitivity and specificity of the testing algorithm. Table B3 shows the calculations of these different quantities over the 2004-2019 period.

Table B3: Estimated HIV prevalence and associated uncertainty in adults testing for HIV

Year	Public sector reported	Public sector adjusted ($Z(t)$)	Private sector proportion ($R(t)$)	Expected total prevalence $\mathrm{E}[\rho(t)]$	Variance of total prevalence $\mathrm{Var}[\rho(t)]$	Expected unadjusted prevalence $\mathrm{E}[\theta(t)]$	Variance of unadjusted prevalence $\mathrm{Var}[\theta(t)]$
2004-05	35.57%	37.81%	45.49%	25.84%	$2.01\%^2$	24.41%	$2.09\%^2$
2005-06	34.89%	37.07%	36.70%	27.60%	$1.59\%^2$	26.05%	$1.78\%^2$
2006-07	31.82%	33.78%	33.37%	25.93%	$1.32\%^2$	24.50%	$1.53\%^2$
2007-08	30.14%	31.99%	29.98%	25.31%	$1.12\%^2$	23.92%	$1.38\%^2$
2008-09	28.44%	30.16%	24.39%	25.04%	$0.86\%^2$	23.66%	$1.20\%^2$
2010-11	16.21%	17.05%	8.22%	16.08%	$0.16\%^2$	15.30%	$0.61\%^2$
2012-13	14.89%	15.64%	8.83%	14.67%	$0.16\%^2$	13.99%	$0.56\%^2$
2015-16	9.45%	9.81%	7.20%	9.32%	$0.08\%^2$	8.99%	$0.38\%^2$
2016-17	8.39%	8.67%	6.61%	8.27%	$0.07\%^2$	8.02%	$0.35\%^2$
2017-18	7.48%	7.70%	6.90%	7.33%	$0.06\%^2$	7.14%	$0.32\%^2$
2018-19	6.41%	6.55%	6.49%	6.25%	$0.05\%^2$	6.14%	$0.29\%^2$

B.3 Proportions of adults ever tested for HIV

Estimates of the proportions of adults ever tested for HIV were obtained from five national household surveys, namely the Human Sciences Research Council (HSRC) surveys in 2005, 2008, 2012 and 2017 [4, 18, 125, 398] and the 2016 Demographic and Health Survey (DHS) [126]. Estimates were stratified by age group, sex and HIV status, as shown in Table B5; these same data have previously been used in the calibration of the Thembisa model, although previous analyses did not include the 2016 DHS data.

Table B5: Proportions of adults reporting having ever tested for HIV

		HIV-negative		HIV-positive	
Year	Age	Males	Females	Males	Females
2005	15-24	11.3% (9.2-13.9)	27.3% (24.1-30.8)	22.4% (9.2-45.1)	37.2% (29.5-45.6)
	25-34	43.4% (34.4-52.7)	48.2% (43.2-53.3)	29.6% (19.9-41.5)	44.1% (35.6-52.9)
	35-44	46.4% (40.3-52.7)	47.3% (42.2-52.4)	34.2% (24.7-45.3)	35.6% (27.0-45.2)
	45-59	36.0% (30.6-41.7)	26.1% (22.7-29.9)	49.4% (31.0-67.9)	25.4% (15.0-39.7)
	60+	17.8% (12.2-25.4)	6.2% (4.2-9.2)	40.6% (11.2-78.8)	3.5% (0.9-12.8)
2008	15-24	25.7% (21.9-29.9)	50.5% (46.8-54.3)	21.2% (8.8-43.0)	73.3% (65.2-80.0)
	25-34	54.8% (48.2-61.2)	80.6% (76.1-84.4)	52.5% (39.9-64.9)	76.4% (68.6-82.7)
	35-44	63.2% (56.4-69.5)	70.7% (66.1-75.0)	63.9% (48.9-76.6)	69.5% (61.4-76.6)
	45-59	59.9% (54.3-65.2)	43.3% (39.1-47.5)	61.2% (44.7-75.5)	57.2% (44.6-69.0)
	60+	28.8% (22.8-35.7)	15.5% (12.3-19.4)	62.7% (32.4-85.5)	26.9% (8.3-59.7)
2012	15-24	38.1% (34.5-41.7)	61.6% (58.4-64.8)	58.9% (46.2-70.6)	78.8% (70.9-85.1)
	25-34	65.3% (60.3-69.9)	91.0% (88.7-92.9)	61.9% (49.8-72.7)	94.1% (91.6-95.9)
	35-44	71.8% (66.1-76.8)	86.3% (83.0-89.0)	75.4% (63.8-84.2)	91.5% (86.0-95.0)
	45-59	71.4% (66.8-75.6)	68.4% (64.8-71.7)	82.4% (70.0-90.4)	77.8% (70.3-83.7)
	60+	46.7% (41.0-52.5)	36.6% (32.6-40.8)	58.2% (36.6-77.0)	57.8% (43.2-71.2)
2016	15-24	59.3% (54.8-63.8)	70.1% (67.7-72.4)	56.6% (34.1-79.1)	87.2% (83.8-90.5)
	25-34	77.7% (71.8-83.5)	92.7% (90.6-94.8)	86.6% (80.8-92.5)	92.8% (89.8-95.7)
	35-44	79.5% (72.9-86.1)	92.9% (90.9-94.9)	88.7% (84.4-93.1)	92.4% (89.9-95.0)
	45-59	77.3% (71.9-82.8)	83.2% (79.1-87.3)	86.5% (80.8-92.3)	94.3% (91.8-96.8)
2017	15-24	50.7% (48.4-53.0)	62.3% (60.2-64.4)	81.3% (71.1-88.4)	90.3% (85.7-93.4)
	25-34	76.3% (73.8-78.7)	93.1% (91.6-94.3)	84.8% (79.1-89.2)	94.9% (92.8-96.4)
	35-44	78.8% (75.7-81.5)	89.9% (87.9-91.6)	87.3% (82.3-91.1)	96.4% (94.4-97.7)
	45-59	79.9% (77.3-82.2)	81.4% (79.5-83.2)	90.1% (85.2-93.6)	95.2% (92.8-96.8)
	60+	63.8% (60.5-67.0)	57.9% (55.3-60.5)	90.6% (79.1-96.1)	89.0% (81.6-93.7)

95% confidence intervals are shown in brackets.

B.4 Rates of HIV testing in patients with opportunistic infections

Published statistics on rates of HIV testing in patients with opportunistic infections are limited, and almost all of the published data relate to HIV testing in tuberculosis (TB) patients specifically. A further problem is that the most recent published data on HIV testing in TB patients (since 2009) do not directly report the fraction of TB patients tested for HIV but rather the fraction of TB patients who know their HIV status (which includes both those tested and those who were previously diagnosed HIV-positive). Another challenge is that even when the reported statistics represent the rate of HIV testing in TB patients who were not previously diagnosed HIV-positive, this represents only testing in the facility in which TB is treated, around the time of treatment. It does not include TB patients who may have been diagnosed with HIV after being referred for HIV testing in other facilities, or TB patients who may have sought HIV testing because their symptoms led them to suspect they may have HIV.

To address these problems, we adjust the reported fractions of TB patients who know their HIV status to take into account these biases, through a two-step process. As in section 3.2, we define $d_i(t)$ to be the fraction of OI patients, with HIV testing history i, who receive HIV testing as a result of their OI, if their OI occurs in year t. (The HIV testing history, i, is 0 if the individual has never been tested for HIV, 1 if the individual has been tested for HIV but not diagnosed positive, 2 if the individual has been diagnosed positive but not started ART, and 3 if the individual has started ART.) We wish to estimate $d_i(t)$ from $\chi_i(t)$, the reported

fractions of TB patients who know their HIV status in year t (these reported fractions are shown in Table B6). We also define $\gamma_i(t)$ to be the fraction of OI patients, with HIV testing history i, who receive HIV testing in the facility where they are treated for their OI, if their OI occurs in year t. This differs from $d_i(t)$, which includes all HIV testing resulting from the OI, regardless of whether it occurs in the health facility where the individual was treated for their OI. The two parameters are assumed to be related to each other by the odds ratio Φ, defined as the ratio of the odds of an OI patient being tested for HIV in the same facility as that in which they are treated to the odds of an OI patient receiving any HIV testing as a result of their OI, i.e.

$$\Phi = \left(\frac{\gamma_i(t)}{1-\gamma_i(t)} \right) \Big/ \left(\frac{d_i(t)}{1-d_i(t)} \right) \tag{B5}$$

In 2009 and subsequent years, the $\chi_i(t)$ and $\gamma_i(t)$ parameters are related to each other by the equation

$$\chi_i(t) = \frac{\sum_s \sum_{i=0}^{1} N_{s,i}(t)\Omega_s \gamma_i(t) + \sum_{s>0} \sum_{i=2}^{3} N_{s,i}(t)\Omega_s}{\sum_s \sum_{i=0}^{3} N_{s,i}(t)\Omega_s} \tag{B6}$$

where $N_{s,i}$(t) is the number of individuals in stage s of HIV infection (s = 0 if the individual is uninfected), with HIV testing history i, at the start of year t, and Ω_s is the annual incidence of opportunistic infections in stage s of HIV infection. (The values of Ω_s for untreated HIV-positive individuals are presented in Table 3.1, and incidence rates for HIV-negative individuals and individuals on ART are set at 0.019 and 0.10 respectively.) This equation represents $\chi_i(t)$ as the proportion of OI patients who have already been diagnosed positive prior to their OI (the second term in the numerator) plus the proportion of other OI patients who are offered HIV testing (the first term in the numerator). From this it follows that

$$\gamma_i(t) = \frac{\chi_i(t) \sum_s \sum_{i=0}^{3} N_{s,i}(t)\Omega_s - \sum_{s>0} \sum_{i=2}^{3} N_{s,i}(t)\Omega_s}{\sum_s \sum_{i=0}^{1} N_{s,i}(t)\Omega_s}. \tag{B7}$$

In the years prior to 2009, it is not clear if the reported proportions are the fractions receiving testing (among those not previously diagnosed positive) or the fractions who know their HIV status (including those who were previously diagnosed). However, early reports tend to suggest the former [135, 141], and hence $\gamma_i(t)$ is set equal to $\chi_i(t)$ in the years before 2009.

Having estimated $\gamma_i(t)$ from $\chi_i(t)$, the next step is to estimate $d_i(t)$ from $\chi_i(t)$, using equation B5. This requires an estimate of Φ, which is unknown. To represent the uncertainty regarding Φ, we assign a vague prior, which is uniform on the interval [0, 1). The posterior analysis (presented in more detail below) yields a posterior estimate of 0.57 for the Φ parameter. Table B6 presents the resulting estimates of the $d_i(t)$, $\gamma_i(t)$ and $\chi_i(t)$ parameters. It is worth noting that in 2010 and subsequent years, the net effect of the two adjustments is relatively small (i.e. there is only a modest difference between $d_i(t)$ and $\chi_i(t)$). However, in the pre-2009

period, $d_i(t)$ is substantially higher than $\chi_i(t)$ because of the assumed change in reporting in 2009 (i.e. the change from reporting the fraction tested to reporting the fraction tested or previously diagnosed).

Table B6: Proportions of OI patients tested for HIV

Year	% of OI patients tested at OI treatment facility or knowing their status ($\chi_i(t)$)		% of OI patients previously undiagnosed, who receive testing at OI treatment facility ($\gamma_i(t)$)	% of OI patients previously undiagnosed, who receive testing as a result of their OI ($d_i(t)$)
	Rate	Sources		
Pre-2004	5%		5%	9%
2004-05	8%	[135]	8%	14%
2005-06	20%	[137]	20%	32%
2006-07	31%	[137]	31%	45%
2007-08	39%	[137, 141]	39%	54%
2008-09	40%	[137, 141]	40%	55%
2009-10	53%	[137]	19%	30%
2010-11	64%	[146] [147]	37%	52%
2011-12	83%	[147]	69%	80%
2012-13	87%	[147]	75%	84%
2013-14	90%	[147]	81%	88%
2014-15	94%	[147]	88%	93%
2015-16	95%*	[147]	90%	94%

* Proportion is assumed to remain constant at 95% in subsequent years (since recent District Health Barometer reports do not report this)..

B.5 Rates of HIV testing in asymptomatic, non-pregnant individuals

Suppose that $G(t)$ is the total number of HIV tests performed in adults aged 15 and older, in year t, as shown in Table B1. As defined in section 3.2, the symbol $\tau_{g,i,s}(x,t)$ is the rate of HIV testing in sexually experienced individuals of age x and sex g, in HIV stage s and with HIV testing history i. The symbol $\tau'_{g,i,s}(x,t)$ is used to represent the corresponding rate of HIV testing in virgins. If $N_{g,i,s}(x,t)$ is the corresponding number of sexually experienced individuals and $V_{g,i,s}(x,t)$ is the number of virgins, at the start of year t, then

$$G(t) \approx \sum_g \sum_i \sum_s \sum_x N_{g,i,s}(x,t)\tau_{g,i,s}(x,t) + V_{g,i,s}(x,t)\tau'_{g,i,s}(x,t). \quad \text{(B8)}$$

(The relation is not exact because the numbers of individuals in the different strata change over the course of the year, so relying only on the values at the start of the year may lead to some bias – it will later be shown that this bias is very small.) The rate of HIV testing in asymptomatic virgins is assumed to be a multiple φ of the rate of HIV testing in asymptomatic girls aged 15 who are sexually experienced and non-pregnant, i.e.

$$\tau'_{g,i,s}(x,t) = b(t)A_2(15,t)r_i(t)\varphi + \Omega_s d_i(t) \quad \text{(B9)}$$

where $b(t)$ and $A_g(x,t)$ are the base testing rate and age-specific adjustment factor respectively (see section 3.2). Substituting equations (3.3) and (B5) into equation (B4), we obtain the following estimate for the base HIV testing rate in year t:

$$\hat{b}(t) = \frac{G(t) - \sum_g \sum_i \sum_s \sum_x N_{g,i,s}(x,t)\left\{\Omega_s d_i(t) + F_{g,s}(x,t)v_i(t)\right\} + V_{g,i,s}(x,t)\Omega_s d_i(t)}{\sum_g \sum_i \sum_s \sum_x N_{g,i,s}(x,t)A_g(x,t)r_i + V_{g,i,s}(x,t)A_2(15,t)r_i\varphi}.$$

This approximation to $b(t)$ is calculated at the start of each year, and is substituted into equation (3.3). Estimates of the $N_{g,i,s}(x,t)$ population totals are updated at monthly time steps, with the estimated values of $b(t)$ being held constant over the course of each year.

The above approach is used to calculate rates of HIV testing at ages 10 and older. In children between the ages of 5 and 10 years, the same equation (B9) is used to calculate the HIV testing rate, except that the second term (representing HIV testing in OI patients) is omitted in children who are either HIV-negative or in the early stage of HIV disease. For children who are in the advanced stage of HIV disease and not yet on ART, this testing rate is multiplied by a factor Q, as the incidence of opportunistic infections in children is not simulated directly. As described in section 5.3, rates of HIV testing in children aged 19-59 months are assumed to be higher than those at 5-10 years (by a factor of J), and the fraction of children tested at 18 months is assumed to have been constant at 20% since 2009. The same formula (adjusted by the factors J and Q) is used to model the rate of PCR testing in children under the age of 18 months who have advanced HIV infection, but no provision is made for PCR testing in children who are HIV-negative or in the early stage of HIV infection, except in the context of early infant diagnosis.

B.6 Model estimates of numbers of HIV tests and HIV prevalence in HIV testers

For individuals aged 15 and older, the total number of HIV-negative test results over the course of year t is calculated as

$$T_0(t) = \sum_{m=1}^{12} \sum_{g=0}^{1} \sum_{i=0}^{1} \sum_{s=0}^{1} \sum_{x=15}^{90} N_{g,i,s}\left(x, t + \frac{m-1}{12}\right) \tau_{g,i,s}(x,t) + V_{g,i,s}\left(x, t + \frac{m-1}{12}\right) \tau'_{g,i,s}(x,t),$$

and the total number of HIV-positive test results is

$$T_1(t) = \sum_{m=1}^{12} \sum_{g=0}^{1} \sum_{i=0}^{3} \sum_{s=2}^{5} \sum_{x=15}^{90} N_{g,i,s}\left(x, t + \frac{m-1}{12}\right) \tau_{g,i,s}(x,t) + V_{g,i,s}\left(x, t + \frac{m-1}{12}\right) \tau'_{g,i,s}(x,t).$$

The latter sum excludes individuals in the acute stage of HIV infection (s = 1), as the model assumes that HIV is (on average) not detectable during the acute phase of HIV infection. The model thus makes an implicit allowance for the sensitivity of the rapid test, assuming that the sensitivity is strongly related to the fraction of testers who are in the acute phase of HIV infection, as has been shown in a recent meta-analysis [399]. However, this sensitivity adjustment is not necessarily equivalent to that described in section B.2, hence the importance of allowing for uncertainty in the latter when calibrating the model to routine HIV testing data. The model estimate of the fraction of adults testing for HIV in year t who test positive is

$$P(t) = \frac{T_1(t) + T_0(t)\left(1 - \mathrm{E}[Sp]\right)}{T_0(t) + T_1(t)},$$

where E[*Sp*] is the same specificity parameter as in section B.2 (0.997). A similar approach is used to calculate the number of antibody tests performed in children under the age of 15, and the HIV prevalence in children tested for HIV. However, as the model does not consider acute HIV infection in children (it would be rare in HIV-positive children over the age of 18 months), there is no exclusion of acutely-infected children, and thus the model implicitly assumes that rapid testing in children has 100% sensitivity.

B.7 Likelihood definition

The likelihood function is defined with respect to two data sources: the proportions of individuals who report having ever tested for HIV in five national household survey (see section B3), and the empirically-derived estimates of the HIV prevalence in individuals tested for HIV in 11 years (see section B2).

