



ASSA2008 AIDS AND DEMOGRAPHIC MODELS

USER GUIDE (beta version)

Originally prepared by Rob Dorrington, Leigh Johnson and Debbie Budlender of the Centre for Actuarial Research, University of Cape Town, for the AIDS Committee of the Actuarial Society of South Africa

(Feedback welcome: Comments to be sent to aid.actuarialsociety.org.za)

July 2010

Table of contents

Table of contents	i
1. Introduction	1
2. Overview of AIDS modelling	2
DISCLAIMER.....	4
3. The structure of the model	6
3.1 DIVISION INTO SUBGROUPS	6
3.2 PROCESS OF INFECTION	7
3.2.1 <i>Introduction</i>	7
3.2.2 <i>Starting the epidemic</i>	9
3.2.3 <i>Infection assumptions</i>	9
3.2.4 <i>Paediatric infections</i>	11
3.3 STARTING POPULATION	11
3.4 MORTALITY	11
3.4.1 <i>Non-HIV mortality</i>	11
3.4.2 <i>Survival of adults with HIV</i>	12
3.4.3 <i>Survival of children with HIV</i>	13
3.5 FERTILITY	14
3.5.1 <i>Non-HIV</i>	14
3.5.2 <i>HIV and fertility</i>	14
3.5.3 <i>Overall</i>	14
3.6 MIGRATION.....	15
3.7 POPULATION GROUPS	15
3.8 INTERVENTIONS AND BEHAVIOUR CHANGE	16
4. Running the model.....	18
4.1 STARTING THE EPIDEMIC	18
4.2 PROJECTING	18
5. Understanding and adapting outputs.....	20
5.1 STANDARD OUTPUTS	20
5.2 ADDING OUTPUT STATISTICS	21
6 The provincial and urban-rural versions of the model	23
6.1 THE PROVINCIAL VERSION	23
6.2 THE URBAN-RURAL VERSION.....	23
7 Changing assumptions and parameters in the model.....	24
7.3 CALIBRATION TIPS	25
Appendices	27
APPENDIX A: SYSTEM REQUIREMENTS	27
APPENDIX B: DESCRIPTION OF WORKSHEETS.....	28
<i>Assumptions</i>	28
<i>Results</i>	29

<i>Interventions</i>	30
<i>Population</i>	30
<i>MortTable</i>	30
<i>SexActivity</i>	31
<i>Male HIVTable and Female HIVTable</i>	31
<i>HIVTable</i>	32
<i>Male and Female Adult Survival</i>	32
<i>Male and Female Staging</i>	33
<i>Paediatric Survival</i>	33
<i>Non-HIV Fertility</i>	34
<i>HIV+ Fertility</i>	34
<i>Male Migration and Female Migration</i>	35
<i>FemPRO</i>	35
<i>FemSTD</i>	35
<i>FemRSK</i>	35
<i>FemNOT</i>	36
<i>MalePRO</i>	36
<i>MaleSTD</i>	36
<i>MaleRSK</i>	36
<i>MaleNOT</i>	36
<i>MaleOLD</i>	36
<i>FemOLD</i>	37
<i>Young</i>	37
<i>AIDS Age Profile</i>	37
<i>HIV Prevalence</i>	37
<i>ANC Age Prevalence</i>	37
<i>ANC Age Profile</i>	37
<i>Calibration</i>	38
<i>Cumulative Deaths</i>	38
<i>Deaths</i>	38
<i>Pyramid</i>	38
<i>Mortality</i>	38
<i>Reported deaths – Female and Male</i>	38
<i>Copyright</i>	38
APPENDIX C: DESCRIPTION OF WORKSHEETS IN PROVINCIAL VERSION WORKBOOK	39
<i>Initializer</i>	39
<i>EC – Common</i>	39
<i>EC – Asian</i>	39
<i>Other worksheets</i>	40
APPENDIX D: SUMMARY OF WORKSHEETS	41
APPENDIX E: ACRONYMS	43

1. Introduction

This guide begins with an overview of modelling of the HIV/AIDS epidemic in South Africa, which is presented in section 2. Section 3 provides information on the structure of the model. It comprises a brief description of the nature and basis of the assumptions, the location of different aspects of the model on the worksheets, and information about which assumptions and values can be changed by the user. Sections 4, 5 and 6 provide instructions on how to use the model. Section 4 describes how to do simple runs so as to get projections for different years. Section 5 describes the standard output, how to interpret it, and how to obtain additional output. Section 6 provides a brief overview of the *provincial* and *urban-rural* versions of the model. Finally, section 7 provides information for the more advanced user who wants to change parameters. This section includes a discussion of the calibration that is necessary when making such changes. As, structurally, the different versions of the model are all based on the *lite* version this manual describes, in the main, the *lite* version of the model. However, where appropriate, details of other versions of the model are discussed.

In addition there are several appendices. Appendix A gives details of the system requirements for running the model. Appendix B lists all the worksheets in the main workbook of the *full* and *lite* versions and explains where to find particular values. Appendix C lists all the worksheets in the workbook which paste the set-up assumptions into the *full* model to create the provincial models and explains where to find particular values. Appendix D is a summary list of all the worksheets, specifying the name and nature of the worksheet, whether it contains values that can be changed by users, and whether its contents change when running the projection engine. Appendix E lists all the acronyms used in the text of this user guide.

While reading the guide, it is useful to have the workbook of the model at hand as there are repeated references to different parts and features of the workbook in the text.

Although some actuarial or demographic background will be helpful in understanding the intricacies of the model's construction, the model is designed to be useful to actuaries and non-actuaries alike. Users who have no need to make changes to the model and its assumptions, for example those who only want to use the output, may want to skip the detail in sections 3 and 5. Users who want to change the assumptions underlying the model must, however, read through this detail as their changes may have unintended effects.

The guide assumes a basic proficiency in using Microsoft Excel. In particular, it assumes familiarity with naming conventions for rows and columns, understanding of concepts such as cells, range names, and macros, and an understanding of the difference between cells containing ordinary data and those containing formulae.

2. Overview of AIDS modelling

Modelling of the AIDS epidemic in South Africa by actuaries began with the so-called Doyle or Metropolitan Life model, which was developed in 1989. The model was based on a population hypothetically divided into four groups that differed in terms of the relative ease with which individuals belonging to each group were expected to contract and transmit the HIV.

The code for the Metropolitan model is proprietary. The Actuarial Society of South Africa (ASSA) felt that it was desirable for people to have access to a non-proprietary programme which users could alter to suit their needs. In 1996, ASSA therefore released the ASSA500 model. This was very similar in structure to the Metropolitan model with some simplifications to ease programming and comprehension and to shorten run times. The model was primarily designed to make users aware of the likely impacts of the epidemic on mortality and morbidity.

In 1998, the AIDS Committee of ASSA decided to develop the model further. There were several reasons for this:

- The ASSA500 model represented the epidemic in the black African population, rather than the population as a whole;
- There were concerns about the accuracy of the preliminary results of the 1996 census and there was a need for national estimates that attempted to correct for suspected deficiencies;
- Many South African demographers were continuing to ignore the impact of AIDS in their projections of the South African population;
- The ASSA500 model had inherited a number of demographic shortcomings from the Metropolitan model, particularly the assumptions of constant fertility, non-HIV mortality over time and the assumption of no international migration.

The result was an Excel 95 workbook called ASSA600, released to the public in early 1999. The model was designed to be appropriate for use as a national population model for the Pattern II (heterosexual) HIV epidemic found in South Africa. The base model contained a scenario that reflected its builders' best estimates of values for the model parameters and was calibrated to fit the antenatal data up to 1997. The naming convention was also changed to allow the user to modify the parameter values, for example for sensitivity analysis and scenario planning. The idea was that alternative version of the model could then be saved as ASSA601, 602, etc.

In 2000, the AIDS Committee felt that a further revision of the model was necessary. The update was needed because of increased knowledge about the epidemic, the availability of new data against which to calibrate the model, and greater awareness of the uses to which the model was being put. It was also decided to change the naming convention to reflect the year of the latest antenatal data used to calibrate the model.

