An investigation into the extent of uncertainty surrounding estimates of the impact of HIV/AIDS in South Africa

Leigh F. Johnson[a*], Rob E. Dorrington[a] and Alan P. Matthews[b]

HIV/AIDS statistics have been the source of much controversy in South Africa, but often the extent of uncertainty around these estimates is ignored. There is need for an assessment of the range of uncertainty around often-quoted HIV/AIDS statistics. This analysis determines ranges of uncertainty around the inputs and outputs of the ASSA2002 AIDS and Demographic model of the South African HIV/AIDS epidemic, using a generalized likelihood uncertainty estimation approach. A sample of 500 parameter combinations was drawn by weighting randomly generated parameter combinations by likelihood functions defined on the basis of four South African HIV/AIDS data sets. The estimated number of HIV infections in mid-2005 was 5.1 million (95% prediction interval: 4.2–6.0 million), equivalent to an HIV prevalence rate of 11.1% (9.1–13.1%). Between mid-2004 and mid-2005, the estimated number of new HIV infections was 490 000 (370 000–590 000) and the estimated number of AIDS deaths was 320 000 (270 000–380 000). The posterior mean HIV survival time was estimated to be 11.5 years (95% credibility interval: 10.0–12.9 years), longer than estimated for elsewhere in the developing world. This analysis confirms that South Africa is experiencing a severe HIV/AIDS epidemic, and suggests that HIV/AIDS epidemiology in the country probably differs from that elsewhere in Africa.

Introduction

HIV/AIDS statistics and models have been a source of much controversy in South Africa. Until recently, most HIV/AIDS models produced single estimates of the expected impact of the epidemic, with no indication being given of the uncertainty around these projections. The conclusions of different agencies have differed substantially, with estimates of the number of HIV infections in South Africa in 2005 ranging between 4.5 million[1] and 6.3 million.[2] The lack of uncertainty analysis has made the estimates of the different agencies difficult to reconcile, which in turn has undermined the credibility of HIV/AIDS models. The lack of uncertainty analysis has also made it difficult for policy makers to anticipate the range of possible HIV/AIDS scenarios that might be expected in future.

Techniques developed to date to assess uncertainty in HIV/AIDS projection models can be divided into two classes: those appropriate to models that are calibrated against HIV prevalence data, and those appropriate to models that are not calibrated against empirical data. Examples of the latter include Latin hypercube sampling,[3,4] Monte Carlo simulation[5,6] and factorial sampling.[7] Examples of the former are bootstrap methods in conjunction with a least squares approach,[8] and deriving prediction intervals from maximum likelihood estimates.[9] These techniques developed for calibrated HIV/AIDS models can be considered frequentist statistical methods, since they do not make use of prior knowledge regarding the parameters being estimated.[10] Bayesian methods, which allow explicitly for such prior knowledge, have been used to a very limited extent in HIV/AIDS projection models.[11–13]

The objective of this paper is to estimate ranges of uncertainty around key HIV/AIDS statistics in South Africa, using a Bayesian approach. This approach acknowledges uncertainty with respect to the principal epidemiological parameters, and integrates data from four South African data sets into a single statistical framework. The statistical framework is developed using the ASSA2002 AIDS and Demographic model†, a deterministic model of the HIV/AIDS epidemic in South Africa.[14] This makes it possible to derive 95% prediction intervals around the model outputs and to obtain credibility intervals for the key epidemiological parameters in the model. Although the approach is described in relation to the ASSA2002 model, it could also be applied to other calibrated HIV/AIDS models.

Method

The ASSA2002 AIDS and Demographic model is a combined cohort component projection and HIV/AIDS model, developed by the Actuarial Society of South Africa (ASSA) to simulate the HIV/AIDS epidemic in South Africa. The model has been described previously[14] and is freely available online.[15] The model, which was originally programmed in Excel and Visual Basic for Applications, was reprogrammed in C++ for the purpose of this analysis. Uncertainty is considered in relation to 24 of the model parameters. These parameters fall into four classes: parameters that determine patterns of sexual behaviour and sexual mixing; parameters determining probabilities of HIV transmission; parameters determining HIV survival in the absence of antiretroviral treatment; and parameters determining the extent of the bias in the data to which the model is calibrated. Prior distributions, representing ranges of uncertainty around these parameters, were specified for each of the 24 parameters, based on a review of the literature. A brief description of these parameters and the ASSA2002 model is provided in the supplementary material online. A more detailed explanation of the parameters and the literature on which the prior distributions were based is given in a working paper.[16]

Four data sets were used in the calibration of the ASSA2002 model. The characteristics of these four data sets are summarized in Table 1. For all four data sets, results are available separately for each sex and each five-year age band. Six sources of bias in these data were considered in the uncertainty analysis: incompleteness of vital registration (under-reporting of deaths); bias towards urban clinics in the early antenatal clinic surveys; exclusion of women attending private health facilities in the antenatal

[a]Centre for Actuarial Research, University of Cape Town, Private Bag, Rondebosch 7701, South Africa.
[b]School of Physics, University of KwaZulu-Natal, Private Bag X54001, Durban 4000.
*Author for correspondence. E-mail: leigh.johnson@uct.ac.za

†Although the ASSA2002 model is not the most recent version of the ASSA AIDS and Demographic models, it does not differ significantly from the more recent ASSA2003 version of the model.

Table 1. Characteristics of data sets used in calibration.

Data set	Years	Age range	Number of records	Ref.
HIV prevalence from antenatal clinic (ANC) surveys	1991–2004	15–39	205 865	2
Reported deaths from the Department of Home Affairs (DHA)	1997–2002	20–54	1 099 010	17
HIV prevalence from Nelson Mandela/HSRC household surveys	2002, 2005	15+	18 112	18, 19
HIV prevalence from RHRU/loveLife household survey	2003	15–24	11 904	20

HSRC = Human Sciences Research Council; RHRU = Reproductive Health Research Unit, University of the Witwatersrand.

surveys; false positive reactions on the blood-based ELISA tests used in the antenatal surveys; false positive reactions on the saliva-based ELISA tests used in the household surveys; and bias in antenatal surveys due to the effect of HIV on fertility.

The uncertainty analysis of the ASSA2002 model was conducted using an approach based on the generalized likelihood uncertainty estimation (GLUE) and sampling/importance resampling (SIR) techniques.[21,22] Prior distributions for 24 model parameters were specified, and 100 000 parameter combinations were sampled from these distributions.