Considering first the likelihood in respect of the household survey data, we define $E_{g,x,h}(t)$ to be the model estimate of the proportion of the population ever tested, in individuals of sex *g* and HIV status *h* (0 = negative, 1 = positive), in age group *x*, at time *t*. The corresponding survey estimate is denoted by $\zeta_{g,x,h}(t)$. For the purpose of calibration, the data are grouped into five age categories: 15-24, 25-34, 35-44, 45-59 and 60+ (except in the case of the 2016 DHS, for which we do not have estimates at ages 60+). The likelihood is calculated on the assumption that the differences between the logit-transformed survey estimates and the logit-transformed model estimates are normally distributed with mean U_h. (The logit transformation is used as it helps to ensure that the assumption of normally-distributed error terms is not violated.) In mathematical terms, we assume

$$\log\left(\frac{\zeta_{g,x,h}(t)}{1-\zeta_{g,x,h}(t)}\right) = \log\left(\frac{E_{g,x,h}(t)}{1-E_{g,x,h}(t)}\right) + U_h + \eta_{g,x,h}(t) + \varepsilon_{g,x,h}(t)$$

where U_h is a reporting bias parameter, $\eta_{g,x,h}(t)$ is a model error term and $\varepsilon_{g,x,h}(t)$ is the survey error term, which is assumed to follow a N(0, $\sigma^2_{g,x,h}(t)$) distribution. The variable $\mathrm{K}_h \equiv \exp(U_h)$ is defined as the odds ratio comparing reported prior testing to actual prior testing. The U_h parameter is estimated using the standard maximum likelihood formula,

$$\hat{U}_h = \frac{1}{48}\sum_g\sum_x\sum_t \log\left(\frac{\zeta_{g,x,h}(t)}{1-\zeta_{g,x,h}(t)}\right) - \log\left(\frac{E_{g,x,h}(t)}{1-E_{g,x,h}(t)}\right),$$

and the $\sigma^2_{g,x,h}(t)$ parameter is estimated from the 95% confidence intervals shown in Table B3. The model error term is assumed to follow a N(0, σ^2_E) distribution, with the variance of the model error being estimated as

$$\hat{\sigma}_E^2 = \frac{1}{96}\sum_g\sum_x\sum_t\sum_h\left(\log\left(\frac{\zeta_{g,x,h}(t)}{1-\zeta_{g,x,h}(t)}\right)-\log\left(\frac{E_{g,x,h}(t)}{1-E_{g,x,h}(t)}\right)-\hat{U}_h\right)^2-\sigma^2_{g,x,h}(t).$$

The likelihood function in respect of the household survey data is then

$$\prod_g\prod_x\prod_t\prod_h\left(2\pi\left(\hat{\sigma}_E^2+\sigma^2_{g,x,h}(t)\right)\right)^{-0.5}\exp\left[-\frac{\left(\text{logit}\left(\zeta_{g,x,h}(t)\right)-\text{logit}\left(E_{g,x,h}(t)\right)-\hat{U}_h\right)^2}{2\left(\hat{\sigma}_E^2+\sigma^2_{g,x,h}(t)\right)}\right].$$

Secondly, for the purpose of defining the likelihood function in respect of the HIV prevalence data, suppose that $P(t)$ represents the model-based estimate of HIV prevalence in adults tested for HIV in year t (as defined in section B.6). It is again assumed that the difference between the model estimate and the empirical estimate is normally distributed on the logit scale. In mathematical terms, we assume

$$\log\left(\frac{\theta(t)}{1-\theta(t)}\right)=\log\left(\frac{P(t)}{1-P(t)}\right)+\varepsilon_\theta(t)$$

where $\varepsilon_\theta(t)$ is the error associated with the empirical derivation (due to uncertainty regarding private sector HIV prevalence and test specificity). The $\varepsilon_\theta(t)$ term is assumed to follow a N(0, $\sigma^2_\theta(t)$) distribution, with $\sigma_\theta(t)$ being estimated as $\text{Var}[\theta(t)]^{0.5}/(\text{E}[\theta(t)]\ (1-\text{E}[\theta(t)]))$, the delta approximation to the standard deviation on the logit scale. The likelihood function in respect of the HIV prevalence data is

$$\prod_t\left(2\pi\sigma^2_\theta(t)\right)^{-0.5}\exp\left[-\frac{\left(\text{logit}\left(\theta(t)\right)-\text{logit}\left(P(t)\right)\right)^2}{2\sigma^2_\theta(t)}\right].$$

B.8 Prior distributions

Table B7 summarizes the parameters included in the uncertainty analysis, the prior distributions assigned to these parameters, and the data sources on which these prior distributions are based. Most of these prior distributions are the same as in our previous analysis of HIV testing rates up to 2012 [124]. Most of the parameters have been defined in this appendix, although some have been defined previously in section 3.2. Gamma distributions are used to represent the uncertainty for the first six parameters and the r_1 and r_2 parameters, while a beta distribution is assigned to represent the uncertainty around the φ parameters, and uniform (0, 1) prior distributions are assigned to the remaining parameters.

Table B7: Prior distributions

Parameter	Symbol	Mean	SD	Data source
Mean age of testing: men*	α_0 / λ_0	42	7	Based on fitting gamma functions to age profile of individuals tested for HIV in Mpumalanga [127], using the provincial age profile [339]
Mean age of testing: women*	α_1 / λ_1	24	4	
SD of age of testing: men*	$\sqrt{\alpha_0} / \lambda_0$	30	5	
SD of age of testing: women*	$\sqrt{\alpha_1} / \lambda_1$	24	4	
Ratio of male to female test uptake at age 25 in 2002	$B_0(2002)$	0.8	0.1	2006 National Communication Survey [128]
Ratio of male to female test uptake at age 25 in 2010	$B_0(2010)$	0.8	0.1	-
HIV test history adjustment:				
Never previously tested	r_0	1	-	Definition of base rate
Previously tested negative	$r_1(0)$	1.5	0.4	[53, 129, 130]
Previously tested negative	$r_1(max)$	1.5	0.7	[53, 129, 130]
Diagnosed HIV+, pre-ART	r_2	0.5	0.30	[393, 400-402]
ART-experienced	r_3	0.5	0.29	[393, 400-402]
OR for HIV testing in OI treatment facility	Φ	0.5	0.29	Vague prior

OI = opportunistic infection. OR = odds ratio. SD = standard deviation.

* For a hypothetical population with uniform age distribution. The actual distribution of HIV testing ages differs when the function is applied to the population pyramid in South Africa.

B.9 Comparison of prior and posterior distributions

Posterior distributions were simulated using Incremental Mixture Importance Sampling [368]. The resulting posterior estimates of the model parameters are presented in Table B8. Consistent with our previous analysis, and consistent with a recent study showing high rates of retesting in recently diagnosed sex workers [403], results suggest a high rate of retesting in individuals who have previously been diagnosed HIV-positive, although retesting appears to be substantially less frequent after ART initiation.

Table B8: Comparison of prior and posterior distributions

Parameter	Prior distribution (mean, 95% CI)	Posterior distribution (mean, 95% CI)
Mean age of testing: men*	42.0 (29.4-56.8)	39.2 (33.8-46.6)
Mean age of testing: women*	24.0 (16.8-32.5)	22.5 (17.5-27.1)
SD of age of testing: men*	30.0 (21.0-40.6)	29.5 (22.2-38.8)
SD of age of testing: women*	24.0 (16.8-32.5)	23.8 (19.4-29.9)
Ratio of male to female test uptake in 2002	0.80 (0.62-1.01)	0.84 (0.71-0.99)
Ratio of male to female test uptake in 2010	0.80 (0.62-1.01)	0.76 (0.65-0.87)
HIV test history adjustment:		
Previously tested negative (baseline)	1.50 (0.82-2.38)	1.66 (1.06-2.24)
Previously tested negative (maximum)	1.50 (0.46-3.12)	5.41 (3.86-7.26)
Diagnosed HIV+, pre-ART	0.50 (0.09-1.24)	1.15 (0.74-1.56)
ART-experienced	0.50 (0.025-0.0975)	0.14 (0.07-0.25)
OR for HIV testing in OI treatment facility	0.50 (0.025-0.0975)	0.57 (0.21-0.87)

OI = opportunistic infection. OR = odds ratio. SD = standard deviation.

* For a hypothetical population with uniform age distribution. The actual distribution of HIV testing ages differs when the function is applied to the population pyramid in South Africa.

The posterior estimates of the θ_0 and θ_1 parameters are 1.43 (95% CI: 1.25-1.59) and 0.69 (95% CI: 0.58-0.78) respectively. This suggests that HIV-negative individuals tend to over-report their past HIV testing, while HIV-positive individuals tend to under-report their past HIV testing. A more detailed discussion of these reporting biases is provided elsewhere [124].

B.10 Comparison of model estimates with data

Figure B1 compares the model estimates of the fraction of adults ever tested for HIV with the household survey data (after adjusting the model estimates to take account of the reporting biases noted previously). Although the model is in good agreement with the data from the 2005 and 2012 surveys, the model tends to under-estimate the reported levels of past testing in the 2008 survey and tends to over-estimate the reported levels of past testing in the 2017 survey (especially in women).

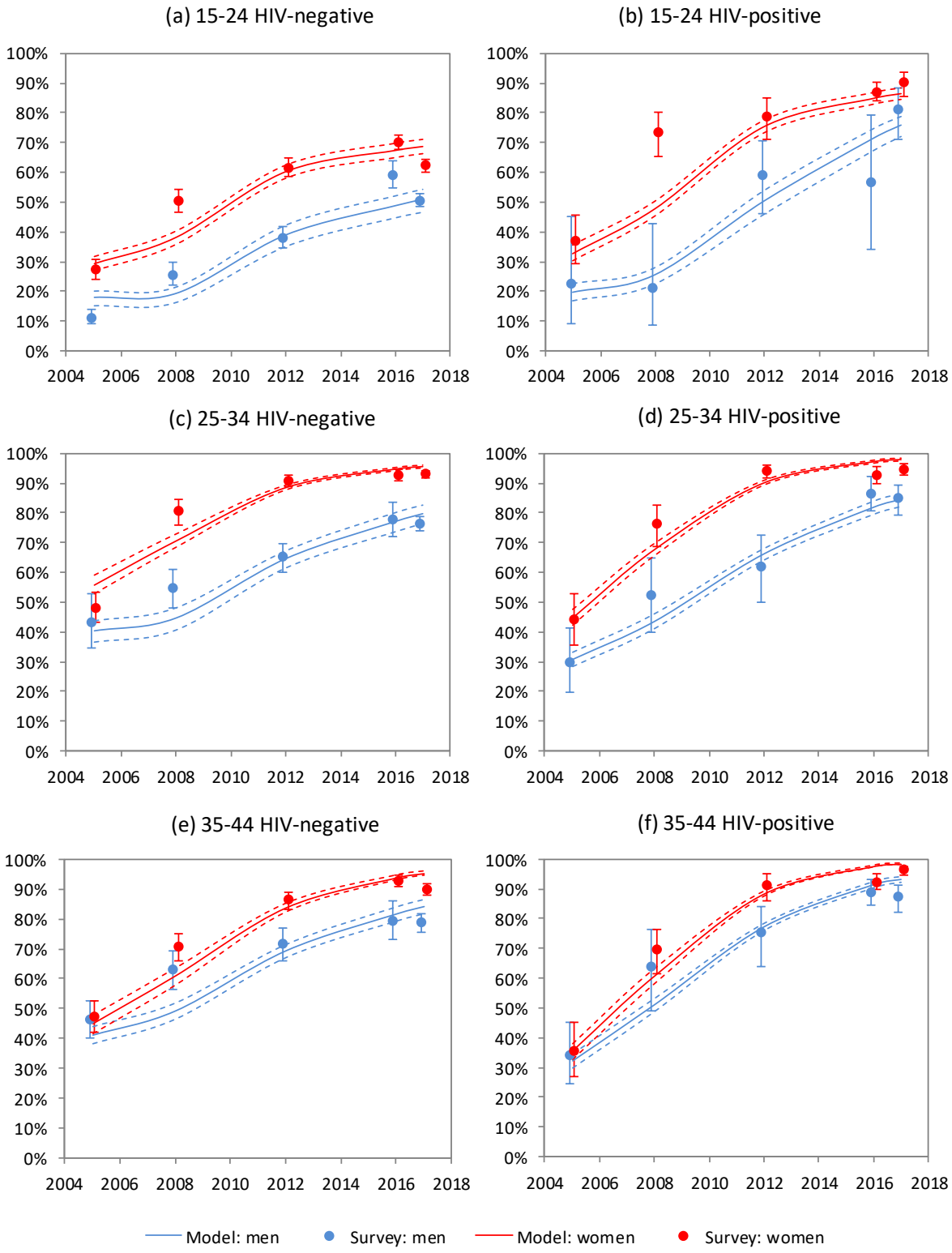


Figure B1: Proportions of adults who report having ever been tested for HIV

Model estimates have been adjusted to reflect expected reporting bias. Dashed lines represent 95% confidence intervals around average model estimates (solid lines), while vertical lines represent 95% confidence intervals around survey estimates (dots).

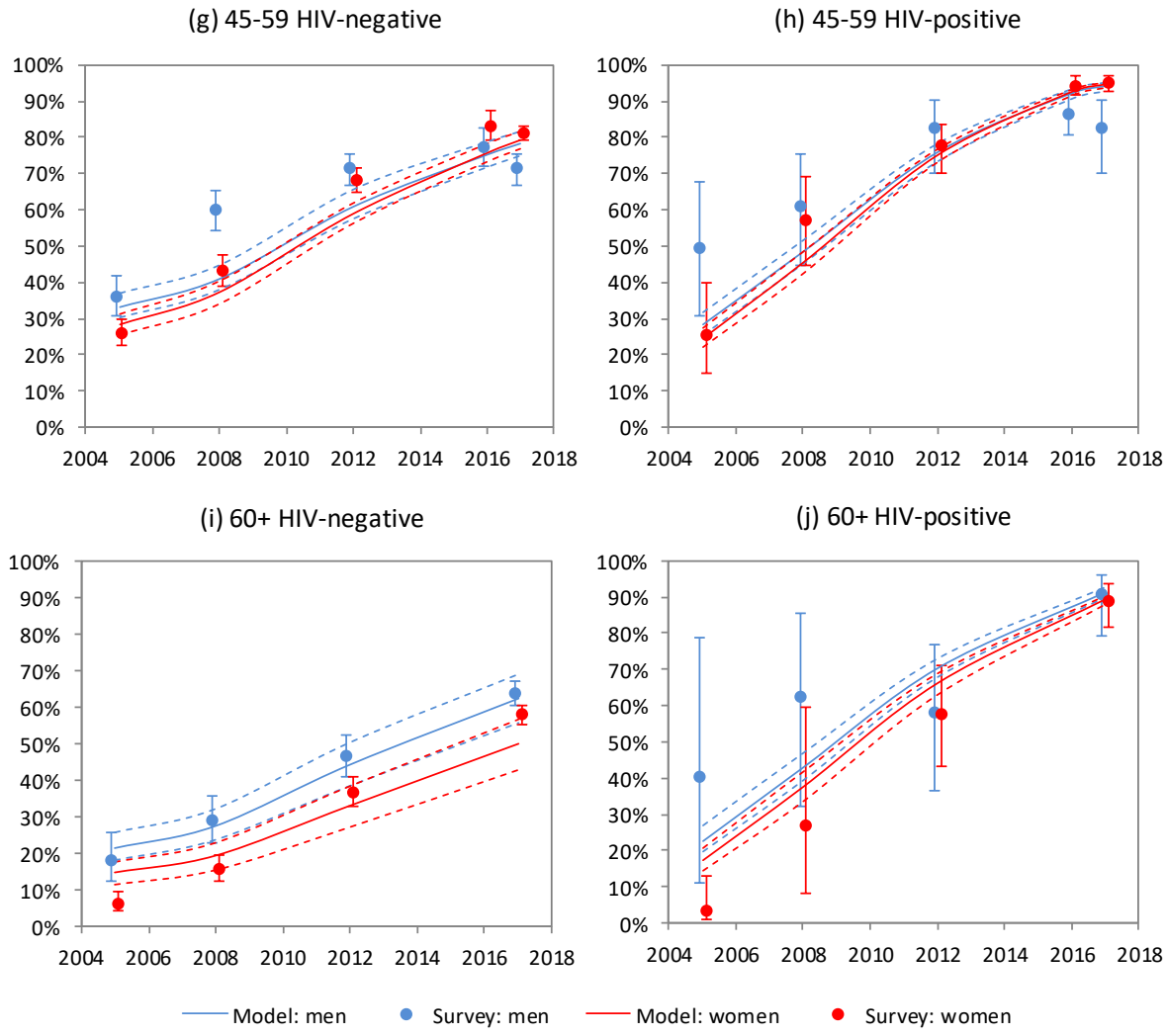


Figure B1 (continued): Proportions of adults who report having ever been tested for HIV
Model estimates have been adjusted to reflect expected reporting bias. Dashed lines represent 95% confidence intervals around average model estimates (solid lines), while vertical lines represent 95% confidence intervals around survey estimates (dots).

Figure B2 compares the model estimates of HIV prevalence in individuals tested for HIV with the corresponding empirically-derived estimates. The model matches the routine testing reasonably closely. The model estimates an increase in prevalence up to 2006-7, reflecting the increasing HIV prevalence in the general population. Thereafter, the model estimates a steady decline in HIV prevalence, in part because of rising levels of diagnosis and ART coverage (since previously diagnosed and treated individuals are less likely to get tested again than those who test negative).

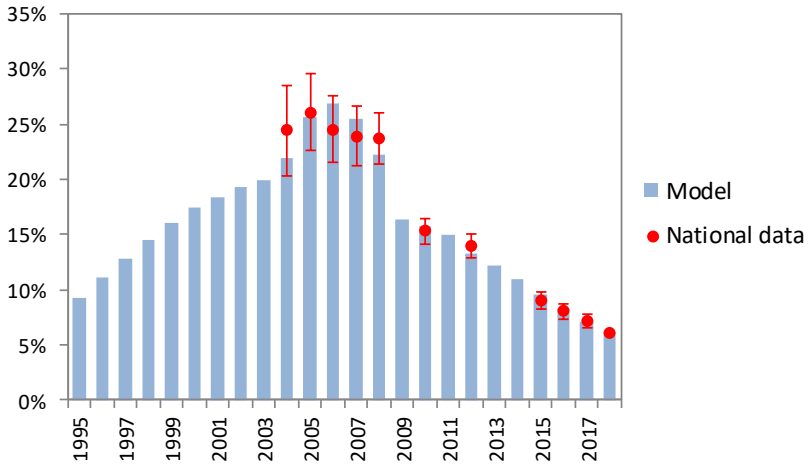


Figure B2: HIV prevalence in adults tested for HIV

Figure B3 compares the model estimates of the total numbers of HIV tests with the empirically-derived totals. There is close agreement between the modelled totals and the empirical estimates, suggesting that any bias due to the assumption of a constant population profile over the course of a year is likely to be minimal. Numbers of HIV tests have been high since the start of national HIV testing campaigns in 2009/10.