The resultant ASSA2000 model incorporated the following adjustments reflecting new or updated information about the epidemic:

- 1988-2000 antenatal clinic (ANC) summary results;
- 1998 South Africa Demographic and Health Survey (SADHS) data, in particular, data on prevalence of STDs and condom usage;
- improved estimates of the population; and
- mortality data on the pattern and level of deaths that suggested, in particular, that non-HIV mortality for adults has not improved over time as expected.

In addition, the model was altered to:

- improve the fit to ANC survey data;
- allow for the possibility of making separate male and female assumptions;
- model the population groups separately;

- limit the trend in mortality and fertility rates over time;
- limit future in-migration;
- change the HIV survival curve to be a function of a Weibull distribution;
- allow for a bimodal distribution of paediatric HIV survival; and
- disaggregate the ‘contagion matrix’ (used in ASSA600) into more measurable and controllable parameters of heterosexual behaviour. These include the probability that a partner comes from a particular risk group, the number of new partners per annum, the number of sexual contacts per partner, the age of the partner and the probability a condom is used.

The date in the Model name refers to the most recent antenatal and mortality data used in the calibration of the model. Since the Department of Health embargo the detailed data needed to calibrate the model for six months after they release their report, and the report on the surveys have been released as late as October the model is released invariably some time after the year to which the data refer.

The ASSA2008 version of the model is the most recent version of the model to be released. The structure of this model is similar to that of ASSA2003. The following are the most significant changes that have been made:

- Interventions
 - PMTCT takes into account slower pace of rollout, and lower uptake of single-dose nevirapine. Roll-out is now in terms of percentage of pregnant women tested and percentage of women on NVP who also receive ART. Also changed the modelling of the impact of interventions on vertical transmission.
 - Separate ART roll-out rates for men, women and children, and in terms of percentage of new AIDS sick who start treatment. Also allowing for greater reduction in viral load on ART (from 1.76 to 2.8 unit reduction in log of viral load and higher rates of retention on ART).
- Changed the way condom usage modelled (and IEC rates of rollout and the factor by which the odds of condom usage increases with 100% rollout)
- Allows for separate HIV survival for adult males and adult females. The survival of untreated adults is now assumed to follow an Exponential distribution rather than a Weibull distribution which leads to a longer mean survival time but with greater variance. Survival of untreated children is assumed to be longer, especially for children infected at or before birth. Because of this the model now allows for the survival of some infected children to adulthood.

Like the ASSA2000 model the ASSA2008 model has been produced as a suite of several versions. The *lite* version, like previous *lite* versions and the ASSA600 model before them, treats the population of the country as one population group. The *full* version models each of the four population groups (Asian/Indian, black African, Coloured and White) separately at a national level, and aggregates to produce results for the population as a whole. Notionally the *provincial* version is the result of the aggregation of the application of the *full* version of the model separately to each of the provinces, although this aggregation is left to the user to do if desired. It would thus allow for geographic differences in the spread of the epidemic.

This user guide is intended for use with all four models. The differences between the *lite* and *full* versions will be noted in the text at relevant places. The approaches in the *provincial* and *urban-rural* versions are described in section 6.

This user’s guide is intended for use with the ASSA2008 model, rather than earlier models, although most parts of the guide are also applicable to the earlier versions of the model. As the

course of the epidemic progresses and more information about it becomes available, the model structure and base scenario will be further updated and future versions of the model will be released.

Any feedback on the model in the form of comments and criticisms would be appreciated and can be sent to aids.actuarialsociety.org.za.

Disclaimer

The model is distributed as a flexible tool to allow researchers to make their own predictions and projections about the HIV/AIDS epidemic. No level of accuracy is implied, nor can the Actuarial Society of South Africa accept any responsibility for the way in which individuals use the model or the results they obtain from it. The model is offered free via the Internet as a public service to anyone who has a use for it.

The ASSA2008 model as disseminated has been calibrated to reproduce the patterns of past antenatal clinic survey data and the number of adult deaths. As such, the model represents the triangulation of data from the population census, antenatal survey and registered deaths by some of the country's top actuaries, demographers and epidemiologists. It is not recommended that users alter the assumptions in the model in any way unless they have a very good reason for doing so. If any of the assumptions are altered in any way, the user must ensure that the model is recalibrated to ensure that it remains consistent with the recorded experience to date. Users who have any questions in this regard can consult with the ASSA AIDS Committee (aids.actuarialsociety.org.za).

Other sections of this guide note where the user can change particular parameters on the worksheets to reflect a change in assumptions. The following points should be observed when making such changes.

In common with previous models one of the features of the ASSA2008 models is the large degree of interdependence of different parameters and assumptions. A change in one will often necessitate a change in others. There are two broad categories of second-level changes.

- In some cases the first change of value will result in another change 'automatically', in that other values are dependent, through a formula of the worksheet or a macro procedure, on the changed value. This happens, for example, where proportions must sum to 100%. In these cases, the user does not need to take any action. However, users should note that the 'automatic' changes will only take effect when the user presses the F9 (CALC) key or runs a projection. The automatic CALC function defaults to OFF in the ASSA2008 model to speed up the projection process.
- In other cases the changed parameter will require a manual change to other parameters and assumptions so as to achieve a fit with the observed values of the ANC surveys, both overall and by age, and the national mortality rates by age. This process of manual changing of parameters to counterbalance previous changes is what we refer to as 'calibration'.

Users must be aware of the nature of the information they are changing. Some of the parameters in the ASSA2008 workbook reflect assumptions or observed data such as numbers in the population. These are entered as ordinary numbers on the worksheet. Many other parameters are based on formulae that draw on values in other cells or cell ranges in the workbook. **The contents of these cells must not be replaced by numbers.** As with ordinary Excel usage, the user can see in the status bar whether a particular cell is the result of a formula or range name reference. In some cases, however, a cell will appear to contain an 'ordinary' number that does not involve a reference or formula but will, in fact, be a record of the previous year's numbers used to project the current year's numbers. This is the case, for example, with all the tables labelled as 'before'. In changing

values on the worksheets, users must be aware of the very different implications of changing a cell containing a simple value and changing a cell containing a formula or reference or a number generated by previous projections. Guidance as to which values can be changed is given by the colour of the figures. By and large calculated values are in black, while assumptions that affect projection – and could thus potentially be changed – are in red. Data such as antenatal survey prevalence figures, which do not impact on the computations of the model, are in blue.

In the past, incorrect results have been attributed to the Actuarial Society of South Africa or the models in public documents and in the press. In order to prevent this from happening in future, we ask that all users adhere to the following guidelines.

- If the results have been generated using the models without any alternation, the user should reference them as ‘results extracted from the ASSA2003 (*lite*, if *lite* version used) AIDS and Demographic model of the Actuarial Society of South Africa as downloaded [date] from [site address]’.
- If the model has been adjusted and recalibrated, the user must, in addition to the full reference to the model, explain exactly how and why it was adjusted. The user must also make it clear that the resulting estimates are not those produced by the Actuarial Society of South Africa. This must be done in such a way that anyone reading the report understands clearly that the user’s results do not represent the views of the Society. Ideally, it should also be done in such a way that another user can replicate the changes and check the projections.

3. The structure of the model

The ASSA model projects year-by-year changes in an initial population profile over a period of years chosen by the user. It does so on the basis of a number of demographic, epidemiological and behavioural assumptions. This section of the guide provides a brief description of the model, its key parameters and assumptions. Any user who wants to change any of the parameters must read this section so as to understand the impact of any proposed changes.

The model projects on a year-by-year basis, with each year's projections reflecting changes between 1 July of one calendar year and 30 June of the following calendar year. For the sake of simplicity, each projection year is referred to by a single calendar year rather than by both of the calendar years. The 'stock' numbers for each year reflect the position as at the middle of the respective calendar year, while the 'flow' numbers reflect the change from the middle of that calendar year to the middle of the following calendar year.

3.1 Division Into subgroups

The model splits the population by sex and also into three distinct age groupings: young (up to age 13), adult (14-59) and old (60 and above). The *full* version of the model also splits the population by population group. The adult group is divided into four risk groups, which are differentiated by their level of exposure to the risk of contracting the HI virus through heterosexual activity. These risk groups are:

- PRO** Individuals whose level of sexual activity is such that it is similar to that of commercial sex workers and the level of condom usage and infection with STDs is similar to that of the STD group.
- STD** Individuals whose level of sexual activity is such that their HIV prevalence is similar to someone regularly infected with STDs.
- RSK** Individuals with a lower level of sexual activity, but who are still at risk from HIV in that they have, on average, one new partner per annum and sometimes engage in unprotected sex.
- NOT** Individuals who are not at risk of HIV infection.