For each data set, a likelihood function was specified, based on the assumption that the differences between the model predictions and the actual observations were normally distributed with zero mean. The likelihood function for the antenatal clinic data set, for example, was specified as

$$L_{ANC} = \prod_{a} \prod_{t=1991}^{2004} \frac{\exp\left(-\frac{n(a,t)(x(a,t)-M(a,t))^2}{2M(a,t)(1-M(a,t))}\right)}{\sqrt{2\pi M(a,t)(1-M(a,t))/n(a,t)}}, \quad (1)$$

where $x(a, t)$ is the antenatal HIV prevalence measured in year t in age group a, in a sample of $n(a, t)$ women, and $M(a, t)$ is the modelled prevalence in pregnant women in year t in age group a. The likelihood function thus represents the degree of correspondence between the observations and the model predictions. An aggregate likelihood function was then calculated as

$$L = {L_{ANC}}^{N_{ANC}} \times {L_{DHA}}^{N_{DHA}} \times {L_{HSRC}}^{N_{HSRC}} \times {L_{RHRU}}^{N_{RHRU}}, \quad (2)$$

where L_u is the likelihood function for data source u, and N_u is a weighting factor applied to data source u. Each N_u weighting factor was set by trial-and-error, in such a way that the 2.5 and 97.5 percentiles of sampled model outputs (drawn from a set of 10 000 model outputs by using sampling weights proportional to ${L_u}^{N_u}$) included approximately 95% of the observations in the relevant data set. Failure to include these N_u weighting factors would have resulted in prediction intervals that were too narrow relative to the variation within each data set. The approach followed in defining the likelihood function and calculating the N_u weighting factors is described in more detail in the full report.[16]

After the aggregate likelihood was defined, it was calculated for each of the 100 000 parameter combinations. A sample of 500 parameter combinations was then drawn (with replacement) from the initial set of 100 000 parameter combinations. The aggregate likelihood values (L) were used as sample weights, so that the most weight was given to those parameter combinations that produced model results consistent with the empirical data. The sample of 500 parameter combinations is therefore one from the posterior distribution of parameter combinations. More detailed model outputs were generated for these 500 parameter combinations, so that average values and 95% prediction intervals could be calculated for selected model outputs. The 500 parameter combinations sampled were also used to determine the posterior distributions for each individual parameter.

Results

Figure 1 compares the levels of HIV prevalence observed in the antenatal clinic surveys, $x(a, t)$, with $M(a, t)$, the levels of HIV prevalence, predicted by the model, in pregnant women attending public clinics. Almost all of the observed prevalence levels are contained in the 95% prediction intervals for $M(a, t)$. The model tends to underestimate recent HIV prevalence rates in the 25–29 age band, although there is reasonable consistency between modelled and observed prevalence levels in other age bands.

Actual numbers of reported deaths were compared with predicted numbers of reported deaths, for each five-year age band, each year and each sex (results not shown). The comparison of actual and modelled deaths in the 20–54 age range is shown in Fig. 2. In aggregate, the predicted trends in reported mortality are consistent with the actual numbers of reported deaths in this age range. Similar comparisons were made with HIV prevalence levels recorded in the household surveys.

The results of the model are presented in Fig. 3, up to 2005. Trends in total HIV infections and HIV prevalence are shown in Fig. 3(a) and (b), respectively. In mid-2005, the average estimated number of HIV infections in South Africa was 5.1 million, with a 95% prediction interval of 4.2 million to 6.0 million. This is equivalent to an HIV prevalence of 11.1% of the total population (9.1–13.1%). Total HIV infections are still rising in most of the 500 scenarios.

Figures 3(c) and (d) show trends in annual numbers of new HIV infections and HIV incidence rates, respectively. In most scenarios, HIV incidence rates peaked during the 1997 to 1999 period, and have been declining since then. The average estimated number of new HIV infections that occurred between mid-2004 and mid-2005 is 490 000 (370 000–590 000). The corresponding average estimated HIV incidence rate over the same period is 1.2% (0.9–1.5%).

The rise in AIDS mortality follows the increase in total HIV infections [Fig. 3(e)]. Between mid-2004 and mid-2005, the average estimated number of AIDS deaths was 320 000 (270 000–380 000). This is equivalent to an increase of 7.0 per 1000 (5.8–8.5 per 1000) in the crude mortality rate, as shown in Fig. 3(f). Associated with the rise in AIDS mortality is an increase in AIDS morbidity, shown in Fig. 3(g) and (h). The average estimated number of AIDS cases in South Africa in mid-2005 is 590 000 (500 000–680 000), and the average estimated percentage of the population sick with AIDS is 1.3% (1.1–1.5%).

Prior and posterior distributions were compared for all 24 parameters. The differences between these distributions were greatest for the average HIV survival time and the factor by which the fertility rate is reduced per year of HIV infection in women who are HIV-positive. These differences are shown in Fig. 4. The posterior distribution for the average HIV survival time, in individuals infected at age 29, has a mean of 11.5 years,

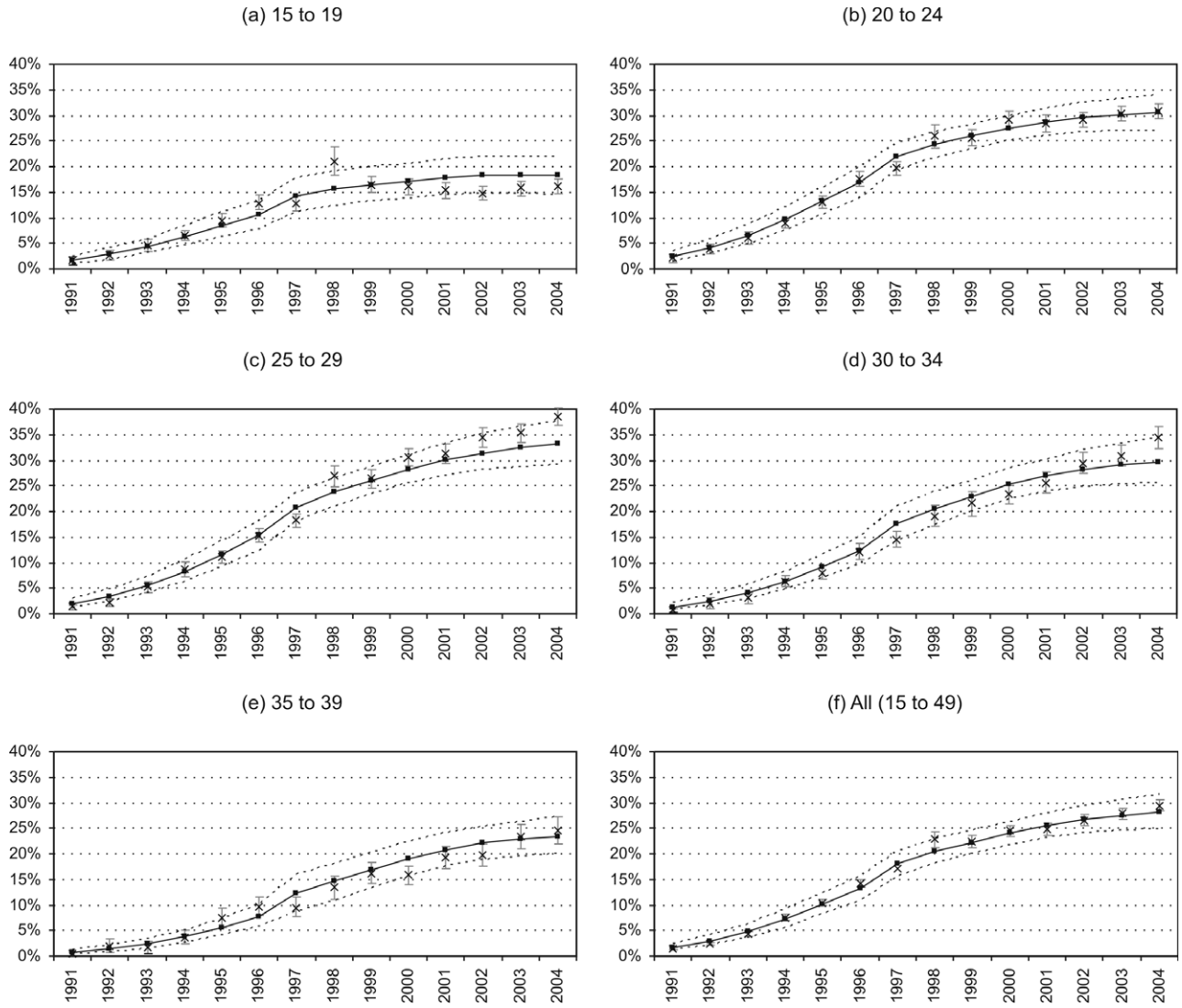

Fig. 1. Levels of HIV prevalence observed and modelled in pregnant women. Observed values are represented by crosses (×), with 95% confidence intervals. Mean of 500 simulated prevalence levels are represented by filled squares (■); 2.5 and 97.5 percentiles of simulated prevalence levels are represented by dashed lines.