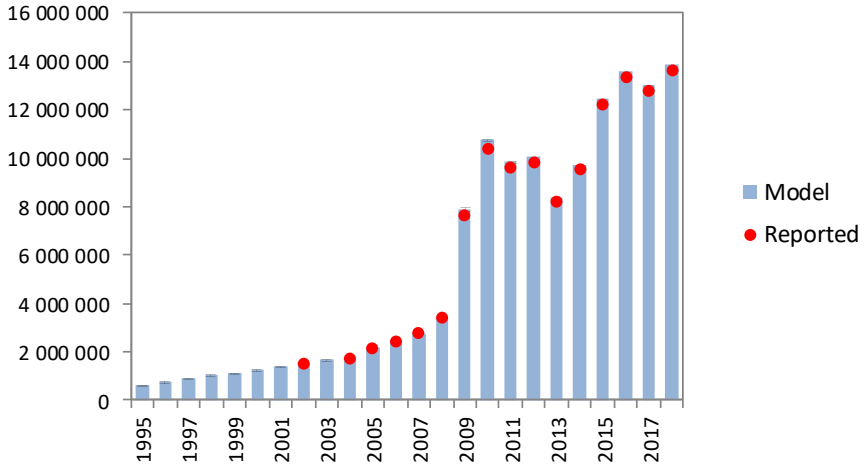


Figure B3: Annual numbers of adult HIV tests

Finally, Figure B4 shows the model estimates of the fraction of HIV-positive testers who are in the acute phase of infection (which is equivalent to 1 – the assumed test sensitivity). In the early phases of the epidemic, this fraction declines over time as HIV incidence declines, but the model estimates that as HIV testing efforts intensify and there is an increasingly high fraction of HIV-positive adults diagnosed, the fraction of HIV-positive testers in the acute phase of infection increases in the more recent years. The fraction in the acute phase of infection ranges between 2.5% and 4.2% over the 2000-2018 period, slightly lower than the rate of 4.7% of positive testers in the acute phase of HIV infection in low-income countries

(4.7%) but slightly higher than the false negative rate in low-income countries (2.3%) in the meta-analysis of Tan *et al* [399].

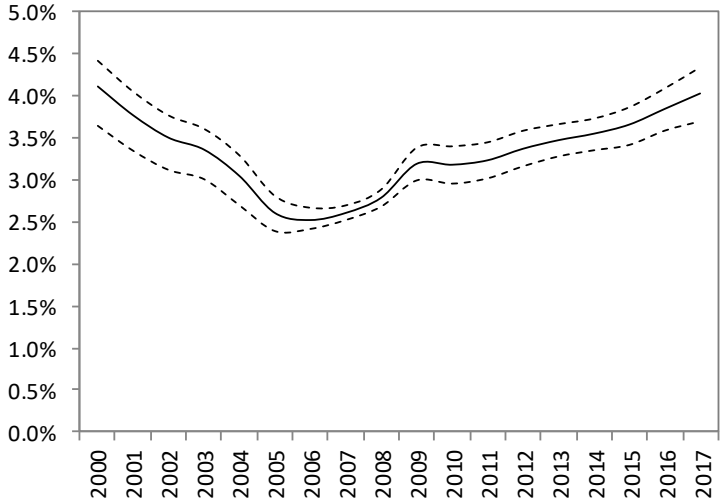


Figure B4: Fraction of HIV-positive testers in the acute phase of HIV infection
Solid line represent model mean; dashed lines represent 95% confidence intervals.

Appendix C: Calibration to HIV prevalence data from key populations

The Thembisa model is fitted to HIV prevalence data from sex workers and MSM, using a Bayesian approach.

C.1 Prior distributions

Prior distributions are assigned to represent the uncertainty in four key parameters. Table C1 summarizes these prior distributions.

The first of these parameters is the male-to-female transmission probability per sex act in sex worker-client interactions. Few studies have estimated the probability of HIV-1 transmission from an infected client to a susceptible sex worker, in African settings. In a study of Senegalese sex workers, Gilbert *et al* estimated that the average probability of HIV-1 transmission per act of sex with an infected client was between 0.00031 and 0.00056, depending on the approach to dealing with missing data on numbers of clients [404]. A similarly low probability of transmission per unprotected sex act with an infected client, 0.00063, was estimated in a cohort of Kenyan sex workers [208]. Data from a South African study of sex workers in KwaZulu-Natal (KZN) can also be used to estimate the probability of transmission per sex act. In this study, an HIV incidence rate of 14.7 per 100 person years was observed [175] in sex workers who reported an average of 23.3 sex acts with clients per week, of which 20.3 were protected [176]. HIV prevalence in truck driver clients was estimated to be 56% [40]. If β is the probability of transmission per act of unprotected sex, and condoms are assumed to reduce this transmission probability by 90% [199], we can crudely estimate the average weekly rate of HIV acquisition as

$$0.56\times\beta\times\left(20.3\times\left(1-0.9\right)+3.0\right).$$

Setting this expression to 0.147/52 and solving for β yields a β estimate of 0.00100. However, the true HIV prevalence in clients is unknown, as truck driver clients might not be typical of clients generally. In a systematic review of HIV risk factors in sub-Saharan Africa, Chen *et al* [405] estimated an average HIV prevalence in sex worker clients of 35%. If the true prevalence in KZN sex worker clients were closer to this average, the estimate of β would be substantially higher. To represent the uncertainty regarding β, we assign a beta prior with a mean of 0.001 (the same as the value from the estimated from the KZN study) and a standard deviation of 0.0005. This prior distribution has a 2.5 percentile of 0.0003 (consistent with the lowest estimate of Gilbert *et al*) and a 97.5 percentile of 0.0022 (a likely upper bound around the transmission probability in the KZN sex worker study).

The second parameter that we consider in the model calibration is the relative rate of client-to-FSW transmission at the start of the HIV epidemic (1985) relative to that in 1995 and subsequent years. The transmission probability specified in the previous paragraph is the parameter that applies in 1995 and subsequent years, but there are a number of reasons why a higher client-to-sex worker transmission probability might be expected in the earlier stages of the HIV epidemic. The most direct evidence of this is the study of Kimani *et al* [208], among Kenyan sex workers, which found a roughly five-fold higher sex worker HIV incidence per

unprotected sex act over the 1985-87 period than the average over the period after 1994, and a roughly linear decline in the acquisition risk between 1987 and 1994. This could have been due to declines in STI prevalence, which followed the same trend as the decline in the HIV transmission risk [208]; similar STI prevalence declines have occurred in South Africa [406]. Alternatively, the reduction in HIV acquisition risk may reflect heterogeneity in HIV acquisition risks among sex workers; many sex workers become infected after only a few exposures to HIV-positive clients, while a minority remain HIV-negative despite long durations of exposure to HIV-positive clients [407]. The reduction in the client-to-FSW transmission risk over the early stages of the HIV epidemic could thus reflect a change in average duration of HIV exposure: the HIV-negative women at the start of the HIV epidemic have only been exposed to HIV for very short durations, whereas the HIV-negative sex workers in the more advanced stages of the HIV epidemic will (on average) have been exposed to HIV for longer durations, and the fact that they have remained HIV-negative means they face a lower HIV acquisition risk, on average. In Thembisa, if β is the average client-to-FSW transmission probability that applies in 1995 and subsequent years, and θ is the factor by which the transmission risk is increased in 1985, the assumed transmission risk in year t is

$$\beta \times (1 + (\theta - 1)\,(1995 - t) / 10),$$

for $t < 1995$. We represent the uncertainty around the θ parameter using a gamma prior with a mean of 5 and a standard deviation of 2, consistent with the Kenyan sex worker study [208].

The third parameter that is allowed to vary in the model calibration is the effect of HIV diagnosis on women's entry into commercial sex. One might expect that a woman who has been diagnosed HIV-positive would be less likely to begin sex work than an HIV-positive woman who is undiagnosed, due to fear of transmitting HIV to others or due to greater concern for her own health. South African studies estimate lower probabilities of HIV diagnosis in HIV-positive sex workers [42, 408] compared to HIV-positive women generally [124], which could be due to a reduced probability of starting sex work following an HIV diagnosis (though other explanations are possible, such as confounding by age and difficulties sex workers may experience in accessing HIV testing). Given the uncertainty regarding the proportionate reduction in the rate of entry into commercial sex after diagnosis, we assign a vague prior, which is uniform on the interval [0, 1), to represent the uncertainty around the reduction.

The fourth parameter for which a prior distribution is assigned is the male-to-male transmission probability per sex act. The only published estimates of these male-to-male transmission probabilities are from high-income settings [181-183], and as heterosexual transmission probabilities in developing countries tend to be higher than those in high-income settings [120], we have chosen a prior distribution for the South African setting with a mean higher than that observed (mean 0.020, standard deviation 0.005). The model does not distinguish transmission probabilities according to the type of sex act.

Table C1: Prior distributions

Parameter	Prior distribution	Mean, standard deviation
Client-to-sex worker transmission probability*	Beta (3.995, 3991)	0.001, 0.0005
Relative rate of client-to-FSW transmission in 1985 (compared to 1995 and later years)	Gamma (6.25, 1.25)	5.0, 2.0
Proportionate reduction in rate of entry into sex work following an HIV diagnosis	Uniform (0, 1)	0.5, 0.29
Male-to-male transmission probability*	Beta (15.66, 767.3)	0.02, 0.005

* Per act of unprotected sex.

C.2 Likelihood definitions: sex worker survey data

Table C2 summarizes the HIV prevalence data from surveys of sex workers, which have been used in model calibration. (We have excluded studies in which HIV status was based on self-report [409, 410], as this has been shown to have low sensitivity in the South African setting [358].) The surveys of commercial sex workers have been conducted in specific communities, and cannot be considered representative of sex workers nationally. It is therefore necessary to allow for potential heterogeneity in HIV prevalence between commercial sex workers surveyed in different communities, using different sampling techniques. We use the notation t_i, n_i and p_i to represent the time of the i^{th} survey, the sample size of the i^{th} survey and the HIV prevalence measured in the i^{th} survey respectively.

Table C2: Studies of HIV prevalence in South African sex workers

Study	Location	Year (t_i)	Sample size (n_i)	Prevalence of HIV (p_i)
Ramjee *et al* [175]	Truck stops between Durban and Johannesburg	1996	416	50%
Dunkle *et al* [47]	Johannesburg	1996	295	46.4%
Leggett *et al* [411]	Johannesburg, Durban, Cape Town	1998*	249	42.6%
Williams *et al* [16]	Carletonville	1998	121	68.6%
Ndhlovu *et al* [412]	Carletonville	2001	101	78%
van Loggerenberg *et al* [50]	Durban	2004	775	59.6%
Luseno & Wechsberg [53]	Pretoria	2005	276	59.1%
Greener *et al* [413]	Durban	2012	349	66.9%
USCF [42]	Johannesburg	2013	764	71.8%
	Cape Town	2013	650	39.7%
	Durban	2013	766	53.5%
Schwartz *et al* [408]	Port Elizabeth	2014	410	61.5%
Black *et al* [414]	Johannesburg	2014	249	75.1%
University of California San Francisco (unpublished)	Johannesburg	2018	542	60.4%
	Cape Town	2018	796	36.5%
	Durban	2018	556	75.5%

* The study date was not stated, and has been assumed to be three years prior to the date of publication, based on average publication delays in other STI prevalence surveys [415].

For the purpose of defining the likelihood function, suppose that $C(t_i, \boldsymbol{\varphi})$ represents the model estimate of HIV prevalence in sex workers in the year of the i^{th} study, where the vector $\boldsymbol{\varphi}$ represents the values of the model input parameters. The difference between the logit-transformed model estimate of HIV prevalence and the logit-transformed observed prevalence is assumed to be composed of a 'random effect' (representing the true difference in HIV prevalence between the HIV prevalence in sex workers nationally and the prevalence

in sex workers in the community being studied) and a 'random error' term (representing the binomial sampling variation due to the limited sample size). More formally, it is assumed that

$$\log\left(\frac{p_i}{1-p_i}\right) = \log\left(\frac{C(t_i,\boldsymbol{\varphi})}{1-C(t_i,\boldsymbol{\varphi})}\right) + r_i + \varepsilon_i, \qquad \text{(C1)}$$

where $r_i \sim N(0,\sigma_r^2)$ and $\varepsilon_i \sim N(0,\sigma_i^2)$. The variance of the random error term, σ_i^2, is estimated by noting that the sample variance of p_i is $p_i(1-p_i)/n_i$, and after logit-transformation, the Taylor approximation to the value of the sample variance of logit(p_i) is

$$\hat{\sigma}_i^2 = \frac{1}{n_i p_i (1-p_i)}. \qquad \text{(C2)}$$

Hence the variance of the random effects term can be estimated using the equation

$$\hat{\sigma}_r^2 = \frac{1}{16}\sum_{i=1}^{16}\left(\log\left(\frac{p_i}{1-p_i}\right) - \log\left(\frac{C(t_i,\boldsymbol{\varphi})}{1-C(t_i,\boldsymbol{\varphi})}\right)\right)^2 - \frac{1}{n_i p_i (1-p_i)}, \qquad \text{(C3)}$$

where 16 is the number of data points used in calibrating the model (Table C2). The likelihood function in respect of the commercial sex worker prevalence data is then

$$L(\mathbf{p}\mid\boldsymbol{\varphi}) = \prod_{i=1}^{16}\left(2\pi\left(\hat{\sigma}_r^2+\hat{\sigma}_i^2\right)\right)^{-0.5}\exp\left[-\frac{\left(\text{logit}(p_i)-\text{logit}(C(t_i,\boldsymbol{\varphi}))\right)^2}{2\left(\hat{\sigma}_r^2+\hat{\sigma}_i^2\right)}\right], \qquad \text{(C4)}$$

where **p** is the vector of p_i values.

C.3 Likelihood definitions: MSM survey data

Table C3 summarizes the HIV prevalence data from surveys of men who have sex with men (MSM), which have been used in model calibration. Only studies that used respondent-driven sampling (RDS) have been used for calibration (i.e. excluding venue-based sampling studies, which tend to be biased towards recruitment of higher-risk MSM). Since none of the surveys are nationally representative, the approach used in defining the likelihood function is the same as for sex worker HIV prevalence data, i.e. based on a random effects model to represent heterogeneity between studies. The only difference is that the model estimates of HIV prevalence have been adjusted so that each model estimate of HIV prevalence in MSM is age-standardized to correspond to the age profile in the survey of MSM (as represented by the fraction of sampled MSM who are aged 18-24, shown in the last column of Table C3). This age standardization is necessary because South African surveys of MSM appear to be biased toward younger MSM, and without appropriate age adjustment, this may lead to the model under-estimating HIV prevalence in MSM [416].

Table C3: Studies of HIV prevalence in South African MSM

Study	Location	Year (t_i)	Sample size (n_i)	Prevalence of HIV (p_i)	% aged 18-24
Cloete *et al* [8]	Cape Town	2012	286	22.3%	67.3%
	Durban	2012	290	48.2%	27.0%
	Johannesburg	2012	349	26.8%	52.1%
Lane *et al* [9]	Gert Sibande district	2012	307	29.4%	29.6%
	Ehlanzeni district	2012	298	15.9%	28.0%
Lane *et al* [10]	Soweto	2008	363	13.6%	31.0%
Rispel *et al* [417]	Johannesburg	2008	202	49.5%	33.3%
	Durban	2008	69	27.5%	33.3%
Tucker *et al* [418]	Cape Town	2010	171	34.5%	42.9%
Sandfort *et al* [382]	Pretoria	2011	480	30.1%	41.0%
Kufa *et al* [419]	Johannesburg	2015	546	43.4%*	45.2%*
	Bloemfontein	2015	525	17.3%*	71.6%*
	Mafikeng	2015	474	14.6%*	71.2%*
	Polokwane	2015	358	22.4%*	59.7%*
Sandfort *et al* [359]	Cape Town	2015	139	31.7%	74.1%
	Johannesburg	2015	171	49.1%	67.8%
Fearon *et al* [420]	Johannesburg	2017	300	37.5%	48%

* Unpublished data. RDS-weighted estimates are not yet available (although an RDS design was used), so only the unweighted estimates are used. Variance estimates were calculated using the RDS design effects described previously.

C.4 Model results

Table C4 compares the prior and posterior estimates of the model parameters. The posterior estimate of the male-to-male HIV transmission probability is substantially higher than the corresponding prior estimate. The Bayesian analysis also suggests that HIV-diagnosed women are on average 52% less likely to begin sex work than HIV-positive women who are undiagnosed.

Table C4: Comparison of prior and posterior distributions

Parameter	Prior distribution (mean, 95% CI)	Posterior distribution (mean, 95% CI)
Client-to-sex worker transmission probability per sex act	0.001 (0.0003-0.0022)	0.0011 (0.0007-0.0018)
Relative rate of client-to-FSW transmission in 1985 (compared to 1995 and later years)	5.00 (1.88-9.62)	5.52 (2.77-9.54)
Proportionate reduction in rate of entry into sex work following an HIV diagnosis	0.500 (0.025-0.975)	0.515 (0.114-0.826)
Male-to-male transmission probability per sex act	0.020 (0.011-0.031)	0.029 (0.025-0.035)

Figure C1 compares the model estimates of HIV prevalence in sex workers and MSM with the levels of HIV prevalence measured in South African surveys. As none of the key population surveys is nationally representative, some degree of divergence between model estimates and survey estimates is to be expected. Survey estimates of HIV prevalence are highly variable, reflecting variation in geographic locations and sampled populations. Due to the absence of any HIV prevalence surveys amongst MSM in the early stages of South Africa's HIV epidemic, confidence intervals around the model estimates of HIV prevalence in MSM during the 1990s are very wide.