By definition, someone from the RSK group will not have sex with someone from the PRO group. Further, those in the NOT group have sex only with others in that group or, if they have sex with individuals from other groups, always take effective precautions.

The numbers in these risk groups are determined initially according to the proportions appearing on the *Assumptions* sheet. These proportions are applied equally to all ages over 24. Before age 25, individuals remain in the NOT group until they become sexually experienced. The assumptions have been determined, where possible, on the basis of empirical evidence. Where this was not possible, either educated guesses were made or the assumptions were determined so that modelled results fit observed data such as the antenatal prevalence figures for past years. The latter method of determining values is part of the 'calibration' process discussed later.

The age and risk classifications divide the population into the following groups, with each group's calculation done on a separate worksheet within the workbook. In the *full* version of the model, there is a set of these worksheets for each population group. In the aggregate of the provinces, there

is a workbook for each province and a set of worksheets for each population group within each province:

- Young:** All individuals aged 0 to 13. The only infections assumed are those arising at birth or from breastfeeding. On their 14th birthday, individuals are allocated to the risk groups according to the assumed proportions on the *Assumptions* sheet. The model assumes that a proportion of those born to HIV+ mothers are HIV+ at birth. Further, the model assumes that a further proportion of non-HIV babies contract the virus from their HIV+ mothers through breast-feeding. The Young age group is shaded yellow on the worksheets.
- FemPRO:** Female members of the PRO risk group up to age 59, subdivided by duration since infection.
- FemSTD:** Female members of the STD risk group up to age 59, subdivided by duration since infection
- FemRSK:** Female members of the RSK risk group up to age 59, subdivided by duration since infection
- FemNOT:** Female members of the NOT group up to age 59
- FemOLD:** On their 60th birthdays all individuals are allocated to the OLD class. The duration since infection classification still applies, but no further infections or fertility occur beyond this age. The OLD worksheets are a run-off of the population. No one is assumed to survive beyond age 90. The OLD group is shaded in grey on the worksheets.
- Male***:** The same structure as the Fem*** worksheets but with no births.

The total population is allocated between male and female and over the age range according to the distributions given in the *Population* worksheet. While it is possible to measure, to some extent, the size of the STD group and, to a lesser extent, the PRO group, the RSK and hence NOT groups are hypothetical constructs whose size is set to reproduce past patterns of prevalence through the calibration process.

3.2 Process of infection

3.2.1 Introduction

Diagram 1 displays how individuals move from state to state under the action of the model. Certain transitions are assumed to be impossible e.g. moving from HIV+ to HIV-; and becoming infected after age 60. The model allows for the inclusion of migrants in all risk groups and at all ages. Migrants are assumed to have the same duration since infection profile and prevalence rate as non-migrants of the equivalent risk group. It should be noted that in the diagram, the category, HIV+Births includes children who were infected perinatally and those that were infected by breast milk.

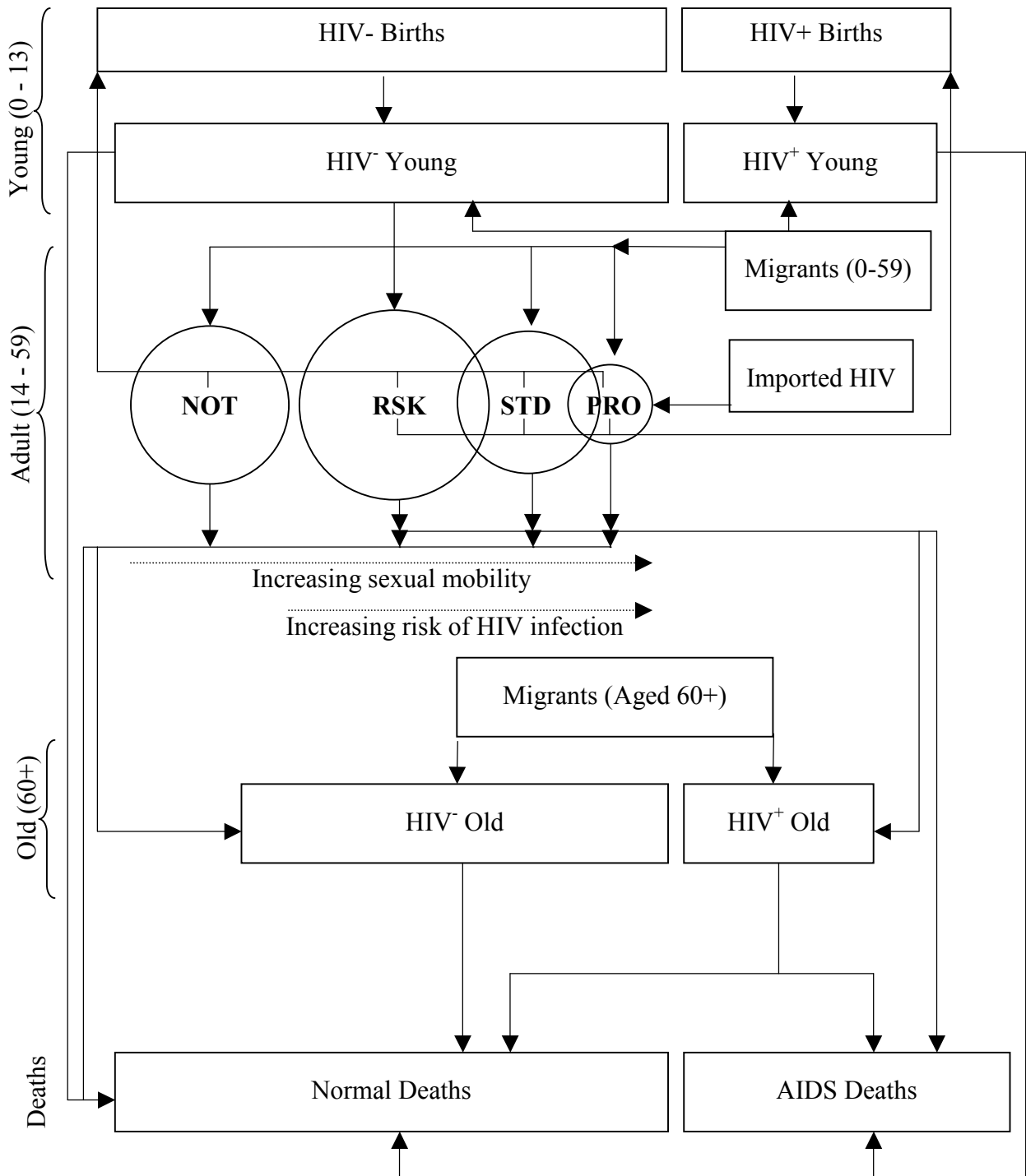


Diagram 1: A schematic diagram of the lite model

The heterosexual interaction and hence spreading of the virus is modelled taking into account the following, given the person is from a particular risk group: *the chance that the partner is from a particular risk group, the chance that the partner is in a particular stage of disease, the number of new partners per year, the number of contacts per partner and the probability of transmission if no condom used (given the risk group of the partner), and the probability that a condom was used.* The number of new partners per year and the number of contacts per partner for females at a particular age are a function of a 'sexual activity' curve. This curve was chosen such that the

pattern of HIV infections of pregnant women assumed by the model to be attending antenatal clinics is more or less the same as the results of the ANC surveys.

3.2.2 Starting the epidemic

Infection is introduced into the PRO risk group via the number of male and female infected ‘imports’ in the *lite* model (cells C27 and D27 on the *Assumptions* sheet). Smaller numbers are used for each of the population groups in the *full* version of the model to allow for both the smaller size of the populations as well as possible lags in the start of the epidemic. These imports are not added to the population, but rather used to create HIV prevalence of partners in the initial years and hence start the epidemic. The number of imported HIV need not be a whole number or even greater than one. This feature allows the model to cater for situations where the starting date of the epidemic is before or after 1985, by simply changing the size of this number.

3.2.3 Infection assumptions

The epidemic spreads through the population at risk by assumed infection of non-infected individuals within and between groups. The rate of spread of the infection is controlled by assumptions about two key factors, namely the *amount of sexual activity*, and a range of factors determining the *probability of infection*.