with a 95% credibility interval of 10.0–12.9 years. This mean is greater than that of the prior distribution, 10.5 years (8.6–12.9). The factor by which fertility is reduced, for each year by which a woman's duration of infection increases, has a posterior mean of 0.97 (0.93–1.00), greater than the prior mean of 0.95 (0.88–0.99).

Discussion

This analysis demonstrates that, despite the biases associated with the HIV prevalence and vital registration data in South Africa, and the uncertainty regarding the many parameters in the ASSA2002 model, there is overwhelming evidence that South Africa is experiencing a severe HIV/AIDS epidemic. With at least 1000 new HIV infections and at least 700 AIDS deaths occurring in South Africa every day, there is an urgent need for concerted action and innovation in HIV prevention and treatment.

An advantage of the approach to uncertainty analysis described in this paper is that it is capable of incorporating data

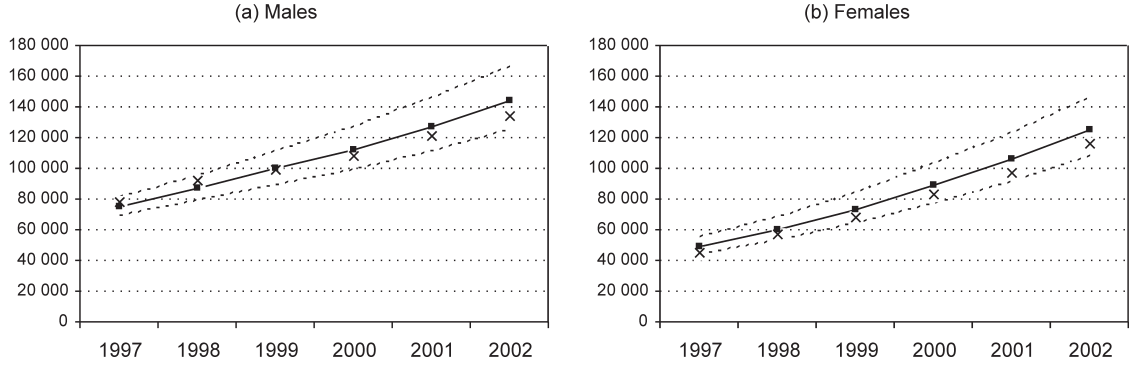

Fig. 2. Numbers of reported deaths in the 20–54 age range, predicted and actual. Actual reported numbers are represented by crosses (×). Mean of 500 simulated mortality levels are represented by filled squares (■); 2.5 and 97.5 percentiles of simulated mortality levels are represented by dashed lines.

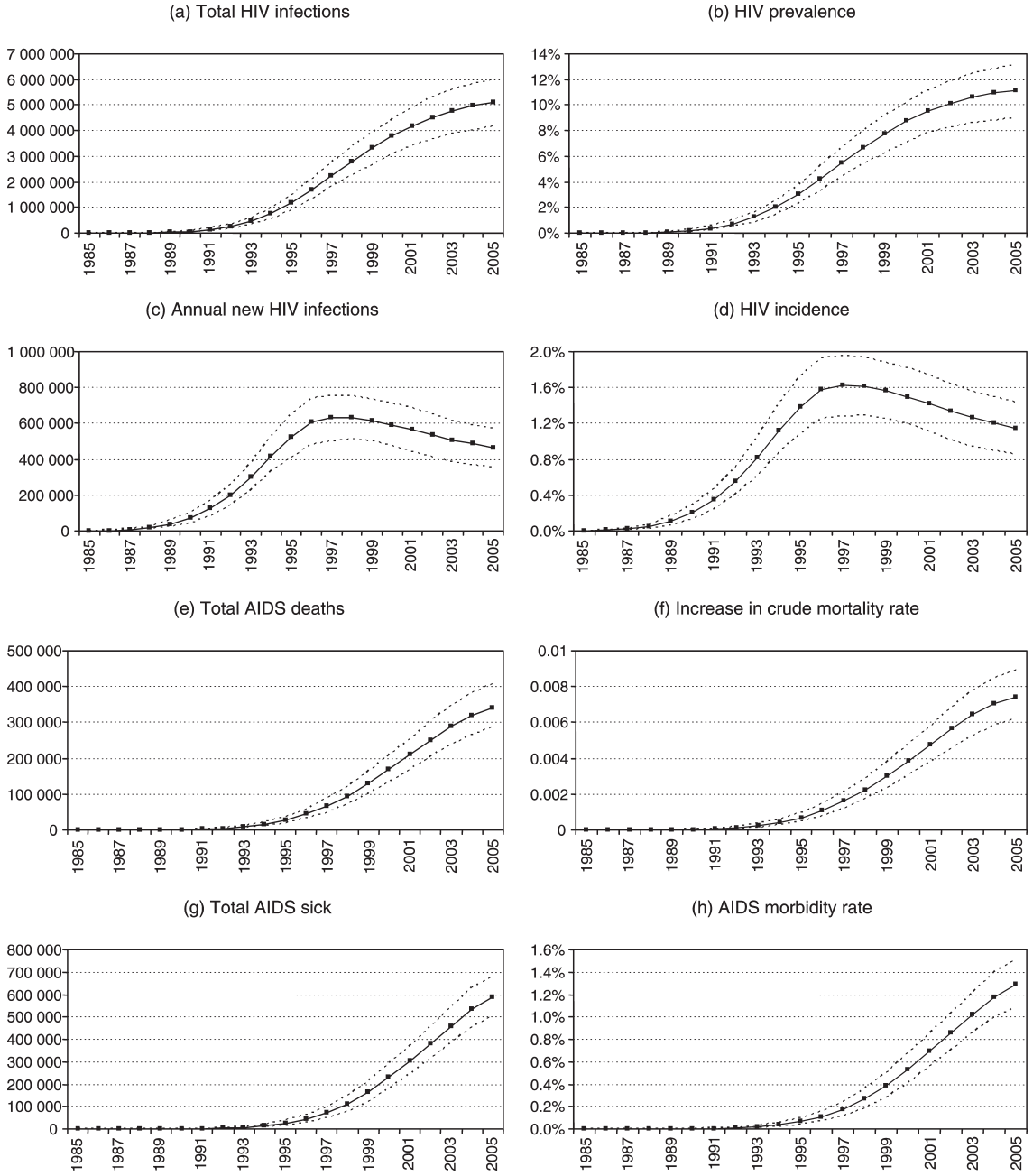

Fig. 3. Trends in HIV prevalence, HIV incidence, AIDS mortality and AIDS morbidity. Mean of 500 estimates are represented by filled squares (■); 2.5 and 97.5 percentiles of estimates are represented by dashed lines.

from multiple sources. Until recently, HIV/AIDS estimates for countries with generalized epidemics were based almost exclusively on data collected from antenatal clinics, and modelling approaches to determining prediction intervals usually did not take into account other empirical data sources.[8,9] Increasingly, however, modellers are having to validate their models against HIV prevalence data from household surveys[23] and mortality statistics.[24] It is therefore important that uncertainty analysis techniques capable of handling multiple data sources be further explored. The Bayesian framework is attractive, as it also allows for prior knowledge of HIV/AIDS epidemiology to be incorporated into the model.