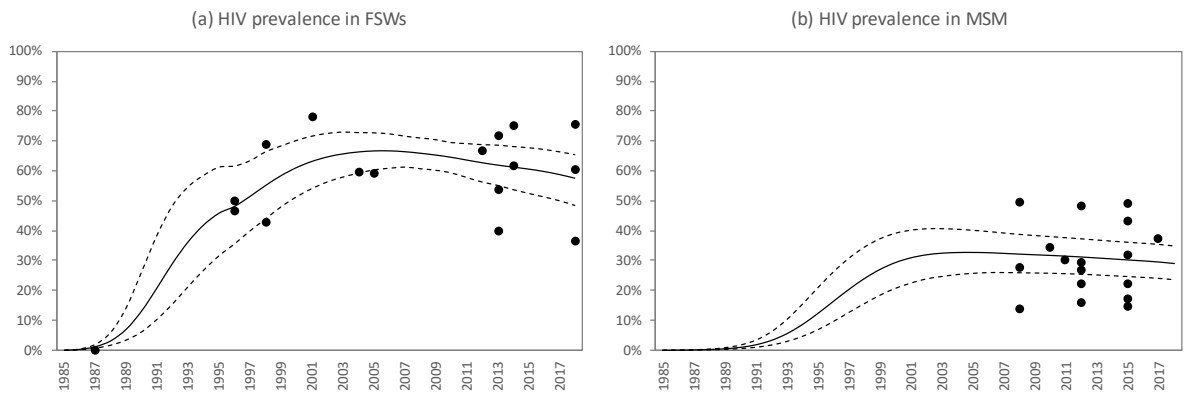


Figure C1: HIV prevalence in key populations

Dots represent survey prevalence estimates (the 1987 data point shown in panel (a) was not included in the definition of the likelihood as the survey found no HIV in sex workers [421], and the likelihood for this observation was thus undefined). Solid lines represent the posterior mean model estimates of HIV prevalence, and dashed lines represent the 95% confidence intervals around these estimates.

Appendix D: The impact of ART on mortality in children

Section 5.3 sets out the general structure of the model of paediatric HIV disease progression, mortality and ART. This appendix provides more detail on two aspects of the model parameterization: the effects of changes over time in rates of ART initiation in late disease (section D.1) and the estimation of mortality rates in children who start ART in early disease (section D.2).

D.1 Modelling the effect of ART on mortality in late disease

A limitation of the Thembisa model is that it groups HIV-positive children with advanced disease into a single state, and does not allow for the possibility that there may be significant heterogeneity in mortality within this state. This becomes particularly problematic when modelling the impact of ART on mortality, as changes in ART uptake lead to changes in the CD4 distributions of untreated children as well as changes in the CD4 distributions of children starting ART. This in turn means that mortality rates in ART-naïve and treated children change as ART uptake increases. To address this challenge, we develop a simple heuristic to adjust the base model assumptions to take account of the effect of ART. This heuristic procedure is very similar to the approach developed for adults, as described previously (see Appendix A of the Thembisa analysis of adult mortality trends in South Africa [94]).

To describe pre-ART mortality at CD4 percentages below 15% (which we will use here as a rough approximation to 'advanced disease'), we assume that the untreated HIV mortality rate in children with CD4 % x is

$$\mu(x) = a\left(b^{x}\right), \qquad \text{(D1)}$$

where a is the mortality rate we would expect in an untreated individual with a CD4 % of zero, and b is the factor by which the mortality rate decreases for a 100% increase in the CD4 % (i.e. ab is the theoretical mortality rate at a CD4 % of 100%). The b parameter is estimated by fitting regression models of the form given in equation (D1) to average mortality levels reported over different CD4 ranges, in different age groups, as reported in a meta-analysis of paediatric survival studies conducted in resource-limited settings prior to the availability of ART [322]. The resulting model fits to the data are shown in Fig D1. Estimates of the b parameter are higher at the younger ages (0.0000305 and 0.0000292 at ages 1-2 and 3-4 respectively) than at the older ages (0.0000005 and 0.0000006 at ages 5-6 and 7 or older, respectively). We set the b parameter in our model at the average of these values, 0.0000153.

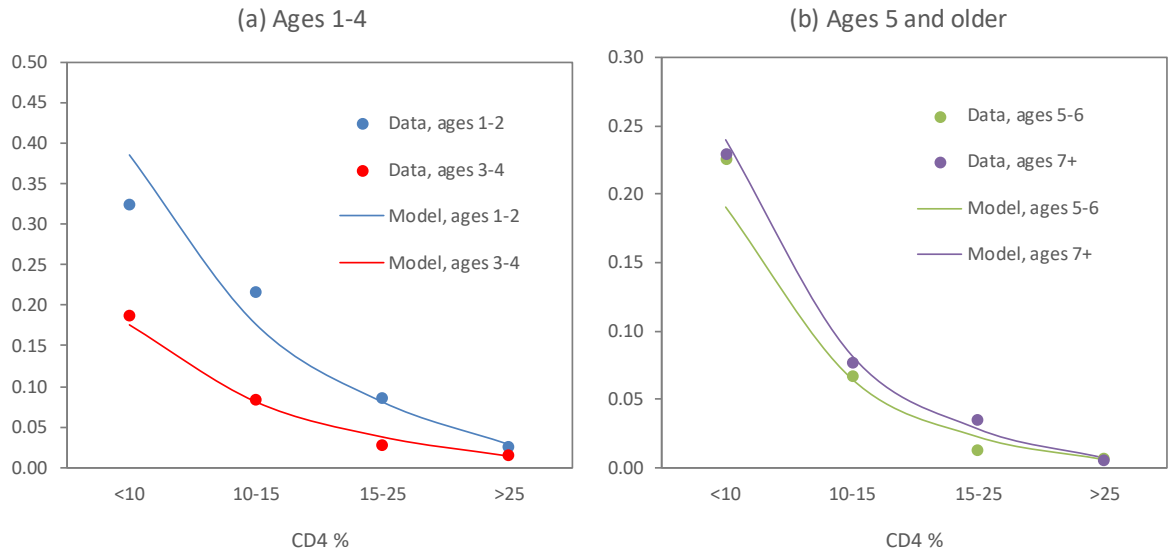


Fig D1: Effect of CD4 count on mortality in the absence of ART

For the purpose of fitting the models to the data points, the average mortality rates reported over different ranges have been taken to apply at the midpoints of the relevant ranges. Mortality data from the >15% CD4 range have been included in order to increase the statistical confidence in the fitted parameters.

Further suppose that $f(x)$ is the distribution of CD4 percentages in a theoretical ART-naïve population, in children who have progressed to advanced disease (which we approximate as CD4 <15% for the sake of simplicity). We assume $f(x) = ke^{\lambda x}$, for $0 \leq x \leq 0.15$, where k is a constant that must satisfy the condition

$$\int_0^{0.15} ke^{\lambda x} dx = 1, \qquad \text{(D2)}$$

from which it follows that $k = \lambda(e^{\lambda \times 0.15} - 1)^{-1}$. In order to estimate the λ parameter, we need to know something about the distribution of CD4 percentages in untreated children with advanced HIV disease. From the same meta-analysis described previously [322], we estimate that roughly 55% of all children who have CD4 percentages <15% have a CD4 <10%. From this it follows that

$$0.55 = \frac{e^{0.1\lambda} - 1}{e^{0.15\lambda} - 1}, \qquad \text{(D3)}$$

and solving this equation gives an estimate of $\lambda = 6.79$.

Having estimated the distribution of CD4 percentages in an ART-naïve population of children with advanced HIV disease, it is possible to estimate the average mortality rate, q_0, as

$$\begin{aligned} q_0 &= \int_0^{0.15} f(x) a b^x dx \\ &= \frac{\lambda a \left(\left(e^{\lambda} b \right)^{0.15} - 1 \right)}{\left(\lambda + \ln(b) \right) \left(e^{0.15\lambda} - 1 \right)}. \end{aligned} \qquad \text{(D4)}$$

Now suppose that q_t is the annual mortality rate in untreated children with CD4 <15%, in year t (this is analogous to the $\mu(x)$ parameter in section 5.3, although for the purpose of this description we are ignoring age effects on mortality). Further suppose that $q_{\min}$ is the minimum mortality rate that we would expect if rates of ART initiation were at their maximum. We would expect q_t to decline as the rate of ART initiation increases, as high rates of ART initiation imply that few individuals will progress to very low CD4 values without starting ART. In modelling q_t we assume it is exponentially related to r_{t-}, the average rate of paediatric ART initiation over the previous three years, subject to the maximum of q_0 and the minimum of $q_{\min}$:

$$q_t = q_{\min} + (q_0 - q_{\min})\exp(-mr_{t-}), \tag{D5}$$

where m is the assumed exponential parameter. This can be written as

$$A_t = \frac{q_{\min}}{q_0} + \left(1 - \frac{q_{\min}}{q_0}\right)\exp(-mr_{t-}), \tag{D6}$$

where $A_t \equiv q_t / q_0$ is an adjustment factor applied to the mortality rate that would be expected in the absence of any ART rollout. The ratio $q_{\min}/q_0$ can be estimated by noting that untreated mortality is at a minimum when all children start ART soon after their CD4 drops below 15%, i.e. $q_{\min} = ab^{0.15}$. From this it follows that

$$\frac{q_{\min}}{q_0} = \frac{b^{0.15}(\lambda + \ln(b))\left(e^{0.15\lambda} - 1\right)}{\lambda\left(\left(e^{\lambda}b\right)^{0.15} - 1\right)}. \tag{D7}$$

Substituting λ = 6.79 and b = 0.0000153 into this equation gives a $q_{\min}/q_0$ estimate of 0.45, which is slightly higher than the corresponding value of 0.31 previously estimated for adults [94].

The m parameter in equation (D6) is difficult to quantify precisely, so a Bayesian approach is adopted to reflect the uncertainty regarding this parameter. Given the lack of information for children, we assign the same prior distribution to this parameter as assumed for adults, viz. a gamma distribution with a mean of 7.5 and a standard deviation of 3.5.

So far we have considered only mortality in untreated children. A similar approach is adopted in modelling the effect of ART-related changes in CD4 distributions on mortality during the first 6 months after starting ART. Suppose that v_t is the annual mortality rate in adults during their first 6 months after starting ART (with baseline CD4 <15%), in year t (this is analogous to $\Phi_0\mu(x)$ in section 5.3, although this is defined for the first 3 months after ART initiation and is age-dependent). Further suppose that v_0 is the corresponding mortality rate that would have been expected in the very early stages of the ART rollout, when rates of ART initiation were very low, and that $v_{\min}$ is the minimum mortality rate that we would expect if rates of ART initiation were at their maximum. We would expect v_t to decrease as the rate of ART rollout increases, as higher rates of ART rollout should lead to higher baseline CD4 counts. As before, we assume a relationship of the form

$$B_t = \frac{v_{\min}}{v_0} + \left(1 - \frac{v_{\min}}{v_0}\right)\exp\left(-mr_{t-}\right), \qquad \text{(D8)}$$

where $B_t \equiv v_t / v_0$. Note that the *m* parameter is assumed to be the same as that in equation (D6), although one could argue that the relationship with the rate of ART initiation may differ depending on whether one is considering pre-ART mortality or treated mortality. (In the interests of obtaining a parsimonious model fit, we use the same parameter value in equations (D6) and (D8), but the model does allow for different values to be assumed.)

For the purpose of estimating the ratio $v_{\min}/v_0$, we will assume that the mortality rate for individuals with baseline CD4 counts of *x*, $v(x)$, is of the form

$$v(x) = z\left(h^x\right). \qquad \text{(D9)}$$

We fit this model to data from a pooled analysis of paediatric ART programmes in South Africa, collected at a relatively early stage in the paediatric ART rollout [324]. After controlling for age, the mortality risk was found to reduce by a factor of 0.895 for each percentage point increase in the baseline CD4 count; this suggests an *h* value of 0.895^{100} = 0.0000144. Of children starting ART with CD4 <15%, 25% had a CD4 <5%, 33% had a CD4 of 5-9% and 42% had a CD4 of 10-14%. Taking this to be the baseline CD4 distribution that would be expected in the early stages of the paediatric ART rollout (i.e. when rates of ART uptake are low), and assuming a roughly uniform distribution of CD4 values within each CD4 category, we can approximate the ratio $v_{\min}/v_0$ as

$$\frac{v_{\min}}{v_0} = \frac{zh^{0.15}}{z\left(0.25h^{0.025} + 0.33h^{0.075} + 0.42h^{0.125}\right)}. \qquad \text{(D10)}$$

Substituting *h* = 0.0000144 into this equation gives a $v_{\min}/v_0$ estimate of 0.43, similar to the estimate of 0.39 previously obtained using adult data [94].

Finally, we define $w(x)$ to be the mortality rate that would be expected at durations >6 months after ART initiation, in children who started ART with a CD4 count <15% (this is analogous to $\Phi_1\mu(x)$ in section 5.3). Similar to the approach adopted with $v(x)$, we assume $w(x)$ can be related to the baseline CD4 % by the equation $w(x) = p(s^x)$, and we estimate the *s* parameter by fitting a regression model to the same paediatric ART dataset as described previously [324]. After controlling for age and ART duration, this regression model estimates that the mortality rate reduces by a factor of 0.938 for each percentage point increase in baseline CD4 %. This is equivalent to an *s* value of 0.938^{100} = 0.00175.

We define w_t to be the annual mortality rate in children in year *t*, who have been on ART for durations >6 months, having started ART with an initial CD4 <15%. As with v_t, we would expect this rate to decline with respect to *t* as rates of ART initiation increase. However, we would expect the decline in w_t to be more moderate than that in v_t, since mortality at longer ART durations is not as strongly related to baseline CD4 count as mortality at early ART durations. We define a relation between w_t to and r_{t-} similar to that in equation (D5):

$$w_t = w_{\min} + (w_0 - w_{\min})\exp(-mr_{t-}), \quad \text{(D11)}$$

where w_0 is the mortality rate that would have been expected in the very early stages of the ART rollout, and $w_{\min}$ is the minimum mortality rate that we would expect if rates of ART initiation were at their maximum. As before, we define $C_t \equiv w_t / w_0$, so that

$$C_t = \frac{w_{\min}}{w_0} + \left(1 - \frac{w_{\min}}{w_0}\right)\exp(-mr_{t-}). \quad \text{(D12)}$$

For the purpose of estimating the ratio $w_{\min}/w_0$, we use the same baseline CD4 distribution as before for the scenario in which there is limited ART rollout:

$$\frac{w_{\min}}{w_0} = \frac{ps^{0.15}}{p\left(0.25s^{0.025} + 0.33s^{0.075} + 0.42s^{0.125}\right)}. \quad \text{(D13)}$$

Substituting s = 0.00175 into this equation gives a $w_{\min}/w_0$ estimate of 0.63, close to the value of 0.61 previously estimated using adult data [94].

A limitation of the approach described above is that it considers only the uncertainty in the m parameter. There is also uncertainty regarding the $q_{\min}/q_0$, $v_{\min}/v_0$ and $w_{\min}/w_0$ ratios.

D.2 Modelling mortality after ART initiation in early disease

To assess the mortality rates in children who start ART in early disease, we analysed data from the South African ART programmes that participate in the International epidemiology Databases to Evaluate AIDS (IeDEA) collaboration. Patients were included if they had a recorded date of starting triple-drug ART, were aged less than 15 years at the time of ART initiation, had a baseline CD4 count or CD4 percentage at the time of ART initiation (within 6 months before and 1 weeks after ART initiation), and were in the 'early' phase of HIV disease at the time of ART initiation. Children were considered to be in the early phase of HIV disease if (to be consistent with the model definition in section 5.3) they did not meet any of the criteria for defining ART eligibility in the WHO 2006 paediatric ART guidelines [311], i.e. they were not in WHO clinical stage IV, and were either aged older than 2 years with a baseline CD4% ≥15%, aged 1-2 years with a baseline CD4% ≥20%, or aged <1 year with a baseline CD4% ≥25%. (In cases where only an absolute baseline CD4 count was available, children were instead classified as being in early disease if they were either aged older than 4 years with a baseline CD4 count of ≥200 cells/μl, aged 3-4 years with a baseline CD4 count of ≥350 cells/μl, aged 1-2 years with a baseline CD4 count of ≥750 cells/μl or aged <1 year with a baseline CD4 count of ≥1500 cells/μl.) The exclusion of children who did not meet the immunological criteria for defining ART eligibility at the time they started ART meant the exclusion of all children who started ART before 2008 (since children starting ART before this time only qualified for ART if they were in 'late' disease), the exclusion of children who started ART before 2012 if they were older than 1 year at ART initiation (since South African guidelines changed to allow for early ART initiation in children under the age of 5 in 2012 [153]), and the exclusion of children who started ART

before 2016 if they were older than 5 years (since the South African guidelines changed to recommend universal ART eligibility in 2016).

Children were classified as dead if there was a death and date of death recorded. In a subset of children for whom civil ID numbers were available, patient records were linked to the National Population Register (NPR), and children were also recorded as dead if a death was recorded on the NPR (in cases where dates of death were recorded on both the NPR and the patient record system, the date of death on the NPR was taken to be the 'true' date of death). For children in whom no death was recorded, follow-up was censored at the time of the last patient contact or (in the case of children who transferred out of the service), the date of transfer.

After applying the exclusion criteria, 2 828 children were included in the analysis. Table D1 summarizes the characteristics of these children at the age of ART initiation. Relatively few of the children included in the analysis started ART in the 5-14 age group, as guidelines only changed to recommend early ART initiation in this group in 2016.