The distribution of female sexual activity by age is represented in the model by the sexual activity curve, which is found on the *SexActivity* sheet. The curve involves an assumption about the relative sexual activity by age for females. The curve is negatively-skewed bell-shaped with the following form:

$$S(x) = \frac{(x-c)e^{-b(x-a)^2}}{c}$$

where a (the position factor) is reflects, in part, the average age of first sexual intercourse
 b (the shape factor) is set, in part, to reproduce the shape of prevalence by age from the antenatal results as well as the age distribution of AIDS cases and AIDS deaths
 c (the scale factor) is set so that the average of $S(x)$ is 1.

The activity of males is a function of that of females and the age of their partners.

By varying the shape of the curve, the distribution of the new HIV infections by age is adjusted (separately for different population groups and province in the two other models). The user can manipulate both a (position) and b (shape) factors. The shape and position factors are recorded in column B of the *SexActivity* sheet.

The default values for a and b for the female curve have been set to reproduce the age distribution of ANC HIV prevalence age distributions over time. The reasonableness of the male curve can be checked against the age distribution of reported male AIDS cases circa 1995. When changing the shape of the curves and the age distribution of partners of females, the user should ensure that the consistency between the number of deaths and number of AIDS cases still holds. This can be checked by looking at the *ANC Age Profile* and *AIDS Age Profile* worksheets and ensuring that the dotted curves (actual figures) are reasonably close to the solid ones (modelled figures).

ASSA2008 models sexual behaviour, and thus the probability of infection, on the basis of a combination of several components, as follows:

- a matrix showing the probability that a male partner is from a particular risk group;

- a matrix of male-to-female transmission probabilities per sexual contact for various combinations of risk group encounters;
- a matrix of female-to-male transmission probabilities per sexual contact for various combinations of risk group encounters;
- a matrix showing the number of new partners per year and the number of contacts per new partner per year;
- a matrix showing the probability that a female partner is from a particular risk group;
- a matrix of condom usage for each risk group by age;
- the effectiveness of condoms; and
- a matrix of relative frequencies of sex, relative odds of condom use and relative levels of HIV infectiousness, in different stages of HIV disease (these stages are defined in sections 3.4.2 below).

These components are all recorded on the *Assumptions* worksheet.

All the above information is then brought together in the following formula:

The probability of someone in risk group i , aged x becoming infected in a year is

$$1 - \left\{ 1 - a(x) \left(1 - \sum_{j=1}^4 w_{ij} \sum_{y=14}^{59} h(y|x) \sum_{t=1}^6 p_{ij}(y) [1 - T_{ij}(y)]^{n_{ij} \sqrt{s(y)D_t}} - \left(1 - \sum_{j=1}^4 w_{ij} \sum_{y=14}^{59} p_j(y) h(y|x) \right) \right) \right\}^{m_i \sqrt{s(x)}}$$

where

$a(x)$ is the multiple by which the probability of HIV infection per partnership increases in women aged x (for women over 25 and for men, the value of this parameter is set at 1)

w_{ij} is the proportion of the individual's partners that are in risk group j

$p_j(y)$ is the proportion of the individual's y -year old partners, in risk group j , that are HIV positive

$p_{ij}(y)$ is the proportion of the individual's y -year old partners, in risk group j , that are HIV positive and in stage t of disease

t is the stage of disease (stages 1 to 4 correspond to the four stages of the WHO Clinical Staging System, stage 5 comprises individuals on anti-retroviral treatment, and stage 6 comprises individuals who have discontinued anti-retroviral treatment)

$h(y|x)$ is the proportion of the individual's partners that are aged y

n_{ij} is the number of sexual contacts the individual is likely to have per partner in risk group j

$s(x)$ is an index of the level of sex activity at age x

m_i is the number of sexual partners the individual has per year

D_t is the factor by which the amount of sex is reduced in stage t of disease

$T_{ij}(y)$ is the probability that an HIV positive y -year old, in stage t of disease and in risk group j , transmits the virus to a partner in risk group i , during a single act of sexual intercourse.

The factor $T_{ij}(y)$ can be expanded as follows:

$$T_{ij}(y) = r_{ij} \cdot I_t [1 - (1 - [1 - f_j(y)] R_t) e]$$

where

r_{ij} is the probability that the individual will be infected if they engage in a single act of unprotected sex with an individual in risk group j

$f_j(y)$ is the probability that the partner uses a condom

e is the effectiveness of condoms in preventing HIV transmission

I_t is the factor by which the risk of transmission (per act of unprotected sex) is increased in stage t of disease

R_t is the factor by which the proportion of sex acts that are unprotected is reduced in stage t of disease.

(All parameters, with the exception of e , $f_j(y)$, I_t , D_t and R_t , are gender-specific. Many of the parameters also change value over time.)

The combination of components allows the model to be used to test the impact of interventions that attempt to change one or more of these variables.

3.2.4 Paediatric infections

Twenty percent of babies born to infected mothers are assumed to be infected at birth. A further 16% of those born to infected mothers not infected at birth are assumed to be infected via breast feeding. The median time of survival of those infected perinatally is 4.74 years and that infected via breastfeeding 14.59 years in 1985, and increases slightly after that. These assumptions were chosen to be consistent both with the distribution of paediatric deaths by age assumed by the UNAIDS/WHO Reference Group on Estimates, Modelling and Projections, and empirical evidence on the proportion of children of infected mothers who get infected.

The user can modify the proportion of births assumed to be infected and the proportion assumed to be infected by mother's milk on the *Assumptions* worksheet (cells X24 and X25). The relevant table is fourth from the top on the left of the worksheet.

3.3 Starting population

The starting population reflects the actual population as at 1 July 1985. This was derived by a process of reconstruction linking estimates of the population in 1970 to those of the census population in 1996, ensuring consistency with estimates of fertility and mortality rates derived independently and between the numbers of males and females in various age groups.

The current provinces did not exist in 1985. For the *provincial* version of the model, it was therefore necessary to reconstruct the base population that could be expected to have been within these boundaries in that year. This was done by taking into account a remapping of the 1991 census into the new boundaries and the patterns of inter-provincial migration between 1985 and 1996.

3.4 Mortality

The mortality data is found in the *MortTable* worksheet. The initial rates of mortality apply on for the year centred on 1 January 1986, to be consistent with a starting population six months earlier at 1 July the previous year.

3.4.1 Non-HIV mortality

The non-HIV probability of death and probability of becoming infected are used in a multiple decrement life table that applies to individuals not infected by HIV.

The model uses a table of estimated mortality rates at each age for each of the years 1985 to 2007. After 2007, mortality rates are projected to trend logistically to ultimate rates at a rate determined by a ‘mortality improvement factor’ using the following formula:

$$(a - b) \times (c^{(CurrYear - 2007)})$$

where

a = the mortality rate in 2007

b = the ultimate mortality rate

c = the mortality improvement factor.

This formula is contained in the lookup formula which can be found in the table of current year non-AIDS mortality in the *MortTable* worksheet.

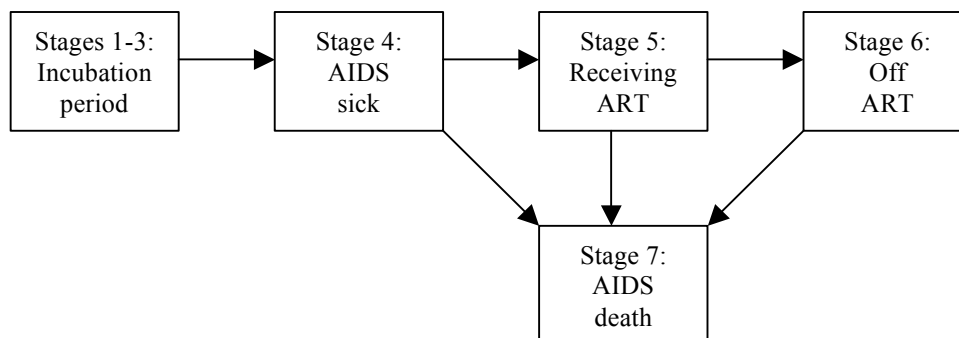
The user can alter the tables of mortality rates for the years to 2007, the ultimate rates of mortality, the mortality improvement factor, and the formulae for interpolating future rates of mortality for all ages on the *MortTable* worksheet. The relevant formulae are found in the table headed ‘Current Year=X Non-AIDS Mortality Rates’, and the ultimate mortality rates and mortality improvement factor. In the *full* version of the model, these changes need to be made to the population group specific *MortTable* sheets.