This analysis suggests that rates of HIV survival in South Africa are significantly higher than those that have been observed elsewhere in the developing world,[25] a hypothesis supported by limited survival data in South Africa.[26,27] This could be due to the unique features of HIV-1 subtype C, the dominant HIV-1 subtype in South Africa. Studies have shown that individuals infected with subtype C develop the synctium-inducing HIV phenotype less frequently than those infected with other subtypes, and this phenotype is associated with more rapid disease progression.[28] In addition, comparison of non-synctium-inducing isolates suggests that subtype C is less fit than subtype B and other subtypes.[29] It has been proposed that the rapid rise in the incidence of subtype C in recent years may be an indication that HIV-1 is evolving towards a more attenuated form, with slower disease progression and thus greater opportunity for transmission.[30] A further possible explanation for the relatively long HIV survival times in South Africa is better access to prophylaxis against opportunistic infections in the local setting.[31]

Another finding from this analysis is the small effect of HIV on fertility in South Africa, relative to that observed in East Africa.[32,33] This is consistent with evidence suggesting that in African settings with low fertility and high contraceptive use, the

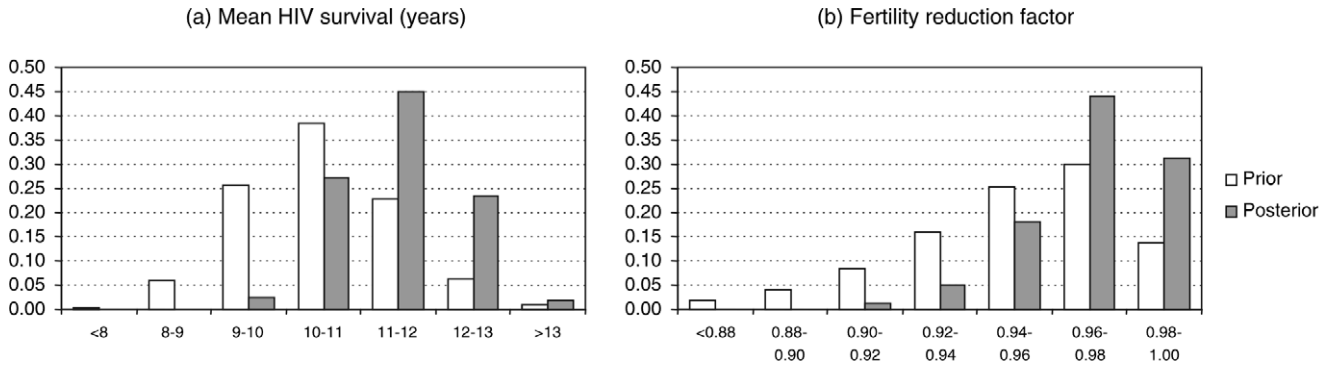

Fig. 4. Comparison of prior and posterior distributions.

impact of HIV on fertility is small.[34,35] Relative to other African countries, South Africa has low fertility and high levels of contraceptive use.[36] A 3% average annual decline in fertility, per year of HIV infection, therefore seems plausible in the South African setting.

The only other modelling group that has thus far produced HIV/AIDS estimates for South Africa together with prediction intervals is UNAIDS.[37] The UNAIDS estimate of total HIV infections in South Africa in 2005 is 5.5 million (4.9–6.1 million), which is slightly greater than the estimate of 5.1 million (4.2–6.0 million) obtained using the ASSA2002 model. However, the UNAIDS estimate of AIDS deaths in 2005 is very similar to the ASSA2002 estimate of AIDS deaths between mid-2004 and mid-2005, both estimates being 320 000 with a prediction interval of 270 000 to 380 000. This similarity is surprising, considering the higher HIV prevalence estimated, and the shorter HIV survival times assumed, by UNAIDS.[38] These differences may be offset by differences between the two models in the modelling of the recent antiretroviral rollout.

A disadvantage of the proposed approach to uncertainty analysis is that the process of setting the weighting factors by trial-and-error is complex, time-consuming and to some extent subjective. In addition, the rescaling of the likelihood by these weighting factors implies that the 'likelihood function' is not a likelihood in the formal statistical sense, but rather a somewhat arbitrary measure of goodness of fit. Other Bayesian techniques, such as Bayesian melding,[39] could be explored as possible alternatives to the GLUE approach employed here. One way in which Bayesian melding generalizes the standard Bayesian approach is by allowing for prior distributions on model outputs, and this may be an alternative means of representing the uncertainty around the bias in the data to which the model is calibrated. In countries in which HIV prevalence data are collected only from sentinel sites, rather than through nationally representative surveys, a hierarchical Bayesian approach may be more appropriate than the one used here.[13]

This uncertainty analysis considers only parameter uncertainty; uncertainty relating to the model structure is not reflected. An example of model uncertainty is the ASSA2002 model assumption that, conditional upon age and sex, non-AIDS mortality and AIDS mortality are independent. If HIV-infected individuals are more exposed to health risks like smoking and alcohol than HIV-negative persons,[40,41] or are of a lower socio-economic status than HIV-negative individuals,[42–44] the independence assumption is likely to be invalid and may lead to some over-estimation of aggregate mortality. Models that relax the independence assumption could be developed, and techniques such as Bayesian Model Averaging could be used to extend the analysis to reflect uncertainty regarding the choice of model.[45]

This uncertainty analysis was limited to a subset of 24 parameters in the ASSA2002 model. Further work is required to incorporate uncertainty regarding mother-to-child transmission of HIV and paediatric HIV survival, as well as uncertainty regarding demographic parameters, which may require alternative statistical techniques.[46] Results have been shown only up to 2005, as uncertainty regarding the future impact of antiretroviral treatment has not yet been incorporated. Projections of uncertainty beyond 2005 will require an assessment of uncertainty relating to antiretroviral treatment and new prevention strategies such as male circumcision, microbicides and vaccines.

This research is funded by the South African AIDS Vaccine Initiative (SAAVI), a consortium with primary funding from the Department of Health, the Department of Science and Technology, and Eskom. We also thank Carl Lombard, Debbie Bradshaw and Lize van der Merwe for their comments on earlier versions of this paper.

Received 12 December 2006. Accepted 10 April 2007.