Table D1: Baseline characteristics of children starting ART in early disease

	%	n
Age at ART initiation		
0-4	85.4%	2414
5-9	6.2%	175
10-14	8.5%	239
Sex		
Male	45.9%	1299
Female	54.1%	1529
Year of ART initiation		
2008-2011	23.4%	662
2012-2015	55.4%	1566
2016-2017	21.2%	600
Prior MTCT drug exposure		
Yes	15.9%	449
No	72.1%	2036
Not recorded	12.0%	340
Baseline CD4%		
15-24%	14.0%	395
25-34%	36.4%	1028
35-44%	16.3%	462
45% or more	9.0%	255
Not recorded	24.3%	688

In total, 51 deaths were recorded over 5814 person years of follow-up, implying a mortality rate of 0.88 per 100 person years (Table D2). The mortality rate was highest in the first year of life (4.57 per 100 person years), and lower in the 1-4 and 5-9 age groups (0.41 and 0.16 per 100 person years, respectively). No deaths were observed in children aged 10-14. Since cause of death was not reliably recorded for most children, and since non-HIV causes are likely to account for a substantial fraction of the deaths (especially in infants), we fit a function of the following form to the recorded death data:

$$\psi(x) = \mu_0(x) + \beta P^x,$$

where $\psi(x)$ is the all-cause mortality rate at age x (in years), $\mu_0(x)$ is the non-HIV mortality rate at age x, β is the HIV-related mortality rate that applies immediately after birth, and P is the factor by which the HIV-related mortality reduces per year of age (as defined in section 5.3). Non-HIV mortality rates are taken from the Thembisa model (using the model assumptions for 2010, and averaging across the male and female rates). The equation is used to calculate the expected number of deaths, for the relevant number of person years, at each age in months. With values of β = 0.06 and P = 0.2, this model produced estimates of mortality roughly consistent with the IeDEA data (Table D2). With these parameters, the model estimates that most deaths in infants who started ART in early disease are HIV-related, but at ages 1 year and older, most of the deaths in children who started ART in early disease are due to non-HIV causes.

Table D2: Mortality rates in children starting ART in early disease

Age group	Person years	Deaths	Mortality rate (per 100 PY)	Modelled mortality rate (per 100 PY)		
				Non-HIV	HIV	Total
0	743.8	34	4.57	1.66	2.53	4.20
1-4	3694.7	15	0.41	0.25	0.22	0.47
5-9	1278.2	2	0.16	0.09	0.00	0.09
10-14	96.9	0	0.00	0.06	0.00	0.06
Total	5813.6	51	0.88	0.39	0.47	0.86

A limitation of this approach is that it does not allow for possible under-reporting of deaths in the IeDEA cohorts. Analyses of adult mortality data from IeDEA cohorts have shown that mortality rates may be substantially under-estimated in the absence of linkage to data from vital registration systems [367]. Since ID numbers were only available for a subset of children, our estimates may therefore under-estimate the true mortality rates in children on ART. However, we have also shown that child mortality does not appear to be under-estimated as substantially as adult mortality in the absence of vital registry linkage [367], and the bias may therefore be small. Another limitation is that the average duration of follow-up is relatively short (2 years), which may lead to mortality rates being over-estimated (since mortality tends to decline as treatment duration increases). The relatively small sample size also prevents the estimation of mortality rates with a high degree of precision, especially in older HIV-positive children.

Appendix E: Calibration to paediatric HIV data

The model is calibrated to paediatric HIV prevalence data, all-cause mortality data, HIV testing and diagnosis data and ART data. In the previous version of Thembisa, we calibrated to paediatric HIV diagnosis data is a separate step, but in the new version of the model, the calibration to all paediatric HIV data sources is fully integrated. This appendix summarizes the results of a recent paper that describes the Thembisa paediatric HIV model and evidence synthesis approach [6].

E.1 Prior distributions

Prior distributions have been specified for 17 of the mother-to-child transmission and paediatric HIV survival parameters. Table E1 summarizes the prior distributions; the justification for the choice of most prior distributions has been presented in section 5, except in the case of the relative fertility of women who have been diagnosed HIV-positive (compared to undiagnosed HIV-positive women), which influences the total number of births to HIV-positive women, and the relative rate of recording of deaths in undiagnosed and HIV-negative children (explained in the sections that follow). These parameters and prior distributions are the same as described previously [6].

Table E1: Prior distributions for parameters considered in calibration to paediatric HIV data

Parameter	Prior distribution	Prior mean, std. deviation
Relative rate of fertility in HIV-diagnosed untreated women	Beta (6.800, 2.914)	0.70, 0.14
Probability of MTCT from chronically-infected mothers, per year of mixed feeding	Beta (26.83, 164.8)	0.14, 0.025
Probability of MTCT from acutely- infected mothers, per month of mixed feeding	Beta (23.73, 124.6)	0.16, 0.03
Ratio of postnatal transmission risk per month of EBF to postnatal transmission risk per month of mixed feeding	Beta (5.056, 5.056)	0.50, 0.15
Children infected at/before birth		
Annual rate of progression to late disease in older children	Gamma (16.00, 40.00)	0.40, 0.10
Excess annual rate of progression to late disease in neonates	Gamma (16.00, 8.00)	2.00, 0.50
Excess progression reduction factor, per year of age	Beta (4.44, 13.13)	0.40, 0.10
Relative rate of progression to late disease if infected postnatally	Beta (3.189, 5.922)	0.35, 0.15
Children in late disease, untreated		
Annual rate of AIDS mortality in older children	Gamma (16.00, 133.3)	0.12, 0.03
Excess annual rate of AIDS mortality in neonates	Gamma (25.0, 7.14)	3.50, 0.70
Excess AIDS mortality reduction factor, per year of age	Beta (3.00, 12.00)	0.20, 0.10
HIV testing rates		
Relative rate of testing in virgin adolescents: 2005	Beta (3.00, 12.00)	0.20, 0.10
Relative rate of testing in virgin adolescents: 2010	Beta (3.00, 12.00)	0.20, 0.10
Relative rate of testing in early disease (relative to late disease)	Uniform (0.00, 1.00)	0.50, 0.29
Effect of ART on mortality		
Relative rate of mortality in ‘stable’ ART phase compared to untreated children with late disease	Beta (3.50, 31.50)	0.10, 0.05
Reduction in mortality* (on log scale) per unit increase in rate of ART initiation (in late disease) over last 3 years	Gamma (4.59, 0.612)	7.5, 3.5
Relative rate of recording of deaths in undiagnosed and HIV-negative children, relative to HIV-diagnosed	Uniform (0.00, 1.00)	0.50, 0.29

ART = antiretroviral treatment; EBF = exclusive breastfeeding; MTCT = mother-to-child transmission.

E.2 Likelihood function: HIV prevalence data

The method used to calculate the likelihood in respect of the paediatric HIV prevalence data has been described previously [258], but has been updated to include data from the 2012 and 2017 national household prevalence surveys [4, 125]. As with adults, we have relied on data from the 2005, 2008, 2012 and 2017 household surveys conducted by the HSRC. For each survey year, sex and age group, the difference between the model estimate of prevalence and survey estimate of prevalence, on the logit scale, is assumed to be normally distributed with zero mean and standard deviation calculated from the reported 95% confidence interval. The likelihood calculation is thus similar to that described in section 7.2.1, but with the omission of the bias and model error terms.

E.3 Likelihood function: all-cause mortality data

Suppose that $N_g(x, t)$ represents the model estimate of the number of deaths (due to all causes) in children of sex g, in age group x, in year t. Let $R_g(x, t)$ be the corresponding number of recorded deaths, as reported by Statistics South Africa [361]. In this analysis, we consider four age groups: <1 year (infants), 1-4 years, 5-9 years and 10-14 years. We consider deaths over the 1997-2016 period.

Let $c_g(x, t)$ be the completeness of death recording, i.e. the fraction of deaths that we would expect to be recorded. The completeness rates are assumed to be the same as assumed in the most recent Rapid Mortality Surveillance report [422], and are shown in Table E2. Completeness is assumed to be the same in boys and girls. Completeness has generally increased over time, although in infants there appears to have been a slight deterioration in completeness since 2011. Completeness also tends to be higher in older children than in younger children, though infants are an exception [367, 423]: most of these infant deaths occur in health facilities, and special interventions (such as the Child Healthcare Problem Identification Programme) have been established to improve the recording of these facility-based deaths [424, 425].

Table E2: Assumed fractions of child deaths that are recorded

	Aged < 1 year	Aged 1-4 years	Aged 5-9 years	Aged 10-14 years
1997	53.3%	38.0%	54.4%	71.7%
1998	61.5%	47.3%	61.2%	75.5%
1999	63.2%	46.3%	61.4%	76.8%
2000	62.7%	47.3%	62.0%	70.1%
2001	62.9%	46.7%	63.9%	73.1%
2002	67.3%	49.7%	66.0%	75.7%
2003	71.8%	53.4%	68.0%	78.1%
2004	76.2%	58.0%	70.0%	80.2%
2005	80.6%	60.1%	71.8%	82.3%
2006	85.0%	63.8%	73.6%	84.1%
2007	85.0%	63.0%	75.1%	85.9%
2008	85.0%	63.6%	76.7%	87.5%
2009	85.0%	63.0%	78.1%	88.9%
2010	85.0%	64.6%	79.5%	90.1%
2011	85.0%	64.6%	80.8%	91.0%
2012	82.0%	64.0%	82.0%	92.1%
2013	75.5%	63.6%	83.2%	93.0%
2014	75.5%	63.2%	84.2%	94.0%
2015	75.5%	62.9%	85.3%	94.7%
2016	75.5%	63.4%	86.2%	95.3%

For the purpose of defining the likelihood function, we assume that the difference between the log-transformed recorded number of deaths and the log-transformed model estimate of deaths (after application of the completeness adjustment) is normally distributed with zero mean and a variance of σ^2. More formally,

$$\ln\left(R_g(x,t)\right) = \ln\left(N_g(x,t)c_g(x,t)\right) + \varepsilon_g(x,t)$$

where $\varepsilon_g(x,t) \sim N(0, \sigma^2)$, i.e. similar to the approach adopted in defining the likelihood when the model is calibrated to the adult mortality data (equation 7.6). The calculation of the variance σ^2 is similar to that for adults (equation 7.7).

E.4 Likelihood function: Paediatric ART data

The age profile of children on ART is important because it provides indirect information about the fraction of children who start ART soon after HIV diagnosis. The data used in the calibration of the model are the reported fractions of children on ART in each age group, over the 2011-2018 period, obtained from the National Health Laboratory Service (NHLS) [426]. The data are summarized in Table E3.

Table E3: Age distributions of children on ART

	2011	2012	2013	2014	2015	2016	2017	2018*
Numbers								
1-4 years	30883	31069	30466	29437	28477	27726	25601	9860
5-9 years	42083	47493	51456	53925	55208	54206	51874	20416
10-14 years	31118	39300	46971	53864	62026	68310	73219	32397
Proportions								
1-4 years	29.7%	26.4%	23.6%	21.5%	19.5%	18.5%	17.0%	15.7%
5-9 years	40.4%	40.3%	39.9%	39.3%	37.9%	36.1%	34.4%	32.6%
10-14 years	29.9%	33.3%	36.4%	39.3%	42.6%	45.5%	48.6%	51.7%

Data supplied by Lise Jamieson, based on previous analysis [426]. * Data for 2018 are incomplete and the absolute numbers are therefore markedly lower than in previous years.

A number of limitations need to be considered when interpreting these data. The ART totals are estimated based on recorded numbers of viral load tests performed in children. This implies that children who are on ART but who do not get viral load tests done are not counted. There is also the risk of double-counting children who have multiple viral load tests in a given year – although this risk is minimized by using a probabilistic matching algorithm to identify laboratory records that relate to the same patient. In infants, however, this probabilistic matching algorithm is considered to be particularly unreliable, as tests are often entered under the mother's name [426], and consequently infants are excluded from the estimates of numbers of children on ART.

For the purpose of defining the likelihood function, $M(x, t)$ is defined as the model estimate of the fraction of children starting ART in year t who are in age group x (0 for 0-4 years, 1 for 5-9 years and 2 for 10-14 years), and $R(x, t)$ is defined as the corresponding reported fraction from the NHLS data (Table E3). The likelihood function is defined on the assumption that the difference between these proportions is normally distributed with zero mean, i.e.

$$R(x, t) = M(x, t) + \varepsilon(x, t)$$

where $\varepsilon(x, t) \sim N(0, (\sigma_a M(x, t))^2)$. Although it might be considered more correct to define the likelihood based on a multinomial likelihood, the normal approximation gives similar results when sample sizes are large, and the purpose of the $\varepsilon(x, t)$ term is to represent the bias in the data rather than the random multinomial error. As explained previously [6], we set the σ_a parameter to be 0.05. This means that under the assumption of normally-distributed error terms, the 95% confidence intervals around the proportions in Table E3 span the interval from 10% below to 10% above the point estimates.

The definition of the likelihood also takes into account reported numbers of children on ART, as described elsewhere [86]. These data were included in the definition of the likelihood to avoid model fits that yield estimates of numbers of HIV infections less than reported numbers of ART patients.

E5. Likelihood function: routine HIV testing data

Total numbers of HIV antibody tests performed in children are available for only three years (2015-2017). Data were supplied by the South African Department of Health (Tshepo Molapo, personal communication), and therefore relate only to the public sector. Relatively

little antibody testing in children is believed to be conducted in the private sector, as the private sector contributes a small fraction of total tests in adults [124], and most of this is testing in workplace programmes and testing for life insurance purposes (neither of which is relevant to children). We assume that the total number of paediatric antibody tests in each year is 3% greater than the corresponding public sector total, to make allowance for the limited testing that occurs in private medical schemes. (The 3% is based on the estimated fraction of adult HIV antibody tests that are conducted in medical schemes [124].) After adjustment, the estimated total numbers of antibody tests performed in children are 880 100 in 2015-16, 1 143 671 in 2016-17 and 1 162 737 in 2017-18.

We define $T'(t)$ as the model estimate of the total number of antibody tests in children in year t (calculated from the $\tau_s(x, t)$ terms defined in section 5.3.3), and $G'(t)$ as the corresponding empirical estimate (based on the adjusted totals in the previous paragraph). For the purpose of defining the likelihood, the difference between the two, on a log scale, is assumed to be normally distributed with zero mean and standard deviation σ_G. Model estimates of the number of HIV tests in children (as a fraction of the number of tests in adults) are constrained to be reasonably stable in recent years because of the assumption that testing rates in children can be expressed as multiples of the testing rates in adults. The variance of the error term is therefore calculated by assessing the extent to which the fraction of HIV tests in children varies over the 2015-2018 period. The ratio of the number of HIV tests in children to that in adults (considering only the available public-sector statistics) is 0.075 in 2015-16, 0.089 in 2016-17 and 0.095 in 2017-18 (average 0.086, standard deviation 0.010). Given that the model is constrained to produce a fairly stable value of this ratio, we should tolerate a similar 'error' in model estimates, i.e. the coefficient of variation in $G'(t)$ values should be 0.010/0.086 = 0.116 for the purpose of calculating the likelihood. Mathematically,

$$\sigma_G^2 = \mathrm{Var}\left[\log\left(G'(t)\right)\right] \approx \frac{1}{G'(t)^2}\mathrm{Var}\left[G'(t)\right] = 0.116^2 .$$

The model is also calibrated to data on the HIV prevalence in antibody tests conducted in children. The approach adopted to defining the likelihood, accounting for the uncertainty around the test specificity and accounting for the uncertainty regarding the HIV prevalence in children tested in the private sector, is the same as described for adults in Appendix B. (Note however that we do not adjust for test sensitivity, as low rapid sensitivity is strongly associated with a high fraction of HIV infections that are in the acute stage of HIV infection [399], and few children over the age of 18 months would have been recently infected.) Table E4 shows the resulting estimates for children, using the same notation as in Appendix B.

Table E4: Estimated HIV prevalence and associated uncertainty in children receiving antibody testing for HIV

Year	Public sector reported	Public sector adjusted ($Z(t)$)	Private sector proportion ($R(t)$)	Expected total prevalence $E[\rho(t)]$	Variance of total prevalence $Var[\rho(t)]$	Expected unadjusted prevalence $E[\theta(t)]$	Variance of unadjusted prevalence $Var[\theta(t)]$
2015-16	3.76%	3.47%	3.16%	3.42%	$0.03\%^2$	3.71%	$0.20\%^2$
2016-17	2.60%	2.31%	2.91%	2.28%	$0.02\%^2$	2.57%	$0.20\%^2$
2017-18	1.92%	1.62%	3.05%	1.60%	$0.01\%^2$	1.89%	$0.20\%^2$

E6. Likelihood function: Proportion of deaths with an HIV-positive diagnosis

The model is also calibrated to data from the Child Healthcare Problem Identification Programme (Child PIP), a mortality audit system focusing on child deaths in health facilities [427]. In a sample of South African health facilities, data are collected on the circumstances leading to each child death and the causes of death. For each death, information is captured on whether the child was known to be HIV-positive, which is useful in estimating (a) levels of HIV diagnosis in HIV-positive children, and (b) levels of AIDS mortality in HIV-positive children. However, there are number of potential sources of bias that need to be considered. The first is that relatively few health facilities contributed data to Child PIP before 2010, and reporting might therefore not be representative of health services generally. We have therefore only used the data from the 2010-2017 period, as the number of Child PIP deaths peaked in 2010 and thereafter started declining [428], in line with national trends in the under-5 mortality rate [422].

The second source of potential bias is that children who have been diagnosed HIV-positive may be more likely to be taken to health facilities when they fall sick than other children who fall sick, because caregivers are more likely to appreciate the urgency of treatment when the child has been diagnosed positive [429]. Alternatively, HIV testing may be more likely to be conducted in a child who is sick in a health facility than in children who do not get taken to health facilities when they fall sick. Because of this potential source of bias, we allow for uncertainty in the quantity e^{γ}, which we define as the relative rate of death recording in Child PIP facilities, for children who have been diagnosed HIV-positive prior to death, compared to the rate of death recording in Child PIP facilities for children who have not been diagnosed positive. We would expect this ratio to be greater than 1 – or equivalently, we would expect the ratio $e^{-\gamma}$ to be less than 1. We therefore assign a uniform (0, 1) distribution to represent the uncertainty in the ratio $e^{-\gamma}$ (Table S6).

Table E5 summarizes the Child PIP data that we use in calibrating the model. For the purpose of calibration we consider only deaths in the 1-4 year and 5-9 year age groups. In infants a high proportion of deaths occurs in the first months of life, and the fraction of deaths with an HIV-positive diagnosis is therefore sensitive to the exact timing of PCR screening and the delay in test turnaround – which we do not model with a high degree of precision. Many of the deaths in the neonatal period are recorded through the Perinatal Problem Identification Programme (rather than Child PIP), which also contributes to inconsistency between model definitions and the data definitions in the first year of life. Child PIP records relatively few deaths in older children (ages 10 years or older), and we have therefore not used these data in calibration.