3.4.2 Survival of adults with HIV

In the absence of anti-retroviral treatment, adults are assumed to progress through four stages of disease before dying from AIDS. These four stages correspond to those defined in the WHO Clinical Staging System. The effects of anti-retroviral treatment (ART) are modelled by introducing a further two stages, which represent people receiving treatment and people who have started treatment but subsequently discontinued treatment. Descriptions of these six disease states are given in the table below. The first two stages are largely asymptomatic. Symptoms occur more frequently in stage 3, and include weight loss and oral infections. People in stage 4 experience a range of more severe conditions, such as pneumonia, extrapulmonary TB and wasting syndrome. These conditions are referred to collectively as AIDS.

HIV stage	Description
1	WHO stage 1: Acute HIV infection
2	WHO stage 2: Early disease
3	WHO stage 3: Late disease
4	WHO stage 4: AIDS
5	Receiving anti-retroviral treatment
6	Discontinued anti-retroviral treatment

The possible transitions between these stages are shown in the figure below. In the absence of treatment, individuals are assumed to progress through each of the four WHO stages sequentially, before dying from AIDS. Individuals who initiate ART are assumed to do so at the time that they experience their first AIDS-defining illness, and move into stage 5 on initiating ART. People may die from AIDS while receiving ART or may discontinue treatment before dying from AIDS.



The time spent in each of the first four stages and the sixth stage is assumed to follow an exponential distribution. The median and shape parameters for the Weibull distributions are specified in the top left table in the *Male* and *Female Adult Survival* worksheets (since the Weibull distributions each have a shape parameter of 1, they are each effectively exponential distributions). The median parameters are assumed to vary according to the age at infection, and the parameters are therefore set separately for three different age bands: 14 to 24, 25 to 34 and 35 or older. For stage 5, rates of transition into stage 6 and rates of AIDS mortality are determined by the “ART Assumptions” and “ART mortality adjustment factors” tables on the *Interventions* worksheet. The “ART Assumptions” table gives annual probabilities of death and transition to stage 6. The “ART mortality adjustment factors” table adjusts the mortality rates faced in stage 5 that apply over each successive year of ART rollout.

On the basis of the above assumptions, the median time from infection to death (in the absence of non-AIDS mortality) is calculated in the *Male* and *Female Adult Survival* sheets. Ignoring the effects of non-AIDS mortality, in 2010, the median survival is roughly 13.81 years for adults (male and female) infected when under the age of 25, 12.11 years for adults infected when between the ages of 25 and 34, and 10.42 years for people infected when over the age of 34. These calculated medians are shown in the *Assumptions* sheet, in the table fourth from the top on the left hand side of the sheet. They should not be changed by the user as they are calculated from values found elsewhere in the workbook.

3.4.3 Survival of children with HIV

The modelling of the survival of children infected with HIV is similar to the modelling of survival of adults. The key difference is that instead of using a four-stage system to classify individuals prior to ART initiation, we use only two stages: pre-AIDS and AIDS. It is assumed that the time spent in the AIDS stage as well as the time spent in the ‘ART discontinued’ stage, follow an exponential distribution. The median and shape parameters for each of these stages are specified in the top left table of the *Paediatric Survival* sheet. It is assumed that children infected perinatally (i.e. infected at birth) experience more rapid disease progression than children who are infected through mother’s milk. Assumed median parameters are therefore specified separately for these two modes of transmission.

The rates of transition from pre-AIDS to AIDS, the rates of AIDS mortality while on ART and the rates at which ART is discontinued are assumed to be constant, except during the first six months on ART. These rates of transition can be altered in columns E and F of the *Paediatric Survival* worksheet (in the second and third tables on the left-hand side).

On the basis of the above assumptions, the median time from infection to death (in the absence of non-AIDS mortality) is calculated in the *Paediatric Survival* sheet. Ignoring the effects of non-

AIDS mortality, in 2010, the median survival is roughly 7.29 years for children infected perinatally and 14.65 years for children infected through breastmilk.

3.5 Fertility

3.5.1 Non-HIV

The parameters relating to fertility of those not infected are found on the *Non-HIV Fertility* worksheet. Overall age-specific fertility rates are determined in a similar way to the mortality rates, with a table of estimated age-specific fertility rates for the period to 2006, after which rates are determined by interpolating between the rates in 2006 and the ultimate rates. The first table in this worksheet provides non-HIV fertility rates for each of the four risk groups for the current year by taking into account the relative fertility of women in each age group and the proportion of women in the various risk groups at that age. The relative fertility factors can be found in the second table from the top left of the *Assumptions* worksheet. The results are shown in columns B, C, D and E in the *non-HIV fertility* sheet. The relative fertility factors for the PRO, STD and RSK groups can be changed on the *Assumptions* sheet. The factor for the NOT group should not be changed as it is derived from the fertility rates for the other three groups.

The model assumes that PROs have a lower fertility rate than STDs, who have a lower rate than RSKs, who, in turn, have a lower rate than NOTs. Although this may seem counter-intuitive, the argument for the assumption is that, in order to maintain a highly sexually active life-style, PROs would probably use contraception or abort fetuses. In addition, there is evidence suggesting that STDs may lead to lower fertility. On the other hand, if awareness of contraception is high, then individuals choosing to have children are more likely to be in stable relationships and therefore at reduced risk of contracting HIV. It is unlikely that the relative fertility rate assumptions have a great impact on the results, with the possible exception of the number of infants born HIV-positive.

3.5.2 HIV and fertility

These parameters are found in the *HIV+ Fertility* worksheet.

The model allows for the impact of the duration of infection on fertility by multiplying the non-HIV fertility rates by a factor determined as follows:

$$\text{adjustment factor} = (a - b) \times c^d$$

where

a = factor allowing for the bias, particularly at the younger ages, arising from the fact that those falling pregnant are those having sex and not using condoms

b = factor allowing for an initial impact of the virus on fertility

c = factor allowing for the impact of the virus on fertility over time

d = duration of infection in years

The first three factors can be changed in the columns entitled 'Start Ratio', 'Initial Impact' and 'Reduction Factor' towards the right of the worksheet (columns AH to AJ).

3.5.3 Overall

The births resulting from this fertility are split into males and females according to the assumed proportion of births that are male. This proportion is found on the *Assumptions* worksheet. The relevant table is fifth from the top, on the left side of the worksheet. The births populate the age

zero cells in the next step of the projection via the *Young* sheet. The proportion of births that are boys can be changed by the user but will have minimal effects on the workings of the model.

3.6 Migration

The migration-related variables are found on the *Male Migration* and *Female Migration* worksheets. For ease of computation, the model assumes that all migration takes place at the end of the relevant projection year. 'Migrant' can refer to both immigrants and emigrants. The figures in the *Male Migration* and *Female Migration* worksheets represent net immigration (i.e. in-migrants less out-migrants).

The starting population was constructed to include all migrants received up to 1985.

Migrants are apportioned to the four risk groups according to the proportions in the *Assumptions* sheet. These are found in the table third from the top on the left side of the worksheet. Currently these proportions are set to be the same as those of the receiving population but the values can be changed by the user.

Migration after 2000 is assumed to fall from its 2000 level, logistically, towards close to zero over a 30-year period.

3.7 Population groups

The ASSA2003 *lite* model does not distinguish between the different population groups defined during the apartheid era, namely African, Coloured, Indian and White. The *full* version of the model allows for separate modelling of the epidemic in the four population groups (sheets identified by 'Black', 'Coloured', 'Asian' and 'White'). This feature of the model was developed in response to user demand. This demand was motivated, in part, by the observation that the impact of the epidemic is very different in the different groups. These differences constitute one of the reasons for differences between the prevalence in the Western and Northern Cape on the one hand and the other provinces on the other. Further, some users wanted to extrapolate the results of the model to socio-economic groups, and geographic sub-regions, and population group is thought to be a useful proxy for these in South Africa. There were thus both demographic and economic reasons for modelling the epidemic in terms of population groups.