1. Statistics South Africa (2005). *Mid-year population estimates, South Africa, 2005.* Online: http://www.statssa.gov.za
2. Department of Health (2005). *National HIV and syphilis antenatal sero-prevalence survey in South Africa 2004.* Online: http://www.doh.gov.za/docs/reports/2004/hiv-syphilis.pdf
3. Blower S. M. and Dowlatabadi H. (1994). Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int. Stat. Rev.* **62**(2), 229–243.
4. Blower S.M., Gershengorn H.B. and Grant R.M. (2000). A tale of two futures: HIV and antiretroviral therapy in San Francisco. *Science* **287**, 650–654.
5. Davenport M.P., Ribeiro R.M., Chao D.L. and Perelson A.S. (2004). Predicting the impact of a nonsterilizing vaccine against human immunodeficiency virus. *J. Virol.* **78**(20), 11340–11351.
6. Law M.G., Prestage G., Grulich A., Van de Ven P. and Kippax S. (2001). Modelling the effects of combination antiretroviral treatments on HIV incidence. *AIDS* **15**, 1287–1294.
7. Boily M., Bastos F.I., Desai K. and Mâsse B. (2004). Changes in the transmission dynamics of the HIV epidemic after the wide-scale use of antiretroviral therapy could explain increases in sexually transmitted infections. *Sex. Transm. Dis.* **31**(2), 100–113.
8. Grassly N.C., Morgan M., Walker N., Garnett G., Stanecki K.A., Stover J., Brown T. and Ghys P.D. (2004). Uncertainty in estimates of HIV/AIDS: the estimation and application of plausibility bounds. *Sex. Transm. Infect.* **80** (Suppl. 1), i31–38.
9. Salomon J.A. and Murray C.J. (2001). Modelling HIV/AIDS epidemics in sub-Saharan Africa using seroprevalence data from antenatal clinics. *Bull. WHO* **79**(7), 596–607.
10. Greenland S. (2000). Principles of multilevel modelling. *Int. J. Epidemiol.* **29**(1), 158–167.
11. Gilks W.R., De Angelis D. and Day N.E. (1999). Bayesian conditional-independence modeling of the AIDS epidemic in England and Wales. *Physica D* **133**(1–4), 145–151.
12. Tan W.Y. and Ye Z.Z. (2000). Estimation of HIV infection and incubation via state space models. *Math. Biosci.* **167**(1), 31–50.
13. Alkema L., Raftery A.E. and Clark S.J. (2007). *Probabilistic projections of HIV prevalence using Bayesian melding.* Centre for Studies in Demography and Ecology, University of Washington. Working paper no. 07-01. Online: http://csde.washington.edu/downloads/07-01.pdf
14. Johnson L.F. and Dorrington R.E. (2006). Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions. *Demographic Res.* **14**, 541–74.

15. Actuarial Society of South Africa (2004). *ASSA2002 AIDS and Demographic Model.* Online: http://www.assa.org.za/default.asp?id=1000000050
16. Johnson L.F., Dorrington R.E. and Matthews A. (2006). *An uncertainty analysis and sensitivity analysis of the ASSA2002 AIDS and Demographic model.* Centre for Actuarial Research, University of Cape Town. Working paper. Online: http://www.commerce.uct.ac.za/Research_Units/CARE/
17. Statistics South Africa (2005). *Mortality and causes of death in South Africa, 1997–2003: Findings from death notification.* Online: http://www.statssa.gov.za
18. Connolly C., Shisana O., Colvin M. and Stoker D. (2004). Epidemiology of HIV in South Africa – results of a national, community-based survey. *S. Afr. Med. J.* **94**(9), 776–781.
19. Shisana O., Rehle T., Simbayi L.C., Parker W., Zuma K., Bhana A., Connolly C., Jooste S. and Pillay V. (2005). *South African National HIV Prevalence, HIV Incidence, Behaviours and Communication Survey, 2005.* HSRC Press, Cape Town. Online: http://www.hsrcpress.ac.za
20. Pettifor A.E., Rees H.V., Kleinschmidt I., Steffenson A.E., Macphail C., Hlongwa-Madikizela L., Vermaak K. and Padian N.S. (2005). Young people's sexual health in South Africa: HIV prevalence and sexual behaviors from a nationally representative household survey. *AIDS* **19**(14), 1525–1534.
21. Beven K. and Binley A. (1992). The future of distributed models – model calibration and uncertainty prediction. *Hydrol. Process.* **6**(3), 279–298.
22. Smith A.F.M. and Gelfand A.E. (1992). Bayesian statistics without tears – a sampling resampling perspective. *Am. Stat.* **46**(2), 84–88.
23. Boerma J.T., Ghys P.D. and Walker N. (2003). Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard. *Lancet* **362**(9399), 1929–1931.
24. Timæus I.M. and Jasseh M. (2004). Adult mortality in sub-Saharan Africa: evidence from Demographic and Health Surveys. *Demography* **41**(4), 757–772.
25. Porter K. and Zaba B. (2004). The empirical evidence for the impact of HIV on adult mortality in the developing world: data from serological studies. *AIDS* **18** (Suppl. 2), S9–S17.
26. Glynn J.R., Sonnenberg P., Nelson G., Bester A., Shearer S. and Murray J. (2005).The effect of HIV on adult mortality: evidence from a large cohort of South African gold-miners with known dates of seroconversion and 10 years of follow-up. International Union for the Scientific Study of Population Conference, Tours, France, 18–23 July 2005.
27. Maartens G., Wood R., O'Keefe E. and Byrne C. (1997). Independent epidemics of heterosexual and homosexual infection in South Africa – survival differences. *Q. J. Med.* **90**, 449–454.
28. Peeters M., Toure-Kane C. and Nkengasong J.N. (2003). Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. *AIDS* **17**, 2547–2560.
29. Ball S.C., Abraha A., Collins K.R., Marozsan A.J., Baird H., Quinones-Mateu M.E., Penn-Nicholson A., Murray M., Richard N., Lobritz M., Zimmerman P.A., Kawamura T., Blauvelt A. and Arts E.J. (2003). Comparing the *ex vivo* fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. *J. Virol.* **77**(2), 1021–1038.
30. Quiñones-Mateu M.E. (2005). Is HIV-1 evolving to a less virulent (pathogenic) virus? *AIDS* **19**(15), 1689–1690.
31. Badri M., Bekker L.G., Orrell C., Pitt J., Cilliers F. and Wood R. (2004). Initiating highly active antiretroviral therapy in sub-Saharan Africa: an assessment of the revised World Health Organization scaling-up guidelines. *AIDS* **18**(8), 1159–1168.
32. Heuveline P. (2003). HIV and population dynamics: a general model and maximum-likelihood standards for East Africa. *Demography* **40**(2), 217–245.
33. Zaba B. and Gregson S. (1998). Measuring the impact of HIV on fertility in Africa. *AIDS* **12** (Suppl. 1), S41–S50.
34. Lewis J.J.C., Ronsmans C., Ezeh A. and Gregson S. (2004). The population impact of HIV on fertility in sub-Saharan Africa. *AIDS* **18** (Suppl. 2), S35–S43.
35. Gregson S., Terceira N., Kakowa M., Mason P., Anderson R., Chandiwana S. and Caraël M. (2002). Study of bias in antenatal clinic HIV-1 surveillance data in a high contraceptive prevalence population in sub-Saharan Africa. *AIDS* **16**, 643–652.
36. Moultrie T.A. and Timæus I.M. (2003). The South African fertility decline: evidence from two censuses and a Demographic and Health Survey. *Pop. Stud.* **57**(3), 265–283.
37. UNAIDS (2006). *2006 Report on the global AIDS epidemic.* Online: http://www.unaids.org/en/HIV_data/2006GlobalReport/default.asp
38. Morgan M., Walker N., Gouws E., Stanecki K.A. and Stover J. (2006). Improved plausibility bounds about the 2005 HIV and AIDS estimates. *Sex. Transm. Infect.* **82** (suppl. 3), iii71–7.
39. Poole D. and Raftery A.E. (2000). Inference for deterministic simulation models: the Bayesian melding approach. *J. Am. Stat. Assoc.* **95**(452), 1244–1255.
40. Furber A.S., Maheswaran R., Newell J N. and Carroll C. (2007). Is smoking tobacco an independent risk factor for HIV infection and progression to AIDS? A systemic review. *Sex. Transm. Infect.* **83**(1), 41–46.
41. Simbayi L.C., Kalichman S.C., Jooste S., Mathiti V., Cain D. and Cherry C. (2004). Alcohol use and sexual risks for HIV infection among men and women receiving sexually transmitted infection clinic services in Cape Town, South Africa. *J. Stud. Alcohol.* **65**(4), 434–442.
42. Shisana O., Zungu-Dirwayi N., Toefy Y., Simbayi L.C., Malik S. and Zuma K. (2004). Marital status and risk of HIV infection in South Africa. *S. Afr. Med. J.* **94**(7), 537–543.
43. Evian C., Fox M., MacLeod W., Slotow S.J. and Rosen S. (2004). Prevalence of HIV in workforces in southern Africa, 2000–2001. *S. Afr. Med. J.* **94**(2), 125–130.
44. Hallman K. (2004). *Socioeconomic disadvantage and unsafe sexual behaviors among young women and men in South Africa.* Population Council, Policy Research Division. Working Paper no. 190.
45. Hoeting J.A., Madigan D., Raftery A.E. and Volinsky C.T. (1999). Bayesian Model Averaging: a tutorial. *Stat. Sci.* **14**(4), 382–417.
46. Lutz W. and Goldstein J.R. (2004). Introduction: How to deal with uncertainty in population forecasting? *Int. Stat. Rev.* **72**(1), 1–4.