Table E5: Child PIP data

Year	Children aged 1-4 years			Children aged 5-9 years		
	Total deaths	HIV-diagnosed	% diagnosed	Total deaths	HIV-diagnosed	% diagnosed
2010	1768	562	31.8%	529	263	49.7%
2011	1496	414	27.7%	413	187	45.3%
2012	1471	337	22.9%	435	168	38.6%
2013	1841	371	20.2%	479	179	37.4%
2014	1947	361	18.5%	482	154	32.0%
2015	1843	337	18.3%	462	157	34.0%
2016	1592	274	17.2%	471	143	30.4%
2017	1372	253	18.4%	414	98	23.7%

For the purpose of calculating the likelihood, we define $R(x, t)$ as the recorded fraction of deaths in health facilities with an HIV-positive diagnosis, for children in age group x in year t (the proportions in Table E5). The model estimate of the fraction of deaths in which HIV is diagnosed prior to death is represented by the symbol $M(x, t)$. The likelihood is then calculated on the assumption that the difference between $R(x, t)$ and $M(x, t)$, on a logit scale, is normally distributed with mean γ, i.e.

$$\ln\left(\frac{R(x,t)}{1-R(x,t)}\right) = \ln\left(\frac{M(x,t)}{1-M(x,t)}\right) + \gamma + \varepsilon_M + \varepsilon_{x,t},$$

where $\varepsilon_M \sim N\left(0, \sigma_M^2\right)$ and $\varepsilon_{x,t} \sim N\left(0, \sigma_{x,t}^2\right)$ represent the 'model error' and 'random error' respectively. The latter represents the binomial error associated with the limited sample size, and the standard deviation is calculated using the normal approximation to the binomial distribution, adjusting to take account of the logit transformation:

$$\sigma_{x,t} = (R(x, t)\,(1 - R(x, t))\,n(x, t))^{-0.5},$$

where $n(x, t)$ is the number of deaths recorded in Child PIP in age group x in year t (Table E5). The 'model error' term takes into account potential mis-specification due to the assumption that the bias (represented by γ) is constant with respect to age and over time. The variance of the model error term is calculated as

$$\hat{\sigma}_M^2 = \frac{1}{16}\sum_{x,t}\left(\ln\left(\frac{R(x,t)}{1-R(x,t)}\right) - \ln\left(\frac{M(x,t)}{1-M(x,t)}\right) - \gamma\right) - \sigma_{x,t}^2.$$

The interpretation of γ as the difference between $R(x, t)$ and $M(x, t)$ on a logit scale, although it may appear to be inconsistent with the previous definition of the ratio $e^{-\gamma}$, is explained elsewhere [6].

E.7 Posterior estimates

Table E6 compares the prior and posterior means for the 17 parameters that are allowed to vary when fitting the model to the paediatric data. In several cases the posterior and prior distributions are similar, although the posterior estimate of the relative rate of fertility in HIV-diagnosed women is closer to 1 than the prior mean, suggesting that HIV diagnosis has relatively little effect on fertility. In addition, the posterior estimate of the annual probability of transmission from breastfeeding mothers (not on ART or in the acute stage of infection) is substantially lower than the prior mean. The posterior mean of the relative rate of testing in early HIV disease, compared to late HIV disease, is 0.034, suggesting a very low rate of antibody HIV testing in children who have been free of HIV-related symptoms. The model calibration to the paediatric HIV data sources is presented elsewhere [6].

Table E6: Comparison of prior and posterior distributions

Parameter	Prior distribution (mean, 95% CI)	Posterior distribution (mean, 95% CI)
RR of fertility in HIV-diagnosed untreated women	0.700 (0.396-0.927)	0.939 (0.908-0.960)
Probability of MTCT from chronically- infected mothers, per year of mixed feeding	0.140 (0.095-0.192)	0.109 (0.101-0.118)
Probability of MTCT from acutely- infected mothers, per month of mixed feeding	0.160 (0.106-0.223)	0.144 (0.124-0.157)
Ratio of postnatal transmission risk per month of EBF to postnatal transmission risk per month of mixed feeding	0.500 (0.213-0.787)	0.611 (0.543-0.678)
Children infected at/before birth		
Annual rate of progression to late disease in older children	0.400 (0.229-0.619)	0.405 (0.338-0.549)
Excess annual rate of progression to late disease in neonates	2.00 (1.14-3.09)	2.50 (2.22-2.77)
Excess progression reduction factor, per year of age	0.400 (0.214-0.602)	0.421 (0.366-0.468)
RR of progression to late disease if infected postnatally	0.350 (0.096-0.666)	0.206 (0.156-0.245)
Children in late disease, untreated		
Annual rate of AIDS mortality in older children	0.120 (0.069-0.186)	0.118 (0.107-0.130)
Excess annual rate of AIDS mortality in neonates	3.50 (2.27-5.00)	3.60 (3.28-3.99)
Excess AIDS mortality reduction factor, per year of age	0.200 (0.047-0.468)	0.310 (0.280-0.352)
HIV testing rates		
RR of testing in virgin adolescents: 2005	0.200 (0.047-0.428)	0.098 (0.059-0.131)
RR of testing in virgin adolescents: 2010	0.200 (0.047-0.428)	0.125 (0.110-0.142)
RR of testing in early disease (relative to late disease)	0.500 (0.025-0.975)	0.034 (0.022-0.046)
Effect of ART on mortality		
RR of mortality in ‘stable’ ART phase compared to untreated children with late disease	0.100 (0.025-0.217)	0.049 (0.033-0.064)
Reduction in mortality (on log scale) per unit increase in rate of ART initiation (in late disease) over last 3 years	7.50 (2.29-15.76)	3.85 (3.14-4.73)
RR of recording of deaths in undiagnosed and HIV-negative children, relative to HIV-diagnosed	0.500 (0.025-0.975)	0.280 (0.233-0.331)

ART = antiretroviral treatment; EBF = exclusive breastfeeding; MTCT = mother-to-child transmission. RR = relative rate.

Appendix F: Estimating rates of viral suppression

In the previous version of Thembisa, a cubic function was used to represent the change over time in the rate of viral suppression in patients starting ART at CD4 counts of less than 200 cells/μl [2]. Viral suppression data from both the TIER database and the National Health Laboratory Service (NHLS) database were used in fitting this cubic function. The model also allowed for potential bias in the published data due to missing viral loads in a substantial fraction of patients, with the extent of the bias being assumed to depend on viral load testing coverage (the fraction of ART patients who have viral load testing data). There are a number of limitations associated with this approach. Firstly, it relies on an implicit assumption that the bias due to missing viral load data depends only on the viral load testing coverage and that the relationship between the true viral suppression and the testing coverage is the same across all provinces. In reality, factors other than testing coverage can influence the extent of the bias, and the relationship between testing coverage and the true rate of viral suppression is therefore not necessarily the same in all provinces. Secondly, the method relies on the assumption that testing coverage can be estimated accurately for both the NHLS and TIER datasets. In reality, testing coverage is imperfectly measured in the NHLS dataset because there is potential double-counting of viral load tests conducted in the same individual, and because the denominator is calculated from DHIS estimates of total numbers of patients on ART (i.e. the numerator and denominator are not obtained from the same source).

In the new version of Thembisa, we adopt a more formal Bayesian approach to estimating the true rate of viral suppression. For each province, we assume that the odds of viral suppression is proportional to the odds of viral suppression estimated using data from South African cohorts participating in the International epidemiology Databases to Evaluate AIDS (IeDEA) collaboration. This constant of proportionality (or odds ratio) is assumed to differ by province. In addition, we assume for each province that there is an odds ratio relating the odds of viral suppression in patients with unrecorded viral loads to that in patients with recorded viral loads. We follow a two-step Bayesian updating approach in estimating these two parameters. In the first step (described in section F.1), only the TIER data are used in the model fitting. In the second step (described in section F.3), the posterior estimate of the IeDEA bias from the first step becomes the prior distribution for the IeDEA bias in the second step, and the Thembisa model is fitted to other HIV data sources. Note that in the second step it is not necessary to include the uncertainty regarding the bias due to missing data, because this bias affects only the interpretation of the TIER data, and the TIER data on viral suppression are not included in the second step.

F.1 Viral suppression in adults

A model of the following form is fitted to IeDEA-SA data on viral suppression in adults:

$$\text{logit}(I_{t,s}) = C + \beta_t + \gamma_s, \qquad \text{(F1)}$$

where $I_{t,s}$ represents the proportion of patients who are virally suppressed in year t, in patients who started ART in baseline CD4 count category s. Full details of the IeDEA-SA dataset and the procedures followed in defining viral suppression are provided elsewhere [189]. The results of the model are summarized in Table F1. Consistent with the previous Thembisa

estimates [2], the results suggest a substantial decline in rates of viral suppression after 2009, followed by a gradual increase in viral suppression after 2013. The results also suggest substantially higher rates of viral suppression in patients who start ART at higher CD4 counts, consistent with previous studies [430-433].

Table F1: Predictors of viral suppression in IeDEA-SA cohorts

Factor	Symbol	Odds ratio (95% CI)
Constant	exp(C)	7.24 (6.22-8.42)
Effect of calendar year (ref. 2005)		
2006	$\exp(\beta_{2006})$	1.05 (0.89-1.24)
2007	$\exp(\beta_{2007})$	0.95 (0.81-1.11)
2008	$\exp(\beta_{2008})$	1.04 (0.88-1.22)
2009	$\exp(\beta_{2009})$	0.91 (0.77-1.07)
2010	$\exp(\beta_{2010})$	0.71 (0.60-0.83)
2011	$\exp(\beta_{2011})$	0.47 (0.40-0.54)
2012	$\exp(\beta_{2012})$	0.53 (0.45-0.62)
2013	$\exp(\beta_{2013})$	0.52 (0.45-0.61)
2014	$\exp(\beta_{2014})$	0.63 (0.54-0.73)
2015	$\exp(\beta_{2015})$	0.82 (0.70-0.96)
2016	$\exp(\beta_{2016})$	0.63 (0.54-0.74)
2017	$\exp(\beta_{2017})$	0.91 (0.78-1.06)
2018	$\exp(\beta_{2018})$	0.93 (0.79-1.09)
Baseline CD4 category (ref. <200 cells/μl)		
200-349 cells/ μl	$\exp(\gamma_{200})$	1.49 (1.44-1.54)
350-499 cells/ μl	$\exp(\gamma_{350})$	1.73 (1.63-1.83)
≥500 cells/ μl	$\exp(\gamma_{500})$	1.92 (1.81-2.05)

We define $V_{t,s}(p)$ as the estimate of the true rate of viral suppression in patients on ART in year t, who started ART in CD4 category s and who currently live in province p. This is calculated as

$$\text{logit}(V_{t,s}(p)) = \text{logit}(I_{t,s}) + \lambda_p, \qquad \text{(F2)}$$

where $\exp(\lambda_p)$ is the odds ratio relating the odds of viral suppression in province p to that in the IeDEA-SA cohorts. The logit transformation is applied to avoid situations in which the regression model predicts a rate of viral suppression <0% or >100%.

The λ_p terms are unknown, and we therefore specify a prior distribution to represent the uncertainty around these parameters. In the analysis of IeDEA-SA adult viral load data, it was noted that the average rate of viral suppression (at a threshold of <400 RNA copies/ml) was 85.7%, and that the rates in each year were consistently between 1% and 5% higher than those reported by the Department of Health at a national level (using the same threshold) [159]. This suggests a prior mean for λ_p of around -0.23 (logit(0.827) – logit(0.857), where 0.827 = 0.857 – 0.03 and 0.03 is the midpoint of the 0.01-0.05 range). We have set the standard deviation of the prior distribution to 0.27, based on fitting the same regression model as shown in equation (F1) to IeDEA-SA data, but allowing for additional terms to represent differences across ART cohorts. The standard deviation of these cohort-specific terms determines the prior standard deviation of 0.27. The prior distribution is thus a normal distribution with a mean of -0.23 and a standard deviation of 0.27.

For the purpose of calibrating the model estimates ($V_{t,s}(p)$) to routinely reported rates of viral suppression from TIER, it is also necessary to allow for uncertainty regarding the bias due to missing viral load data. We define θ_p as the ratio of the odds of viral suppression in patients with missing viral load measurements to that in patients with recorded viral loads, in province p. We consider the following estimates in defining plausible ranges for θ_p:

- In the analysis of IeDEA-SA data, missing viral load measurements were imputed for each individual, based on their clinical and demographic characteristics, their outcomes (death or loss to follow-up) and viral load measurements taken at other visits (Tameryn Pillay, manuscript under review). On average, the imputed viral loads were lower than the recorded viral loads, although the difference was small (aOR 0.96, 95% CI: 0.94-0.98).
- Euvrard *et al* [434] compared viral suppression in South African patients whose viral loads were not recorded with those in patients whose viral loads were recorded and found the latter to be significantly higher (aOR 1.43, 95% CI: 1.30-1.59). This might be considered an upper bound on the 'true' value of the θ_p parameter, since 16% of patients did not have any viral load done (regardless of whether it was recorded), and these patients who did not have their viral loads done were probably patients who were not attending clinics as frequently and therefore poorer treatment adherers with higher viral loads.
- Another study in Zambia found that rates of viral suppression in patients who had 'unofficially transferred' to another ART service were substantially lower than those who had remained in care at their original ART facility (50.2% versus 81.9% respectively, OR 0.22) [435]. Such unofficially transferred patients are to some extent representative of patients with missing viral loads, since South African reporting systems exclude individuals who have transferred in from other health services when reporting on viral suppression [434], and because unofficial transfers are often preceded by poor clinic attendance and negative patient experiences [436] (i.e. circumstances that might make a patient likely to miss their viral load test). Nevertheless, the OR of 0.22 is probably a lower bound on the 'true' value of the θ_p parameter, since 'unofficial transfer' was a self-reported outcome, and it is likely that some of the individuals who claimed to still be on ART were no longer receiving ART (our model definition excludes individuals who have stopped ART completely).
- Other studies have documented extremely high odds ratios when comparing untested to tested patients [437], but in the context of low coverage of routine viral load testing, and therefore not applicable to South Africa, where routine viral load testing has been policy since the start of the public sector ART programme.

Based on these observations, we assign a gamma prior to represent the uncertainty around the θ_p parameter, with mean 0.96 and standard deviation of 0.25. This distribution has 2.5 and 97.5 percentiles of 0.53 and 1.51 respectively, i.e. roughly consistent with the range of empirical estimates that we have noted.

Suppose the $R_t(p)$ is the reported rate of viral suppression in province p at time t, and that the corresponding testing coverage on which this is based is $\delta_t(p)$. For the purpose of defining a likelihood function, we define

$$\text{logit}\left(\delta_t(\text{p})R_t(\text{p}) + \frac{1-\delta_t(\text{p})}{1+\frac{1\text{-}R_t(\text{p})}{R_t(\text{p})\theta_{\text{p}}}}\right) = \text{logit}\left(\sum_{s=1}^{4}\pi_{t,s}(p)V_{t,s}(p)\right) + \varepsilon_t(p), \qquad \text{(F3)}$$

where $\pi_{t,s}(p)$ is the model estimates of the fraction of ART patients who started ART in CD4 category s, and $\varepsilon_t(p) \sim \text{N}(0, \sigma^2)$. Our assumption is thus that if the model provides a good fit to the data, the difference on the logit scale between the reported rates of viral suppression (after adjustment for missing viral load data) and the modelled rates of viral suppression (after adjustment for differences in viral suppression across baseline CD4 categories) should be close to zero on average, with relatively small variance (i.e. low σ^2). The $\pi_{t,s}(p)$ terms are approximated from the previous version of Thembisa (version 4.2).

The variance term is approximated using the maximum likelihood formula:

$$\sigma^2 = \frac{1}{n_p}\sum_{\text{t}\in T_{\text{p}}}\left[\text{logit}\left(\delta_t(\text{p})R_t(\text{p}) + \frac{1-\delta_t(\text{p})}{1+\frac{1\text{-}R_t(\text{p})}{R_t(\text{p})\theta_{\text{p}}}}\right) - \text{logit}\left(\sum_{s=1}^{4}\pi_{t,s}(p)V_{t,s}(p)\right)\right]^2,$$

where T_p is the set of time points for which we have viral suppression estimates and n_p is the number of $R_t(p)$ data points for province p.

The model is fitted to several datasets for each of the nine provinces (and for the country as a whole). The following sources are combined in creating the dataset for each province:

- Province-specific viral load data from the TIER database for 2013-14, considering patients who had been on ART for 6 months and 48 months, i.e. a total of 18 data points (information on viral suppression was not available for other years or for other ART durations) [159].
- National viral load data from the TIER database for patients who had been on ART for 6 months (for each year from 2005-2014) and for 48 months (for each year from 2009-2014), i.e. a total of 16 data points [159].
- National viral load data from the TIER database for patients on ART in 2013, reported at 12-month intervals from 12 to 108 months after ART initiation, i.e. a total of 9 data points [170].
- National and provincial viral load data from the TIER database, for patients who had been on ART for 6 months, for each of the four quarters in the 2017-18 and 2018-19 fiscal years, i.e. a total of 80 data points (Thapelo Seatlhodi, personal communication).

Other viral load data are available, but this analysis is limited to those data sources for which there was information on both the proportion of patients with viral load measurements and the proportion of those measurements that were suppressed. Suppression was defined in all cases as a viral load of <400 RNA copies/ml.

Having specified the prior distributions and the likelihood function, the final step in the Bayesian analysis is the simulation of the posterior distribution. We follow a Sampling

Importance Resampling approach to approximate the posterior [438]. Since there are only two parameters being estimated in the analysis (λ_p and θ_p), it is sufficient to use a small sample size (1000) in both the sampling and resampling steps.

Table F2 summarizes the results of the Bayesian analysis for each province (and for the country as a whole). The standard deviation of the model errors is included as a measure of 'goodness of fit'. In most provinces, the standard deviation is around 0.2 or lower, but a notable exception is Gauteng, where there has been a steep increase in viral suppression over the last two years (steeper than suggested by the IeDEA data), leading to a poor model fit. In the Free State, KwaZulu-Natal and Western Cape, levels of viral suppression appear to be slightly higher than the rates estimated from the IeDEA data (i.e. positive posterior estimates of λ_p), but in all other provinces viral suppression appears to be lower than that in the IeDEA-SA cohorts. Finally, the posterior estimates of the θ_p parameter are in most cases not very different from the prior mean (0.96); the main exception is Mpumalanga, where the posterior mean (1.30) is substantially higher than the prior mean.