Developing this aspect required a number of demographic assumptions. The disaggregation presents significant challenges during calibration of the model. Unfortunately, very limited data exist about the impact of the epidemic in the different population groups. Mortality data have not been published on a population group basis since 1990, and the ANC data have not been published in disaggregated form since 1995. The mortality data have, since 1998, again been collected on the basis of population group but have not yet been published. The ANC survey continues to collect information on population group. This information is not publicly available, but ASSA's AIDS Committee has been given access to data for 1997 to 2001.

The model has been fitted to these data and to data from private sector company testing and insurance testing by taking into account differing levels of STDs and condom usage.

3.8 Interventions and behaviour change

The model allows for the effect of interventions (prevention and treatment programmes) on sexual risk behaviours, probabilities of HIV transmission and HIV survival. Currently, five interventions are modelled:

- improved treatment for sexually transmitted diseases (STDs);
- information and education campaigns (IEC) and social marketing;
- voluntary counseling and testing (VCT);
- mother-to-child transmission prevention (MTCTP); and
- anti-retroviral treatment (ART).

The user can choose which of these interventions are introduced in the projection by entering ‘Yes’ or ‘No’ for each intervention in the top left table in the *Interventions* worksheet. In the table to the right of this, the user can specify the year in which each intervention is assumed to be introduced, and the rates at which each intervention is assumed to be phased in. These rates can be thought of as the proportions of the population that have access to the prevention or treatment programme considered in each year after the initial introduction of the programme. In the default scenario, it is assumed that all five interventions are introduced, but at different times and at different rates of phase-in.

The *Interventions* sheet also contains assumptions about the effect of each intervention. The table below summarises the effects of each intervention in terms of the key epidemiological parameters in the model. It is assumed that all individuals participating in an MTCTP or ART programme would receive counselling and testing prior to joining the programme (the extent of this can be set by the user). The behavioural changes that occur under VCT are therefore assumed also to occur under MTCTP and ART scenarios. Improved STD treatment lowers the probability of HIV transmission, because other STDs enhance the risk of HIV transmission when present in either the HIV-negative or HIV-positive partner. Anti-retroviral treatment also lowers the probability of HIV transmission, because this treatment lowers the concentration of HIV in the body, and hence renders recipients less infectious.

Intervention	Condom usage	Frequency of sex	Probability of sexual transmission	Probability of mother-to-child transmission	Survival with HIV
IEC and social marketing	✓				
Improved STD treatment			✓		
VCT	✓	✓			
MTCTP	✓	✓		✓	
Anti-retroviral treatment	✓	✓	✓		✓

The table second from the top shows adjustment factors that apply to the rates of mortality shown in the “ART Assumptions” table (lower down on the Interventions worksheet) over each year of ART rollout.

Information and education campaigns are assumed to be the main reason for the dramatic increases in condom usage that have been observed in South Africa over the last decade. It is currently assumed that with the maximum possible roll-out of these campaigns, a 35-fold increase in the odds of condom usage from 1985 levels would occur. This assumption can be changed in the table third from the top in the *Interventions* worksheet.

Substantial improvements in the treatment of STDs have occurred since 1994, particularly with the introduction of syndromic management protocols in the public health sector. Since STDs occur more frequently in the higher-risk groups, the effect of the improved treatment is assumed to be greater in the high-risk groups than in the low-risk groups. These assumptions can be changed in the table fourth from the top in the *Interventions* sheet.

The assumed effects of VCT are specified separately for people who are not at risk of infection, people who are at risk of HIV infection, and people in each stage of HIV disease. Individuals who are at risk of infection are generally more likely to be tested than those who are not, and behaviour change following VCT tends to be more significant when people test positive – particularly in the later stages of disease. The user can specify the rates at which untested individuals receive testing when there is 100% access to VCT services, and the reductions in unsafe sex for individuals who receive VCT. In addition, the model allows for individuals who test negative to be retested at a later stage, by moving a proportion back into the ‘untested’ population each year. There is also allowance for a ‘wearing off’ of the benefits of VCT over time, as individuals gradually forget what they have learnt through the VCT programme and revert to their former sexual practices. All of these assumptions are specified in the table fifth from the top of the *Interventions* worksheet.

The rate at which VCT is accessed is increased if an ART or MTCTP programme is introduced. Individuals participating in these programmes would be required to undergo counselling and testing, as explained previously. In addition, allowance can be made for greater utilisation of VCT services among individuals who are not yet eligible for ART (because they are HIV-negative or still in the early stages of disease), if ART is available. In the default scenario, however, we assume that there is no such increase.

Mother-to-child transmission prevention is assumed to reduce both the proportion of children infected perinatally and the proportion infected by breastmilk. These assumptions, together with the assumed proportions of women agreeing to participate in the MTCTP programmes, are specified in the table sixth from the top of the *Interventions* worksheet.

The most significant effect of anti-retroviral treatment is its effect on HIV survival. The reduction in AIDS mortality is modelled by introducing two further disease states, as described in sections 3.4.2 and 3.4.3. A further benefit of ART is that it reduces the frequency of AIDS-defining illnesses in adults and children by approximately 75%. The model also allows for the reduction in viral load that results from ART, and the associated decline in HIV infectiousness for people on ART. The only potentially negative impact of ART is that people may practise less safe sex if they perceive HIV/AIDS to be less of a threat when they have access to ART. The model therefore allows for a degree of reversal of the benefits from the social marketing programme when ART is available. (In the default scenario, however, it is assumed that there is no adverse behaviour change as a result of ART.) All of these assumptions regarding the effects of ART are specified in the table seventh from the top in the *Interventions* worksheet. The rates of mortality and discontinuance while on ART are shown in the table. The rates of transmission at the other stages are shown in the top left tables in the *Male and Female Adult Survival* worksheets and *Paediatric Survival* worksheets.

4. Running the model

4.1 Starting the epidemic

The model has a start date of 1 July 1985. At this date, each risk group is assumed to be free of HIV. An initial level of prevalence is then introduced by setting the number of ‘imported HIV’ on the *Assumptions* sheet to start the epidemic. HIV levels and patterns in later years are generated by the assumptions built into the model.

To run the model, open the appropriate ASSA2003 workbook. When prompted, click on the ‘Enable Macros’ button. If you click on ‘Disable Macros’, the workbook will open, but you will not be able to run any projections.

The copyright notice will appear automatically after you click ‘Enable Macros’. Click on the ‘Accept’ button (provided you accept the copyright conditions). If you are using the *lite* version, the workbook will open with the *Assumptions* worksheet on the screen. If you are using the *full* version, the workbook will open with the *Assumptions – All* worksheet on the screen. Clicking on the ‘Decline’ button will close the workbook.

Before running the model, click the ‘Reset Projection to Zero’ button on the *Assumptions* or *Assumptions – All* worksheet. This runs the ‘ResetProj’ macro, which allocates the initial population into age bands and the four risk groups according to the proportions on the *Population* worksheet and the *Assumptions* worksheet.

4.2 Projecting

The model allows the user to project in two different ways. The first is by making use of buttons which run the model for periods of one, five, ten or forty years at a time. Clicking more than one button one after the other produces a projection for the sum of the specified years.

For example, to estimate the situation as at July 1998, 13 years after the start date, perform the following actions after having reset the model as described above:

- Click the ‘Project 10 years’ button on the *Assumptions* worksheet;
- Click the ‘Project One year’ on the *Assumptions* worksheet three times, allowing time for the programme to run each year between clicks.

After each click, the bottom left-hand corner of the screen will report on the year-by-year progress of the projection. After the four clicks described above, the sheets, in particular the *Results* and *Population* worksheets, will reflect the picture 13 years after the start date of July 1985.

An alternative method of obtaining the same output is to specify the end year. To do this, perform the following actions after having reset the model:

- Type in the end year required in the cell to the right of the ‘To specific year’ button on the *Assumptions* worksheet and <ENTER>;
- Click the ‘To specific year’ button.

Again, the bottom left-hand corner of the screen will record the progress of the projection.

For simple runs, it makes little difference which method of running the projection is chosen. For some users, however, one way may serve the purpose better than others. For example, a user who wants to change a particular assumption as from a certain date, will want to use the first method. Similarly, a user who wants to record results at particular points, for example, every five years, should use the first approach.