Supplementary material to:

Johnson L.F., Dorrington R.E. and Matthews A.P. (2006). An investigation into the extent of uncertainty surrounding estimates of the impact of HIV/AIDS in South Africa. *S. Afr. J. Sci.* **102**, 135–140.

ASSA2002 model description and summary of parameters

The ASSA2002 AIDS and Demographic model is a combined cohort component projection and HIV/AIDS model, developed by the Actuarial Society of South Africa (ASSA) to describe the HIV/AIDS epidemic in South Africa. Two modes of HIV transmission are modelled: heterosexual transmission and mother-to-child transmission. The model contains a large number of parameters, which can be broadly grouped into the following classes:

- Demographic parameters
- Sexual behaviour parameters
- Heterosexual HIV transmission parameters
- Adult HIV survival parameters
- Paediatric HIV transmission and survival parameters
- HIV prevention and treatment parameters
- Calibration parameters.

The non-demographic parameters are described in detail elsewhere;[1] the purpose of this appendix is to provide only an overview of the model and its parameters. The uncertainty analysis is restricted to a subset of the sexual behaviour parameters, heterosexual HIV transmission parameters, adult HIV survival parameters, and calibration parameters. These parameters and the prior distributions chosen for these parameters are summarized in Table A1.

Sexual behaviour is modelled by dividing the sexually active population, between the ages of 14 and 59, into four risk groups. The 'PRO' group consists of sex workers and their regular clients, while the 'STD' group consists of other individuals who are regularly infected with sexually transmitted diseases (STDs). Although not regularly infected with STDs, individuals in the 'RSK' group are assumed to be at risk of HIV infection. Individuals in the 'NOT' group are assumed not to be at risk of HIV infection, either because they are abstaining from sexual intercourse or because they are in monogamous long-term relationships with partners who remain faithful to them. Individuals in the STD group can interact sexually with individuals in the PRO and RSK groups, but no other interactions between risk groups are assumed to occur. The 'sexual mixing' parameter determines the extent to which the STD group interacts with the PRO and RSK groups, and can take on any value between 0 and 1 (0 implying no mixing between risk groups, and 1 implying random mixing[2]). Different annual numbers of partners and frequencies of sex are assumed for each risk group. A two-parameter function is used to determine the 'sex activity curve' for women, which is a measure of the relative frequency of sex and the relative rate of partner change at different ages. The two parameters in this function determine the earliest age at which sexual activity begins (the 'position factor') and the amount of sex at young ages relative to older ages (the 'shape factor'). Most sexual behaviour parameters are specified for females, and male parameters are calculated to be consistent with these.

Adults infected with HIV are assumed to progress through four stages of disease before dying from AIDS, in the absence of antiretroviral treatment. These four stages correspond to the four stages of the WHO Clinical Staging System,[3] with the first two stages representing asymptomatic infection, the third stage representing pre-AIDS symptoms, and the fourth stage representing AIDS. The time spent in each stage of disease is assumed to follow a Weibull distribution, and the Weibull shape parameters of these distributions are assumed to be linearly related to the Weibull means, with a 'gradient parameter' determining the slope of this linear relationship. For a specified mean survival

Table A1. Prior distributions for the 24 parameters included in the uncertainty analysis.

Parameter	Prior distribution	Prior (mean, 95% interval)
Sexual behaviour parameters		
% of initial population in STD risk group	Beta (19.8, 79.2)	20% (13–28%)
% of initial population in RSK risk group	Beta (19.0, 63.7)	23% (15–33%)
Annual average number of partners (female PRO)	Gamma (25, 10)	250 (162–357)
Annual average number of partners (female STD)	Gamma (25, 0.48)	12.0 (7.8–17.1)
Annual average number of partners (female RSK)	Gamma (25, 0.04)	1.00 (0.65–1.43)
Sexual mixing parameter	Beta (12, 12)	0.50 (0.31–0.69)
Shape factor (of sex activity curve)	Beta (24.9, 4949)	0.0050 (0.0032–0.0071)
Coital reduction factor (*K*)	Uniform (0, 1)	0.50 (0.025–0.975)
% reduction in annual number of partners due to social marketing programmes	Beta (0.8, 7.2)	10.0% (0.1–36.7%)
Heterosexual HIV transmission parameters		
Probability of HIV transmission per act of sex		
Male to female (both partners in RSK group)	Beta (24.95, 12449)	2.0 (1.3–2.9) per 1000
Female to male (both partners in RSK group)	Beta (24.97, 24949)	1.0 (0.6–1.4) per 1000
Male to female (partners in STD/PRO groups)	Beta (24.82, 3521)	7.0 (4.5–10.0) per 1000
Female to male (partners in STD/PRO groups)	Beta (24.87, 4949)	5.0 (3.2–7.1) per 1000
Increase in infectiousness per log increase in HIV viral load	Gamma (6.25, 0.28)	175% (66–337%)
Increase in susceptibility per year below age 25	Gamma (16, 0.0075)	12.0% (6.9–18.6%)
Adult HIV survival parameters		
Mean HIV survival time (for individual infected at age 29), in years	Gamma (110, 0.095)	10.5 (8.6–12.5)
Decrease in mean survival time per 10-year increase in age at HIV acquisition, in years	Gamma (25, 0.06)	1.50 (0.97–2.14)
Gradient parameter	Gamma (4, 0.075)	0.30 (0.08–0.66)
Calibration parameters		
% of adult deaths reported in 1999	Beta (179.8, 34.2)	84% (79–89%)
Ratio of antenatal HIV prevalence in all public clinics to that in sampled clinics, in 1990	Beta (19.2, 4.8)	0.80 (0.62–0.93)
Odds ratio for HIV in women attending private clinics (controlling for age and race)	Beta (5.83, 2.50)	0.70 (0.37–0.94)
Axsym false positive rate	Exponential (50)	2.0% (0.0–7.4%)
Orasure false positive rate	Exponential (50)	2.0% (0.0–7.4%)
Factor by which fertility reduces per year of HIV infection	Beta (49.19, 2.59)	0.95 (0.88–0.99)