Table F2: Posterior estimates of viral suppression parameters

	Sample Size (n_p)	Standard deviation of model errors (σ)	Difference in viral suppression relative to IeDEA (λ_p)	OR for viral suppression if VL not recorded (θ_p)
EC	10	0.097	-0.43 (-0.64 to -0.29)	1.03 (0.61-1.45)
FS	10	0.170	0.22 (-0.03 to 0.46)	1.06 (0.60-2.03)
GT	10	0.406	-0.45 (-0.70 to -0.18)	1.14 (0.62-1.73)
KZ	10	0.137	0.23 (0.02 to 0.42)	0.88 (0.55-1.30)
LM	10	0.224	-0.59 (-0.79 to -0.39)	1.02 (0.62-1.52)
MP	10	0.176	-0.15 (-0.34 to 0.01)	1.30 (0.77-2.20)
NC	10	0.260	-0.28 (-0.58 to -0.05)	1.09 (0.67-1.70)
NW	10	0.225	-0.18 (-0.39 to 0.03)	0.96 (0.60-1.42)
WC	10	0.134	0.15 (-0.06 to 0.36)	0.81 (0.51-1.30)
SA	33	0.249	-0.18 (-0.40 to 0.03)	0.94 (0.61-1.33)

Figure F1 shows an example of the model calibration to the data at a national level. The data shown in Figure F1 are unadjusted since the posterior mean of θ_p is close to 1 at a national level. The model results are calculated for the posterior mean (λ_p = -0.18) and the 2.5 and 97.5 percentiles (λ_p = -0.40 and 0.03 respectively, as represented by the dashed lines). The model results generally appear consistent with the data, with one notable outlier in 2013 (which related to patients who had been on ART for 9 years, i.e. a relatively small cohort since the public-sector ART programme only started in 2004). The model is also validated by two data points from 2016 (not included in the model calibration because there was no associated information on the testing coverage). Both the model and the data suggest a dip in rates of viral suppression around 2010-2011, followed by a gradual increase thereafter.

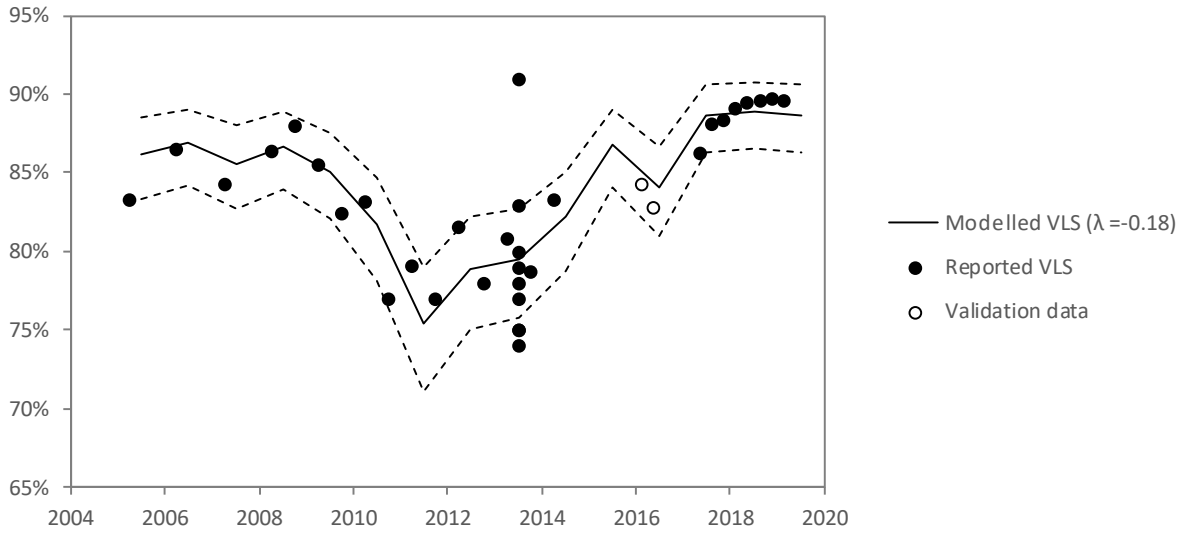


Figure F1: Viral suppression trends (% of ART patients with VL <400 RNA copies/ml)
In 2013 there are several data points, as rates of viral suppression were reported at a number of different ART durations; in most other periods the data represented relate to viral load testing around 6 months after ART initiation. Dashed lines represent 95% confidence intervals calculated using the 2.5 and 97.5 percentiles of λ_p. Model results are calculated using the regression coefficients in Table F1, adjusted by λ_p (equation F2) and weighting by the Thembisa estimates of the fraction of ART patients starting ART in each CD4 category in the corresponding year.

F.2 Viral suppression in children

For children, we lack reliable nationally representative data on rates of viral suppression. The District Health Information System (DHIS), which summarizes the data from TIER, reports overall rates of viral suppression, but not disaggregated by age group. Our approach is therefore to use IeDEA paediatric ART data to estimate time trends in viral suppression [189], and then to adjust these rates using the same adjustment factors that we use in adults (i.e. using the same formula as in equation F2). This means that we are not attempting to 'fit' routine viral suppression data, in the way we do for adults. Table F3 summarizes the IeDEA-SA estimates of viral suppression in children, and shows the implied rates of viral suppression for different values of λ_p. The range of model estimates differ from other published sources. For example, Joseph Davey *et al* [377] found that among children receiving ART in 5 districts in 2016, the most recent viral load was suppressed in 68.2% of children. On the other hand, Lilian *et al* [439] found rates of viral suppression of 48-52% among children receiving ART in a rural district in Limpopo. However, neither of these data sources is nationally (or even provincially) representative, and there is thus substantial uncertainty regarding true rates of viral suppression in children at a national level. NHLS data have suggested rates of viral suppression in children of 64-67% over the 2015-18 period (Kimberley Perez, personal communication), but these are based on a threshold of <1000 RNA copies/ml (not 400) and may be biased by double-counting of tests conducted in the same patient.

Table F3: Viral suppression in children (<400 RNA copies/ml)

	IeDEA-SA	Model estimate		
Year	estimate	λ_p = -0.18	λ_p = -0.40	λ_p = 0.03
2005	73.7%	70.1%	65.3%	74.3%
2006	78.2%	75.0%	70.6%	78.7%
2007	78.0%	74.8%	70.4%	78.5%
2008	82.0%	79.1%	75.3%	82.4%
2009	79.5%	76.4%	72.2%	80.0%
2010	73.5%	69.9%	65.0%	74.1%
2011	67.9%	63.9%	58.7%	68.6%
2012	68.0%	64.0%	58.8%	68.7%
2013	64.3%	60.0%	54.7%	65.0%
2014	70.4%	66.5%	61.4%	71.0%
2015	67.7%	63.7%	58.5%	68.4%
2016	66.2%	62.0%	56.7%	66.8%
2017	71.9%	68.1%	63.2%	72.5%
2018	70.6%	66.7%	61.7%	71.2%

F.3 Updating the assumptions about viral suppression in Thembisa

In the calibration of the Thembisa model to provincial/national HIV prevalence and mortality data, we allow for uncertainty regarding the λ_p parameter, as the extent of viral suppression influences both the trends in HIV incidence and the trends in HIV-related mortality. The posterior distributions in Table F2 become the prior distributions in this calibration process, i.e. using a Bayesian updating process to estimate λ_p. (In the first step, the estimates of λ_p are determined by province-specific or national viral suppression data, and in the updating step the λ_p parameters are determined by the province-specific or national HIV prevalence and mortality trends.)

We define $V'_{t,s}(p)$ as the Thembisa estimate of the true rate of viral suppression in patients on ART in year t, who started ART in CD4 category s and who currently live in province p. This is calculated as

$$\text{logit}(V'_{t,s}(p)) = \text{logit}(0.5 \times [I_{t,s} + I_{t+1,s}]) + \lambda_p,$$

for $t \geq 2005$ and $t \leq 2018$. This is similar to the equation in (F2), but we average across calendar years t and $t + 1$ for the purpose of calculating the Thembisa estimates because the projection years in Thembisa run from mid-year to mid-year. In the period from 2020 onward, rates of viral suppression are assumed to increase as a result of the introduction of dolutegravir to replace efavirenz and nevirapine in first-line ART regimens. A recent network meta-analysis estimates that patients receiving dolutegravir are significantly more likely to achieve viral suppression than patients receiving efavirenz (OR 1.87, 95% CI: 1.34-2.64) [190]. We have used this odds ratio to determine the rates of viral suppression in 2020 and subsequent periods, since efavirenz has been the main first-line antiretroviral drug in South Africa up to 2019. In the 2019 projection year (which runs from mid-2019 to mid-2020) we assume the rate of viral suppression to be mid-way between that in the 2018 and 2020 projection years, since this is a period of transition from efavirenz to dolutegravir. No dolutegravir adjustments are made in the case of children, as this drug is only recommended for individuals over 20 kg.

Appendix G: Modelling treatment interruptions

This appendix describes the approach used to estimate the frequency of treatment interruption and treatment resumption after an interruption. Figure G1 presents an overview of the theoretical model that we apply to South African data sources. Patients are assumed to disengage from care at a constant rate λ, but only a proportion θ of these disengagements are assumed to be true ART interruptions. The reasons for other disengagement will vary from study to study, depending on the methods used to classify patients 'lost to follow-up' (LTFU), but will most commonly include patients who have transferred to other ART services (so-called 'silent transfers') and patients who have died without their death being recorded by the clinic at which they were receiving ART. The patients who interrupt ART are assumed to resume ART at a constant rate ρ, with a proportion ϕ_1 of these patients resuming ART at a different clinic within the same province, a proportion ϕ_2 resuming ART in a different province, and the remainder resuming ART at the same clinic at which they originally received ART. For the sake of simplicity, we do not consider mortality while interrupting ART, as studies suggest that after excluding the deaths that occur soon after LTFU (which in most cases represent failure to record mortality rather than mortality after a treatment interruption) this is a relatively infrequent occurrence [440, 441].

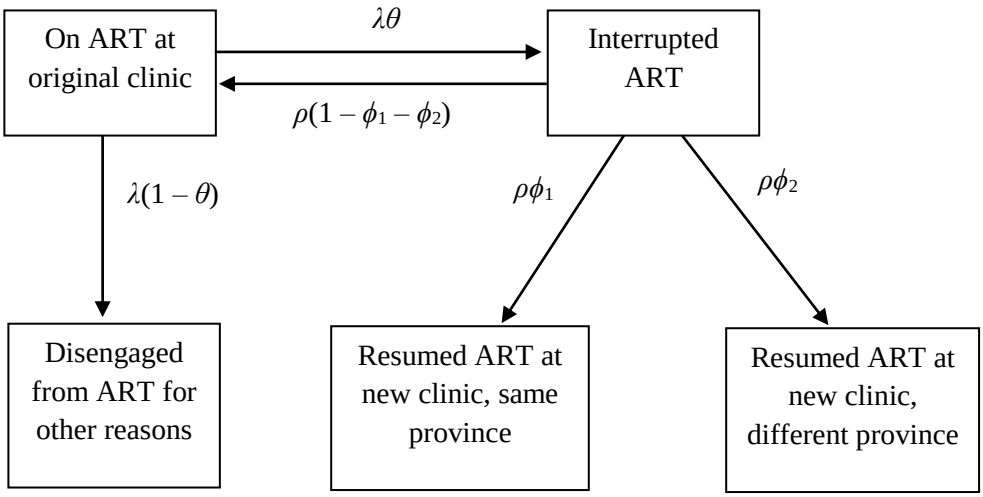


Figure G1: Model of ART interruption and return to care

We attempt to estimate the parameters of this model using data from different studies. Three South African studies were identified that estimated either rates of treatment interruption or rates of treatment resumption [441-443]. Each study follows a different design, and no single study estimates all of these parameters, but by drawing on the estimates from different studies it is possible to determine plausible ranges for the parameters. Of primary interest are the parameters λ, θ and ρ. The sections that follow describe each of the studies and the process followed in estimating the model parameters. The findings of the different studies are then synthesized in the final section.

G.1 Patients on ART in Khayelitsha (Western Cape)

Kaplan *et al* [441] conducted the largest and most recent of the three studies. Patients were included in the study population if they were receiving ART in Khayelitsha clinics between 2013 and 2014. Patients who were lost to follow-up (more than 6 months without a visit and no documented transfer or death) were tracked to determine if they had received ART elsewhere in the province (by linking their patient ID to the provincial patient record system) or had died (by linking their civil ID to the vital registration system). Disengagement from care was defined as a gap of more than six months in receiving ART, without the patient receiving ART or dying within the 6-month window. The definition of disengagement therefore excludes silent transfers (where the patient received ART at another service within 6 months of dropping out of ART care) and unrecorded deaths. The resulting estimates of the treatment interruption rate, which varied between 0.10 and 0.12 per annum, are thus directly comparable with the modelled quantity $\lambda\theta$. This rate of treatment disengagement was found to be roughly constant with respect to the time since ART engagement, after controlling for baseline covariates. This rate of treatment disengagement is probably an under-estimate, since individuals who disengaged from ART but returned to the same clinic at which they were previously treated are not included in the definition of treatment interrupters.

The study found that cumulative rates of return to care after a treatment interruption increased from 27% 1 year after the interruption to 50% by 2.5 years after the interruption (Table G1). However, it is important to note that this represents only patients who returned to other ART facilities within the Western Cape, and (as noted in the previous paragraph) individuals who returned to care at the original ART clinic were not included in the definition of treatment interrupters. It is also important to note that because of the 6-month window used to define LTFU, the quoted durations of follow-up after interruption are 6 months longer than the period in which individuals could actually return to care (if they had returned to care within 6 months then they would not have met the study definition of an interrupter). The model quantity that should be compared to the 27% rate of return after 1 year is thus

$$\frac{\phi_1}{\phi_1 + \phi_2} \exp\left(-\rho / 2\right),$$

and similar equations are defined at the other durations of follow-up shown in Table G1. We adopt a Bayesian approach in estimating these quantities. A vague prior (uniform on the interval [0, 1]) is assigned to represent the uncertainty in the quantity $\exp(-\rho)$. The quantity $\phi_1/(\phi_1 + \phi_2)$ can be estimated approximately from a similar study conducted in Gauteng [443], which found that out of 103 patients who resumed ART at a different ART clinic after interrupting treatment, 69% did so in the same province and the remaining 31% started ART in a different province. Although the rates of clinic movement might be different in Gauteng and the Western Cape, both are predominantly urban provinces, and some similarity would therefore be expected. We therefore assign a beta prior to represent the uncertainty in the quantity $\phi_1/(\phi_1 + \phi_2)$, with mean 0.69 and standard deviation 0.046 (consistent with the sample size of 103).

A likelihood function is defined with reference to the data in Table G1, assuming that the data are drawn from a normal distribution with the mean given by the model and a standard deviation calculated assuming a coefficient of variation of 0.1 around the data (standard

deviations were not reported in the Kaplan study and the coefficient of variation of 0.1 is therefore taken from another study conducted in the Western Cape, the Kranzer study described below [442]). The posterior estimates were generated by drawing a sample of 10 000 parameter combinations from the prior distributions, calculating the likelihood values for each parameter combination, and then drawing a second sample of 1000 parameter combinations from the original set of parameter combinations using the likelihood values as weights (i.e. a sampling importance resampling procedure [438]). Table G1 shows that the resulting model estimates of the fraction returning to care are roughly consistent with the data.

Table G1: Proportions of ART interrupters returning to care at different clinics in the Western Cape

Months after interruption	Observed % returned to care	Standard deviation (assumed)	Posterior model mean (95% CI)
12	27%	2.7%	23.7% (20.5-27.7%)
18	38%	3.8%	38.7% (34.8-43.0%)
24	45%	4.5%	48.3% (44.4-52.4%)
30	50%	5.0%	54.3% (49.4-58.9%)

The posterior estimate of the quantity $\phi_1/(\phi_1 + \phi_2)$ is 0.66 (95% CI: 0.57-0.75), roughly consistent with the prior mean. The posterior estimate of the ρ parameter is 0.92 (95% CI: 0.67-1.32).

As noted previously, the $\lambda\theta$ estimates of 0.10-0.12 are likely to be under-estimates because interrupters who return to the same clinic are not included in the definition of interruption. However, they are also likely to be under-estimates of the true rates because interrupters who resume ART at a different clinic within 6 months of an interruption are also excluded from the definition. Having estimated $\rho = 0.92$, it is possible to estimate the fraction of interrupters who do not resume ART within 6 months, which is $\exp(-\rho/2) = 0.63$, and thus the reported rates of interruption should be inflated by factors of 1/0.63 in order to obtain corrected estimates (0.16-0.19).

G.2 Patients on ART in Masiphumelele (Western Cape)

Kranzer *et al* [442] measured rates of ART interruption and return to care in the Masiphumelele community. This study differed from the previous study in a number of important respects. Treatment interruptions were defined as periods of more than 30 days without a pharmacy refill visit, but individuals who returned to the original clinic were not excluded from the definition of treatment interrupters. Using this definition yielded an incidence of treatment interruption of 0.128 per annum (95% CI: 0.114-0.144).

The study was limited in its ability to isolate 'true' interrupters; patient records were not linked to other health facilities and there was no linkage to the vital registration system to determine whether the patient had died. However, three attempts were made to contact patients who were late in collecting their pharmacy refills, and in addition, home visits were made in the case of half of the LTFU cases to determine whether the patient had died or left the area (such patients were censored at the date of death or migration in the calculation of rates of ART resumption). We define θ to be the fraction of LTFU patients who remained in the same community (Masiphumelele) after stopping treatment and who were not LTFU due

to death (i.e. θ is the fraction of interrupters who could reasonably resume ART in Masiphumelele). Although the θ parameter is unknown, a recent systematic review of studies that had traced patients LTFU after starting ART in sub-Saharan Africa estimated the proportion of individuals not on ART (as distinct from dead or transferred to other ART services) to be 29% (estimates varied between 4% and 82% across studies) [444]. The 29% is probably an under-estimate of the θ parameter, considering that (a) some attempts were made by Kranzer *et al* to trace patients who were LTFU and thus exclude unrecorded deaths and transfers/migrations, and (b) the tracing studies included in the systematic review would mostly not have included patients who interrupted ART but returned to their original clinic in their definition of LTFU. We therefore assign a beta prior to represent the uncertainty around the θ parameter, with mean 0.55 and standard deviation 0.15. The 2.5 and 97.5 percentiles of this distribution are 0.25 and 0.83 respectively. The mean and standard deviation have thus been chosen such that the lower limit is close to the 29% average (which we consider a likely lower bound) and the upper bound is close to the maximum of 82% estimated in the studies included in the systematic review.