By default, the model projects forward from July 1985. Some users will be more interested in projections into the future from the current year. The model will project the epidemic to July of the current year if you click the 'To current date' button on the *Assumptions* worksheet. The current year for the projection is derived from the system date on the computer on which the projection is performed. As before, progress is recorded at the bottom left-hand corner of the screen. The end point of this projection can then be used for projections into the future.

To run repeated projections from the current date, or from any other specified date, the user should perform the following actions:

- Run the projection to the desired starting date by one of the methods described above;
- Click the 'Store Population' button to store the results for the last year of the projection;
- Click the 'Reset Projection from Store' button to establish the position in the last year of the previous projection as the new starting point;
- Run the new projections by one of the methods described above.

As is the case with all projections, the further into the future one projects, the less reliability one can place on the results. This is especially true when projecting the impact of an epidemic about which there is still much uncertainty. For this reason it is not recommended to use the ASSA model to project beyond 40 years. However, some users will want longer projections, for example, to estimate cohort life expectancies. Such users will find that the program tends to become very large and slow on longer projections irrespective of which of the above methods is adopted to run the model.

The user can circumvent the problem by storing the population after 40 years, resetting the base situation to what has been stored, and projecting forward from there, as follows:

- Click the 'Store Population' button after projecting for 40 years.
- Click on the 'Reset Projection from Store' button to get ready for further projections.

These steps need to be repeated if the projection is longer than 80 years. Unless the user is only interested in the end point, he or she should save the output they are interested in (say mortality rates over time) before resetting.

5. Understanding and adapting outputs

5.1 Standard outputs

Key statistics for each year of the projection can be placed in the *Results* worksheet to record their passage over time. Currently the recorded statistics include:

- numbers of people as at 1 July in each sex/risk group/HIV classification and in total;
- number of deaths in each year (July-June) due to AIDS or other reasons;
- HIV prevalence rates for different groups;
- fertility and birth rates;
- male and female mortality rates by age;
- crude extra mortality i.e. additional mortality resulting from AIDS;
- numbers of people accessing prevention and treatment programmes; and
- numbers of adults and children in different stages of HIV disease.

The *Results* worksheet contains results for each year of the projection.

Further population results, such as numbers alive and dead by HIV status for each age, can be found in the *Population* worksheet. The *Population* worksheet contains information only on the start and end dates of the projection. Where several projections are performed one after the other, the start date is the original start date i.e. July 1985 unless the 'Store Population' and 'Reset Population from Store' functions have been used.

To view the *Results* or *Population* worksheets, simply click on the appropriate tabs at the bottom of the screen.

Other worksheets contain charts of some of the generated information.

- The *AIDS Age Profile* worksheet shows the distribution of AIDS cases by age;
- The *HIV prevalence* worksheet shows prevalence by risk group and sex for each year from the first year of the projection to the final year;
- The *ANC Age Prevalence* worksheet shows the trends in antenatal prevalence levels observed since 1991, by five-year age group, compared with the prevalence levels projected by the model for each five-year age group.
- The *ANC Age Profile* worksheet shows the projected prevalence of HIV in women attending public antenatal clinics, by five-year age groups for the projected year, as well as the observed prevalence levels for 1995, 1999, 2000 and 2001.
- The *Cumulative Deaths* worksheet contains a chart of the cumulative number of people who have died from AIDS in each year from the start of the projection to the end;
- The (non-cumulative) *Deaths* worksheets contains a chart of the number of people becoming infected with HIV, dying from AIDS and dying from other causes in each year from the start of the projection to the end;
- The *Pyramid* worksheet contains a chart of the population pyramid for each sex in the last year of projection in five-year age groups;
- The *Mortality* worksheet contains a chart of the projected mortality curves for each sex in the last year of projection with and without AIDS ('No AIDS');
- The *Reported deaths – Female/Male* worksheets compare the number of total adult deaths by age for females and males separately with those estimated to have actually occurred in South Africa in selected years.

The user can obtain a full-screen picture of any of these charts (except the reported deaths) by clicking the appropriate button on the *Assumptions* worksheet after running a projection. The chart can be printed in the normal way, by choosing 'File' and then 'Print' through the Excel menu

system. To copy the chart to another document, for example in MSWord, the user can use the 'Edit' menu and 'Copy' function of Excel. By then using the 'Paste Special' and 'Picture' options in the MSWord, the chart will take up less disk space and memory.

To return to the *Assumptions* worksheet from any of the charts, click the 'Return' button at the bottom right of the worksheet.

5.2 Adding output statistics

Although the results of the projection are recorded in each of the worksheets, in most worksheets this is only done for one year at a time. The *Results* worksheet allows the user to record desired output over time.

The user can erase items not of interest from the *Results* worksheet. In doing so, however, the user should note that some items are used to produce the charts described above, so if these are needed, then those specific rows must not be erased. The user can also add statistics that have not been included on the standard *Results* worksheet. References or formulae for desired statistics should be inserted in the column headed 'LATEST' (column B). However, the user must ensure that all rows are inserted above the line marked 'NB: Any additions must be inserted above this line!'

In the simple case, the user can add statistics computed directly from data already included in the model. In the more complicated case, the user can add data from other sources and calculate additional output data from this and existing data.

A very simple example would be adding as output the population in the age group used for calculations of the economically active population in each year. In South Africa, the age group 15-65 is usually used for this purpose. To include the output on the *Population* worksheet, the user would perform the following actions:

- Insert a row in the *Results* worksheet anywhere above line marked 'NB: Any additions must be inserted above this line!';
- Label the row in the column headed 'STATISTIC' (column A), for example give it the label '15-65';
- Insert in the 'LATEST' column the formula for calculating this number in a year, i.e. '=SUM(Population!L19:L67)'

When the projection is run, the number of economically active people in each year of the projection will be recorded in the *Results* sheet.

An alternative way of achieving the same result is to perform the calculation on the *Population* worksheet, and then refer to the calculated cell on the *Results* spreadsheet. The detailed steps for doing this are as follows:

- Insert the formula to calculate the population aged 15-65 years on the *Population* worksheet at the bottom of the 'Total Population' column. For this example, we assume this is done in cell C97;
- On the *Results* worksheet, insert and label a new row '15-65' as described above;
- On the *Results* worksheet, in the 'LATEST' column, type the reference '=Population!C97'.

When inserting the above and other formulae, remember that the cells will not necessarily be updated until after you run a projection, as ASSA2003 sets the default of the CALCulation function from Automatic to Manual. To see the updated calculation without running a projection, press the F9 key on the keyboard.

Another simple example would be to calculate the death rate of male and female PROs. To do this, the user would perform the following actions:

- Insert four rows in the *Results* worksheet, for example in the 'Deaths' section;
- Label the rows in the column headed 'STATISTIC', with the labels 'AIDS Male PRO deaths', 'Normal Male PRO deaths', 'AIDS Female PRO deaths' and 'Normal Female PRO deaths';
- Insert in the 'LATEST' column of these rows the appropriate references for these statistics, namely 'MalePRO!AO51', 'MalePRO!AL51', 'FemPRO!AO51' and 'FemPRO!AL51';
- Insert two rows in the *Results* worksheet, for example in the 'Fertility and Growth' section;
- Label the rows in the column headed 'STATISTIC', with the labels 'Male PRO Death Rate' and 'Female PRO Death Rate'
- Insert in the 'LATEST' column of these rows the appropriate formulae for calculating the death rates, namely $=(B39+B40)/(B10+B11)$ AND $=(B41+B24)/(B12+B13)$;

When the projection is run, the death rates for male and female PROs in each year of the projection will be recorded in the *Results* sheet.

A more complicated example would be adding as output the number of workdays lost to HIV/AIDS-related disease. To obtain this output, the user would need to perform the following actions:

- On the *HIVTable* worksheet, create a table setting out the estimated number of workdays lost per annum. This could be indexed by any or all of age, sex, stage of disease, duration since infection and risk group. These data would have to be obtained from elsewhere.
- On the *Results* worksheet, insert a new row, for example above the mortality tables.
- In the new row, in the column labelled 'LATEST', type in the formula that will calculate the new statistic based on the table and the population worksheet.

The new statistics will be recorded on the *Results* sheet for each step of the projection. Charts using any of the new statistics can also be produced.