Table A2. Parameters not included in the uncertainty analysis.

Demographic parameters

- Initial population size and initial age- and sex-distribution of the population (in the year in which the epidemic starts)
- Non-AIDS mortality rates (by age, sex and year)
- Net migrants into South Africa (by age, sex and year)
- Fertility rates (by age and year)
- Proportion of births that are male

Sexual behaviour parameters

- % of sexually active individuals initially in the 'PRO' group
- Number of HIV-infected individuals temporarily entering the PRO group per annum, from outside South Africa
- Proportion of migrants to/from South Africa in each risk group
- Year in which HIV-infected migrants first enter South Africa ('start year')
- Rates at which youth become sexually experienced
- Correlation between age at sexual debut and propensity for sexual risk behaviour
- Coital frequencies within partnerships (by risk group of both partners)
- Rates of condom use (by age and risk group)
- 'Position factor' for sex activity curve
- Age of partner preferences

Heterosexual HIV transmission parameters

- Average level of HIV viral load in each disease stage
- Condom effectiveness

Adult HIV survival parameters

- % of adult HIV survival time spent in each disease stage

Paediatric HIV transmission and survival parameters

- Rates of mother-to-child transmission (before or at birth and through breastfeeding)
- Rates of HIV survival and rates of progression to AIDS in children infected at or before birth
- Rates of HIV survival and rates of progression to AIDS in children infected through breastfeeding

HIV prevention and treatment parameters

- Extent to which social marketing programmes have been phased in
- Extent of improvement in condom usage due to social marketing programmes
- Phase-in of syndromic management of STDs and effectiveness in reducing HIV transmission
- Phase-in of VCT and effectiveness in increasing condom use and reducing coital frequencies (by HIV stage)
- Phase-in and effectiveness of PMTCT
- Phase-in of HAART and levels of mortality and morbidity after starting HAART
- Change in sexual behaviour after starting HAART
- Effect of HAART on VCT uptake and condom use in general population
- Log reduction in HIV viral load after starting HAART

Calibration parameters

- % increase in proportion of deaths that are reported, per annum
- Sensitivity of HIV antibody tests
- Correlation between HIV risk behaviour and pregnancy at young ages
- % of women attending private health facilities (by age and race)

time, the gradient parameter determines the variance of the survival time; an increase in the gradient parameter implies a reduction in the variance. The mean HIV survival time is assumed to depend on the age at which individuals become HIV-infected, with individuals infected at older ages progressing to death more rapidly than those infected at young ages.[4]

The probability of heterosexual transmission of HIV, per act of sex with an infected partner, is assumed to depend on several factors. The most important of these are the sex of the susceptible partner and the risk groups of the susceptible and infected partners. It is also assumed that the disease stage of the infected partner affects the probability of HIV transmission. The levels of infectiousness in the different stages are estimated by assuming average levels of HIV viral load in each disease stage,[5,6] as well as a factor by which the probability of HIV transmission increases per log increase in viral load.[7,8] It is also assumed that in young women, age affects HIV susceptibility, with susceptibility increasing by a particular factor for each year below the age of 25.[9,10]

Four HIV prevention programmes are allowed for in the model: social marketing, syndromic management of STDs, voluntary counselling and testing (VCT), and prevention of mother-to-child transmission (PMTCT). These interventions are assumed to be introduced at rates consistent with public health sector statistics.[11–13] Social marketing programmes are assumed to lead to increases in condom usage and reductions in average annual numbers of partners.

The effects of highly active antiretroviral treatment (HAART) are modelled by adding two stages to the basic four-stage model of adult HIV survival: one representing people receiving HAART and the other representing people who have discontinued treatment. Treatment is assumed to be started at the time of the first AIDS-defining illness, and assumptions about proportions of individuals starting HAART have been set to be consistent with reported numbers of individuals receiving HAART.[14] The

frequency of sex in the different stages of disease is assumed to depend on both the proportion of adults who know their HIV status and the severity of symptoms in the different disease stages.[15–18] Symptoms are assumed to be most severe in individuals who have untreated AIDS and individuals who have discontinued HAART. Beta priors are used to represent the extent of uncertainty surrounding the reduction in coital frequencies due to HIV symptoms in each stage, but to ensure that the reductions in the different stages vary proportionally to one another across the different simulations, the same coital reduction factor, *K,* is used to sample from each of these beta distributions.

For the purpose of calibrating the model to HIV prevalence data and vital registration data, it is necessary to take into account various sources of bias associated with these data. In the case of the vital registration data, comparison of the modelled and actual deaths is only valid if the modelled deaths are adjusted to make some allowance for incompleteness of the reporting of deaths.[19] In the case of HIV prevalence data, it is necessary to allow for possible false positive reactions produced by the HIV tests used,[20–22] if there is no confirmatory testing of positive test results. The blood test used in the antenatal survey (the Abbott Axsym ELISA) differs from the saliva test used in the 2002 HSRC and 2003 RHRU surveys, which is based on the Orasure collection device. Separate priors have therefore been specified for the false positive rates on these two tests. Antenatal survey data are also biased due to the effect of HIV on fertility,[23] and the model therefore assumes that fertility in HIV-infected women reduces exponentially, relative to that in HIV-negative women, the longer they have been infected with HIV.

In the case of the antenatal clinic survey data, it is also necessary to allow for two sources of sampling bias: bias towards urban antenatal clinics in the early years of the survey, and the effect of not including women seeking antenatal care in private health facilities. The former source of bias is significant because HIV prevalence is substantially higher in urban areas than in rural areas, particularly in the early stages of an HIV/AIDS epidemic. It is therefore assumed that the ratio of antenatal prevalence in all public clinics to that in sampled clinics, in 1990, is less than one, and that this ratio increases linearly to one in 1999, by which time a new survey protocol had been introduced to remove the urban bias.[24] The effect of not including women attending private antenatal clinics can be estimated if assumptions are made about the proportion of pregnant women attending private clinics in each age and race group, and the independent effects of race and private clinic attendance on a woman's odds of HIV infection. As the latter effect is difficult to estimate reliably, it has been included in the uncertainty analysis.