Our model estimates that of individuals who interrupted ART, the proportion who resumed ART at their original clinic within t years is $\theta \exp(-\rho(t - 1/12))$, where ρ is defined as the rate of ART resumption within the Masiphumelele community, among the individuals who were still alive and untreated in the Masiphumelele community. The 1/12 adjustment is included to represent the 1-month window used to define interruption. The parameters θ and ρ can be estimated roughly by fitting the model to the observed cumulative proportions of patients who resumed ART (Table G2). As in the previous section, we assign a vague prior to the $\exp(-\rho)$ quantity. We define a likelihood function based on the assumption that the data are drawn from a normal distribution with the means given by the model and the standard deviations calculated from the published confidence intervals around the data. The posterior distribution is simulated in the same way as in the previous section.

Table G2: Proportions of ART interrupters returning to care in Masiphumelele

Months after interruption	Observed % returned to care (95% CI)	Standard deviation (assumed)	Posterior model mean (95% CI)
12	26.7% (21.7-32.7%)	2.8%	25.8% (21.0-31.5%)
24	37.1% (31.1-43.9%)	3.3%	37.5% (33.4-41.4%)
36	42.1% (35.2-49.7%)	3.7%	42.5% (36.6-48.5%)

The resulting posterior estimates of the θ and ρ parameters are 0.47 (95% CI: 0.37-0.65) and 0.94 (95% CI: 0.46-1.65) respectively. The estimate of the rate of ART resumption is thus similar to that estimated in the previous section (0.92). However, the estimate of the quantity $\lambda\theta$ is 0.06 (0.128×0.47), which is substantially lower than the estimates of 0.16-0.19 obtained using the Kaplan study data. This is probably because the patients included in the Kranzer study were followed much more intensively (as noted before, three attempts were made to contact patients if they were late in returning for pharmacy refills), and a subset of the patients were enrolled in a clinical trial. In addition, the estimate of 0.06 per annum excludes individuals who moved out of the study area (who may have been more likely to interrupt ART).

G.3 Pregnant women in Gauteng

Clouse *et al* [443] attempted to trace women who were classified lost to follow-up after initiating ART during pregnancy. Patient records were linked to the National Health Laboratory Service (NHLS) database to identify instances where women had initiated ART at other facilities (outside of the facility at which ART had been initiated during pregnancy), but the definition of treatment resumption excluded women who returned to the same clinic at which they originally resumed ART. Of 274 women who were LTFU, 103 (37.6%) were found to have received ART at other sites. However, it is not clear how many of the women died (since there was no linkage of civil identifiers to the vital registration system), which is a concern, as some of the LTFU might actually have been due to mortality that was not captured by the original clinic at which women started ART. In addition, it was noted that 30% of the women were not from South Africa, and it is possible that some of these women returned to their country of origin soon after giving birth (which means that even if they resumed ART, their records would not be picked up by the NHLS).

We therefore specify parameter θ_1 to represent the fraction of women who were LTFU as a result of death, and parameter θ_2 to represent the fraction of women who left the country after giving birth. We assign a uniform prior distribution to represent the uncertainty regarding the latter, assigning equal probabilities to all values on the range [0, 0.3], the upper bound being the fraction that would be expected if all immigrants returned to their country of origin. We assign a beta prior distribution to represent the uncertainty around the parameter θ_1, with mean 0.3 and standard deviation 0.10. This prior distribution is based on a systematic review of tracing studies [445], which found that the fraction of LTFU patients who had died varied between 12% and 87% in African studies, and varied inversely in relation to the cumulative LTFU rate, being around 30% at a cumulative LTFU rate of 38% (the LTFU rate in the Clouse study). The 30% is also similar to the results of a more recent systematic review, which found that in patients who were LTFU after 2007, the fraction who were found to have died after tracing was 30% (95% CI: 21.2-38.9%) [444].

Unlike in the two previous studies, the cumulative proportion of women who had returned to care was not specified at fixed durations after the interruption occurred – the 37.6% referred to previously therefore represents the average across different follow-up durations in the study. Women started ART between the start of January 2012 and the end of July 2013, and LTFU was reported as occurring a median of 3 months after ART initiation (equivalent to a mean of 0.36 of a year). If we assume for simplicity that all women who were LTFU were LTFU after exactly 0.36 of a year, and that the distribution of ART initiation dates was uniform over the interval [2012, 2013.58], then the distribution of LTFU times must have been uniform over the interval [2012.36, 2013.92]. In September of 2015, the NHLS database was assessed for evidence of a return to care in South Africa. At this time, we would expect the cumulative fraction returned to care to be

$$\frac{(1-\theta_1)(1-\theta_2)}{2013.92-2012.36}\int_{2012.36}^{2013.94}(1-\exp(-\rho(2015.75-t-\delta)))dt,$$

where ρ is the rate of ART resumption after an interruption, and δ is the LTFU window (in this study, women were only considered LTFU if they had no clinic visit for 3 or more months, i.e. $\delta = 0.25$). The above expression simplifies to

$$(1-\theta_1)(1-\theta_2)\left(1-\frac{1}{\rho\times 1.58}\left(\exp(-1.56\rho)-\exp(-3.14\rho)\right)\right).$$

The likelihood is calculated by comparing this modelled proportion to the observed proportion (0.376), assuming that the observed proportion is normally distributed with a mean given by the above expression, and a standard deviation of 0.029 (the standard deviation is calculated from the data). As in the previous sections, we assign a vague prior to the exp(-ρ) quantity. The posterior distribution is simulated by drawing a sample of 10 000 combinations of the θ_1, θ_2 and exp(-ρ) quantities, calculating the likelihood for each, and then drawing a second sample of 1000 parameter combinations from the first sample, using the likelihood values as weights.

The resulting posterior estimates have wide confidence intervals. The estimated annual rate of ART resumption (ρ) is 0.70 (95% CI: 0.28-2.60), while the estimated proportions leaving South Africa and dying (θ_1 and θ_2) are 17% (95% CI: 1-29%) and 34% (95% CI: 15-54%) respectively.

Having obtained estimates of θ_1 and θ_2, it is possible to approximate the 'true' rate of ART interruption, $\lambda\theta^*$, where λ is the LTFU rate and $\theta^* = (1 - \theta_1)(1 - \theta_2)$. From the above information, we would expect the cumulative fraction LTFU at the end of August 2015 to be

$$\frac{1}{2013.58-2012}\int_{2012}^{2013.58}\left(1-\exp(-\lambda(2015.67-t-\delta))\right)dt$$

Setting this expression to 38.1% (the observed cumulative fraction LTFU) yields an estimate of $\lambda = 0.18$, and the corresponding estimate of $\lambda\theta^*$ is 0.10. This is likely to be an under-estimate of the true rate of ART interruption, as individuals who resumed ART at the same facility were not included in the definition of LTFU. On the other hand, it is possible that the estimate may be exaggerated, considering that some of the women who resumed ART at other clinics might have done so soon after their last visit at the original clinic (i.e. they might not have actually interrupted ART). However, Clouse *et al* reported that the median time out of care among those resuming ART was 305 days (IQR: 100-582), which suggests that the number starting ART elsewhere without any gap in ART is likely to be relatively small.

G.4 Summary

Table G3 summarizes the estimates of the rate of ART interruption and resumption from the different studies reviewed previously. None of the studies provides a truly unbiased measure of the rate of treatment interruption. The Kaplan and Clouse studies are both likely to under-estimate the true rate because they exclude individuals who interrupted ART and then resumed ART at the same clinic. The Kranzer study also is likely to under-estimate the true rate, as it excludes patients who left the study area (who may have been more likely to interrupt ART), and it followed the patients who remained in the study area intensively (i.e. follow-up may not have been typical of that in the general public health sector). For the purpose of modelling ART interruptions in Thembisa, we assume a value of 0.25 for the annual rate of ART interruption. Further work is required to consider the uncertainty around this parameter, but for now we note that this choice of parameter value leads to model estimates of the difference between cumulative and current ART enrolment (a crude measure

of the extent of ART interruptions) that are consistent with differences between reported cumulative and current enrolment (Figure 8.10).

Table G3: Summary of South African estimates

Parameter and source	Estimate (95% CI)	Comment
Rate of 'true' ART interruption ($\lambda\theta$)		
Kaplan *et al* [441]	0.16-0.19	Likely under-estimate
Kranzer *et al* [442]	0.06	Likely under-estimate
Clouse *et al* [443]	0.10	Likely under-estimate
Rate of ART resumption (ρ)		
Kaplan *et al* [441]	0.92 (0.67-1.32)	-
Kranzer *et al* [442]	0.94 (0.46-1.65)	-
Clouse *et al* [443]	0.70 (0.28-2.60)	-

Pooling the rates of ART resumption estimated in Table G3 and weighting by the inverse of the variance gives an average ART resumption rate of 0.91 (95% CI: 0.68-1.21). We therefore assign a value of 0.90 to the ρ parameter, again noting the need for further work to consider uncertainty around this parameter.

Now consider a simplified model in which mortality after ART initiation is the same regardless of whether individuals remain on ART or interrupt ART. We define $I(t)$ to be the fraction of patients who started ART t years ago who are currently interrupting ART. This fraction can be calculated using the differential equation

$$\frac{dI(t)}{dt} = (1 - I(t))\lambda\theta - I(t)\rho$$

with the initial condition $I(0) = 0$. The analytic solution to this equation is

$$I(t) = \lambda\theta\,(1 - \exp(-(\lambda\theta + \rho)t))/(\lambda\theta + \rho)$$

Substituting the previously-assumed values of 0.25 and 0.90 for $\lambda\theta$ and ρ respectively into the equation, we get estimates of $I(t)$ of 0.054 for 3 months after ART initiation, 0.149 for 12 months, 0.196 for 24 months, 0.210 for 36 months and 0.215 for 48 months. These are the values assumed for the ι_d parameter in section 4.6. The values are slightly different from those estimated in a recent analysis, based on African tracing studies, which estimated that the fraction of survivors who had stopped or interrupted ART was 0.100 12 months after ART initiation and 0.265 5 years after ART initiation [446]. As this study excluded transient ART interruptions, the lower estimated fraction off ART at 12 months (0.100 compared to 0.152 in our model) is to be expected.

Appendix H: Switching from first-line ART to second-line ART

H.1 Adults

In an analysis of patients starting ART over the 2000-2008 period in five South African ART programmes, Fox *et al* [196] found that by five years after ART initiation, 10.1% of adults had switched from first to second line ART. This is equivalent to an annual switch rate of 2.3% when excluding the first 6 months after ART initiation (in which switches are rare). In an earlier analysis of a subset of these cohorts, Keiser *et al* [447] estimated that 9.8% of patients had switched to second-line by 3 years after ART initiation, equivalent to an annual switch rate of 4.0% (again excluding the first 6 months after ART initiation). Similarly, in an analysis of African ART programmes in which there was routine virological monitoring, Haas *et al* [448] estimated the cumulative fraction who had switched by 5 years after ART initiation was 14.0%, equivalent to an annual switching probability of 3.3%. These rates are similar to the average switching rate of 2.65 switches per 100 person years in a meta-analysis of switching rates in sub-Saharan African ART cohorts [449]; the average switching rate in this meta-analysis increased to 3.33 switches per 100 years when the analysis was limited to settings in which there was routine virological monitoring (as has been the case in South Africa).

Studies suggest that switching is more frequent in patients who start ART in more advanced HIV disease. In the analysis of Fox et al [196], the majority of patients (92%) started ART at a CD4 count of less than 200 cells/µl and although the authors did not directly report the effect of baseline CD4 count on the rate of switching, they did report that the rate of virologic failure was significantly lower in patients with higher baseline CD4 counts (for example, the rate of failure in patients with a baseline CD4 count of 350 cells/µl or higher was 0.45 times (95% CI: 0.17-1.20) that in patients with a baseline CD4 of 100-199 cells/µl). In a more recent study, based on data from the PopART trial in the Western Cape province of South Africa, Fatti *et al* [432] found that there was a more than 10-fold difference in rates of confirmed virological failure when comparing patients who started ART at CD4 counts ≥500 cells/µl with those starting at CD4 counts <200 cells/µl. Similar effects of baseline CD4 count on rates of switching/failure have been noted in other African ART cohorts [209, 450].

Evidence suggests that switching tends to be more frequent in younger adults. Adults in Khayelitsha were found to have significantly higher risks of virologic failure if they started ART below age 25 [209]. Fox *et al* [196] also found that rates of virological failure were significantly lower at older ages than at younger ages, although age did not significantly affect the rate of switching after a failure. In an analysis of Ugandan ART patients who had experienced virologic failure, older age was associated with significantly lower rates of switching in univariable analysis though not in multivariable analysis [450]. In an analysis of IeDEA data from across sub-Saharan Africa, Haas *et al* [448] also found that older age was significantly associated with a lower rate of switching.

In the Thembisa model, we assume that the annual rate of switching (after the first 6 months of ART) is 0.0230 in patients who start ART at CD4 <200 cells/µl, 0.0092 in patients with a baseline CD4 count of 200-349 cells/µl, 0.0041 in patients with a baseline CD4 count of 350-499 cells/µl, and 0.0015 in patients starting ART at CD4 counts of 500 cells/µl or higher. The

rates of switching in patients with baseline CD4 counts of 200 cells/μl or higher are calculated from relative rates of switching of 3.49, 1.40, 0.60 and 0.23 in patients starting ART at CD4 counts of <200, 200-349, 350-499 and ≥500 cells respectively, as estimated in the analysis of Fatti *et al* [432]. (We consider this a more reliable source than the earlier analysis of Fox *et al* [196] as this earlier analysis was conducted at a time when relatively few patients started ART at higher CD4 counts.) The resulting assumed fractions of patients who are on second-line ART at different ART durations are shown in Table H1. The model does not currently allow for age differences in rates of switching. The assumption of a switching rate of 0.023 per annum in patients with a baseline CD4 count of <200 cells/μl is consistent with the rate of 0.023 estimated by Fox *et al* [196], but lower than the rates estimated in other studies, which mostly included data from outside South Africa.

Table H1: Proportion of adult ART patients who are on second-line ART

Baseline	Time since first ART initiation				
CD4 count	6 months	18 months	30 months	42 months	54 months
<200	0.0%	3.6%	7.1%	10.5%	13.7%
200-349	0.0%	1.5%	2.9%	4.3%	5.7%
350-499	0.0%	0.6%	1.3%	1.9%	2.5%
500+	0.0%	0.2%	0.5%	0.7%	1.0%

With these assumptions, the Thembisa model estimates approximately 316 000 adults on second-line ART in November 2018. This is roughly consistent with programme data: the total number of adult ART patients on second-line ART in the public sector was 277 000 in November 2018 (Ruth Lancaster, personal communication), and the corresponding number in the private sector has been stable at around 33 000 in recent years [169], suggesting a total (for the public and private sectors combined) of around 310 000.

H.2 Children

An early analysis of South African paediatric ART data, based on children who started ART up to 2008, found that by 3 years after ART initiation 6.2% of children had switched to second-line, although there was minimal switching in the first year of ART [451]. Although this analysis was conducted at a time when there was limited South African guidance on switching in children, more recent analyses have produced similar results: in an analysis of data from South African and Botswana, Collins *et al* [452] found that the cumulative probability of switching by 3 years after ART initiation was 5.4%.

There is limited data on the effect of age on the rate of switching. Collins *et al* [452] found that in South Africa and Botswana, the cumulative probability of switching by 3 years after ART start was 3.7% for children who started ART below age 3 years, compared to almost 7% in children who started ART at 3 years or older. Although South African guidelines in 2010 recommended that all children initiating ART younger than 3 years start on lopinavir and subsequently switch to efavirenz on reaching the age of 3 [335], this guideline is followed only in a minority of cases (Mary-Ann Davies, personal communication), and Collins *et al* [452] found few cases of it being applied in South Africa. In an analysis of IeDEA data from South Africa, age was not found to be significantly associated with either the rate of virologic failure or the rate of switching after virologic failure, in multivariable analysis [451]. However, in an analysis of IeDEA data from across the African and Asia-Pacific regions,

older age was found to be associated with a significantly higher rate of switching to second-line in children [453].

It is also unclear if baseline disease severity affects the rate of switching. The early analysis of IeDEA data in South Africa found that although severe immune suppression and WHO clinical stage 3 or 4 at ART initiation were significantly associated with increased rates of virologic failure, both variables ceased to be significant predictors of rates of failure in multivariable analysis [451]. A more recent IeDEA analysis found that both low baseline CD4 count and advanced WHO stage at ART initiation were significantly associated with higher rates of ART failure, but failure was defined differently across cohorts, and the low overall rate of switching in this study (around 0.4% by 3 years after ART initiation) suggests limited generalizability of results to South Africa [453]. A recent analysis from the CIPHER collaboration found that children who started ART with severe immunodeficiency had a significantly higher rate of switching to second-line (aHR 1.40, 95% CI: 1.21-1.62) than in those who started ART in earlier stages of immunodeficiency [452].

In Thembisa, we assume that there is no switching from first to second line ART during the first 6 months after ART start. Thereafter the assumed annual rate of switch is 0.026 in children who start ART in late disease (based on the early analysis of switching in IeDEA South African cohorts [451], i.e. $0.026 = -\ln(1 - 0.062)/2.5$) and 0.018 in children who start ART in early disease (based on the relative rate of switch in early disease in the recent CIPHER analysis [452], i.e. $0.018 = 0.026/1.40$). Because the evidence regarding the effect of age on switching is inconsistent, we do not include any age effects in the modelling of switching.

H.3 Limitations

A limitation of the way in which we have modelled switching in Thembisa is that we assume switching is independent of mortality and ART interruptions. In reality, virological failure is more likely to occur in patients who have interrupted ART [196, 209], and virological failure is a precursor to switching to second line, which means that patients who have interrupted ART and returned to care are more likely to be on second line. Another limitation is that we lack reliable programme statistics to validate the assumptions about switching to second-line ART in children, and our estimates for children should therefore be treated with caution.