The prior distributions for the 24 parameters included in the uncertainty analysis are summarized in Table A1. In general, beta priors are used for those parameters restricted to the range [0, 1] and gamma priors are used for those parameters restricted to the range [0, ∞). Exponential priors are used for the false positive rates, as the exponential distribution matches the strongly skewed distribution of false positive results more closely than the beta distribution. Values are sampled from the different prior distributions independently of one another, so that there is no dependency between the prior distributions.

A large number of ASSA2002 parameters have not been included in this uncertainty analysis. In most cases, the excluded parameters are omitted because they are believed to have little effect on the correspondence between model outputs and observed levels of HIV prevalence and mortality, or because the empirical evidence used to determine the parameters is sufficiently reliable to warrant treating the parameters as known. As this analysis considers uncertainty only in projections up to 2005, parameters that are only of significance beyond this date (e.g. HAART rollout and effectiveness) are also excluded from the analysis. The parameters that have not been included in the uncertainty analysis are summarized in Table A2.

1. Johnson L.F. and Dorrington R.E. (2006). Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions. *Demographic Res.* **14**, 541–74.
2. Garnett G. and Anderson R. (1996). Sexually transmitted diseases and sexual behaviour: insights from mathematical models. *J. Infect. Dis.* **174** (Suppl. 2), S150–S160.
3. WHO International Collaborating Group for the study of the WHO Staging System (1993). Proposed 'World Health Organization staging system for HIV infection and disease': preliminary testing by an international collaborative cross-sectional study. *AIDS* **7**, 711–718.
4. Collaborative Group on AIDS Incubation and HIV Survival (2000). Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *Lancet* **355**, 1131–7.
5. Kassa E., Rinke de Wit T.F., Hailu E., Girma M., Messele T., Mariam H., Yohannes S., Jurriaans S., Yeneneh H. and Coutinho R.A. (1999). Evaluation of the World Health Organization staging system for HIV infection and disease in Ethiopia: association between clinical stages and laboratory markers. *AIDS* **13**, 381–389.
6. Hubert J., Burgard M., Dussaix E., Tamalet C., Deveau C., Le Chenadec J., Chaix M., Marchadier E., Vilde J., Delfraissy J., Meyer L. and Rouzioux C. (2000). Natural history of serum HIV-1 RNA levels in 330 patients with a known date of infection. *AIDS* **14**, 123–31.
7. Quinn T., Wawer M., Sewankambo N., Serwadda D., Li C., Wabwire-Mangen F., Meehan M., Lutalo T. and Gray R. (2000). Viral load and heterosexual transmission of human immunodeficiency virus type 1. *N. Engl. J. Med.* **342**, 921–929.
8. Fideli U.S., Allen S.A., Musonda R., Trask S., Hahn B.H., Weiss H., Mulenga J., Kasolo F., Vermund S.H. and Aldrovandi G.M. (2001). Virologic and immunologic determinants of heterosexual transmission of human immunodeficiency virus type 1 in Africa. *AIDS Res. Hum. Retroviruses* **17**(10), 901–910.
9. Carpenter L.M., Kamali A., Ruberantwari A., Malamba S.S. and Whitworth J.A. (1999). Rates of HIV-1 transmission within marriage in rural Uganda in relation to the HIV sero-status of the partners. *AIDS* **13**(9), 1083–1089.
10. Gray R., Wawer M., Brookmeyer R., Sewankambo N., Serwadda D., Wabwire-Mangen F., Lutalo T., Li X., van Cott T. and Quinn T. (2001). Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet* **357**, 1149–1153.
11. Ramkissoon A., Kleinschmidt I., Beksinska M., Smit J., Hlazo J. and Mabude Z. (2004). *National Baseline Assessment of Sexually Transmitted Infection and HIV Services in South African Public Sector Health Facilities.* Reproductive Health Research Unit, University of the Witwatersrand, Johannesburg.
12. McCoy D., Besser M., Visser R. and Doherty T. (2002). *Interim findings on the national PMTCT pilot sites: lessons and recommendations.* Health Systems Trust, Durban.
13. Myer L. (2005). Barrier methods. In *HIV/AIDS in South Africa,* eds S.S. Abdool Karim and Q. Abdool Karim. Cambridge University Press, Cape Town.
14. Department of Health (2006). *National Comprehensive HIV and AIDS Plan Statistics.*
15. Ross A., Van der Paal L., Lubega R., Mayanja B.N., Shafer L.A. and Whitworth J. (2004). HIV-1 disease progression and fertility: the incidence of recognized pregnancy and pregnancy outcome in Uganda. *AIDS* **18**, 799–804.
16. Terceira N., Gregson S., Zaba B. and Mason P.R. (2003). The contribution of HIV to fertility decline in rural Zimbabwe, 1985–2000. *Pop. Stud.* **57**(2), 149–164.
17. Hankins C., Tran T. and Lapointe N. (1998). Sexual behaviour and pregnancy outcome in HIV-infected women. *J. Acquir. Immun. Defic. Syndr.* **18**, 479–487.
18. Wilson T.E., Gore M.E., Greenblatt R., Cohen M., Minkoff H., Silver S., Robison E., Levine A. and Gange S.J. (2004). Changes in sexual behaviour among HIV-infected women after initiation of HAART. *Am. J. Public Health* **94**(7), 1141–1146.
19. Dorrington R.E., Moultrie T.A. and Timæus I.M. (2004). *Estimation of Mortality using the South African Census 2001 Data.* Monograph 11. Centre for Actuarial Research, University of Cape Town.
20. Nkengasong J N., Maurice C., Koblavi S., Kalou M., Yavo D., Maran M., Bile C., N'guessan K., Kouadio J., Bony S., Wiktor S.Z. and Greenberg A.E. (1999). Evaluation of HIV serial and parallel serologic testing algorithms in Abidjan, Côte d'Ivoire. *AIDS* **13**(1), 109–117.
21. Meda N., Gautier-Charpentier L., Soudré R.B., Dahourou H., Ouedraogo-Traoré R., Ouangré A., Bambara A., Kpozehouen A., Sanou H., Valéa D., Ky F., Cartoux M., Barin F. and van de Perre P. (1999). Serological diagnosis of HIV in Burkina Faso: reliable, practical strategies using less expensive commercial test kits. *Bull. WHO* **77**(9), 731–739.
22. Urassa M., Godoy K., Killewo J., Kwesigabo G., Mbakileki A., Mhalu F. and Biberfeld G. (1999). The accuracy of an alternative confirmatory strategy for detection of antibodies to HIV-1: experience from a regional laboratory in Kagera, Tanzania. *J. Clin. Virol.* **14**(1), 25–29.
23. Lewis J.J.C., Ronsmans C., Ezeh A. and Gregson S. (2004). The population impact of HIV on fertility in sub-Saharan Africa. *AIDS* **18** (Suppl. 2), S35–S43.
24. Department of Health (2001). *National HIV and syphilis sero-prevalence survey of women attending public antenatal clinics in South Africa: 2000.* Pretoria.