6 The provincial and urban-rural versions of the model

6.1 The provincial version

The *provincial* version of the ASSA2003 model consists of the application of the *full* ASSA2003 model to each of the provinces. By projecting and calibrating each province separately, the *provincial* model has the potential to allow for differences between the provinces in respect of particular characteristics, such as different demographic composition of the population, different levels of STDs, different prevention and treatment programmes and different starting times for the epidemic.

The *provincial* version of the ASSA model utilises two workbooks – the workbook for the *full* version of the model, together with a workbook that contains the initial assumptions relating to each province. The user creates the workbook for a particular provincial model by first loading the *full* model and the AssumptionsProv workbook and then running a macro on the AssumptionsProv workbook which pastes in the province-specific starting assumptions into the *full* version of the model. The *full* version can then be operated in the usual manner, as described above, to give results pertaining to a particular province and can be save as the *provincial* version of the model.

The workbook of provincial assumptions consists of an *Initializer* worksheet where the user specifies the province, destination file, and whether the model allows for a change scenario. There are a further nine *Prov – Common* worksheets, and 36 *Prov – Population Group* worksheets containing initial assumptions which are pasted into different worksheets of the *full* version worksheet once the user has specified the province and asked that the paste operation take place. The cells of the workbook consist of a mix of absolute values and formulae. Some of the latter appear to give invalid or incorrect values when viewed in the AssumptionsProv workbook. However, once the formulae are pasted into the *full* version worksheet, the cells referred to in the formula take effect and the values become valid.

6.2 The urban-rural version

The *urban-rural* version allows for the case where available data on HIV prevalence consists of inconsistent and irregular samples surveys from urban and rural clinics. The user can input data from all the sites, specify each site as either urban or rural, and indicate how the data should be aggregated to produce average urban, rural and national figures.

Calibrating the *urban-rural* model is somewhat more difficult than calibrating the *full* version of the model. Not only does one have to calibrate all three levels (urban, rural and national) but in many situations one has no data for setting the size of the risk groups (STD in particular) and a number of the other parameters and these have to be set by trial and error to produce the best fit.

7 Changing assumptions and parameters in the model

As far as possible, the parameters of the ASSA2003 model have been set by reference to studies of empirical evidence. This was possible, for example, in respect of the size of the STD group, the probability of transmitting the virus, age of the partners, and condom usage. Where it was not possible, parameters are set within bounds of reasonableness to produce output comparable with observations of antenatal seroprevalence levels and estimates of the actual number of deaths based on the registered deaths.

Calibration involves adjustments of parameters that have not been estimated independently so that the ‘results’ of the model more or less match observed reality. In particular, the model results should match the results of the annual ANC surveys both in terms of overall level and by age. The model results should also match the number of adult deaths estimated on the basis of those recorded by the Department of Home Affairs on the population register after adjusting for an estimate of under-recording.

Calibration can be done visually, using the charts on the *Calibration*, *ANC Age Profile* and *Reported deaths* worksheets.

The *Calibration* worksheet looks at the fit with ANC survey data. In judging the fit, it must be borne in mind that the ANC points before 1997 are probably biased upwards, perhaps as a result of an urban bias in the location of data collection sites over a period when prevalence levels were higher in the urban areas. The ‘target’ points are therefore the ones that should be used when calibrating the model. Further, the 1998 ANC survey value is incorrect as the Department of Health made an error in weighting the provincial results to get the national estimate. The correct value should be some 2% lower than the published value.

The user can obtain a full-screen calibration graph of the antenatal clinic prevalence each year by clicking the ‘Calibration data’ button on the *Assumptions* worksheet after running a projection. The aim is to get the solid curves to follow the points marked by the data as closely as possible. These points correspond to HIV prevalence among:

- ANC attendees;
- Commercial sex workers (CSWs); and
- STD clinics;

in respect of each:

- actual surveyed prevalence; and
- ‘target’ expected prevalence.

The ‘target’ prevalence is based on the patterns found in other African countries where the epidemic has been experienced for longer than in South Africa.

The variables for HIV prevalence among women aged 15-49 are found on the *HIVTable* worksheet.

The procedure used in the initial calibration of the model was first to fit the prevalence figures for the PRO group. This group is roughly equivalent to CSWs but not identical in that some CSWs will be at less risk than PROs as defined in the model because, for example, they use condoms and have STDs treated. The next step after fitting the PRO group, was to fit the STDs and finally the ANC attendees. It was assumed that most sexual activity occurs between partners in the same risk group. Further, by definition, members of the RSK group do not have sexual contact with members of the PRO group, and members of the NOT group do not have unprotected sexual contact with members of any of the other groups.

Based on evidence that suggests that women only visit a clinic late in their second trimester, the projection intervals run from the middle of one year to the middle of the next year. This assumes that women are, on average, six months pregnant when they first visit the clinic.

In order to model the prevalence of women attending public antenatal clinics rather than that of all pregnant women, the ASSA2003 model increases the age-specific rates of all pregnant women in the population by a multiple to allow for the fact that the prevalence of women attending public antenatal clinics is likely to be higher than that of pregnant women in general.

Adopting this approach, it is impossible to get anything but a crude fit to the PRO and STD target points. Modelled prevalence plateaus at a far higher level than the observed and 'target' points suggest. One possible explanation for this mismatch is that the survey and target CSW prevalences may include people from the (lower risk) RSK and STD risk groups, while the model definition of the PRO group is narrower, including only people at higher risk.

The *ANC Age Profile* and *ANC Age Prevalence* sheets both compare the prevalence by age from the antenatal survey with that generated by the model. The *ANC Age Profile* sheet can be accessed by clicking the 'ANC Age profile' button on the *Assumptions* sheet. Calibration should aim to ensure as close a fit to the age profile of the antenatal prevalence data as possible by changing such factors as the female curve of sex activity and the age distribution of partners.

The *Reported deaths* worksheets contain charts which look at the fit with deaths recorded by the Department of Home Affairs. There are two such worksheets, one for male deaths and one for females. The user can view these worksheets by selecting the appropriate tabs on the workbook after running a projection. A full screen view is obtained by choosing the 'Full screen' option on Excel's View menu.

The charts show the estimated number of deaths in the country based on the registered deaths by age for each of the years 1996 (calendar), 1997/98, 1998/99, 1999/2000 and 2000/01, as well as deaths projected by the model for the last year of the projection. The aim in calibration is to match as closely as possible the estimates of the number of deaths based on the registered deaths.

7.3 Calibration tips

Theoretically the user can change a wide range of parameters on the worksheet. In practice, however, there are only a limited number of combinations of parameters for which changes could provide a potentially realistic picture of the AIDS epidemic in South Africa. Further, because of the interdependence of different parts of the model, it is sometimes not easy to predict the effects of a particular change. This sub-section therefore provides some tips as to type and extent of changes in output that are likely to result from changing particular parameters.

The following pointers suggest the effect of changing some of the key parameters on the outcome:

- The size of the NOT group determines to a large extent the level at which the antenatal prevalence is expected to plateau in the long run.
- Provided it remains fairly small (around 1% of the population), the PRO group does not impact much on the long-term course of the epidemic.
- The STD group is conceptually defined to be those infected with STDs to such an extent that the average transmission probability is expected to be that recorded on the *Assumptions* sheet. In constructing the model, the proportions were set to be consistent with those recorded in the 1998 Demographic and Health Survey, on the assumption that people in the PRO and STD groups experience STD symptoms about 25% of the time.

- The age profile of prevalence is determined largely by the sex activity, age of partner and age distribution of condom usage assumptions.

Appendices

Appendix A: System requirements

Pentium III 500Mhz (or better from a speed perspective)

Windows 98 or later

Microsoft Excel 98 or later

64MB Ram

Appendix B: Description of worksheets

Assumptions

This is the worksheet that appears on the screen when the user opens the model. At the centre it contains a number of buttons that allow the user to run projections, jump to worksheets containing output graphs and calibration checks, and reset the model.

- The 'Project One Year', '5', '10' and '40' buttons allow the user to run the model for that number of years cumulatively, in other words, in addition to the point to which the model has already been run.
- The 'To Date' button runs the model to the current year as determined by the clock on the machine running the model.
- The user can specify the year to which he or she wants to project, by entering the year and clicking on the 'To specific year' button.
- The 'Reset Projection to zero' button returns the workbook to the start year.
- The 'Store Population' and 'Reset Projection from Store' allow the user to run the projection to a point in time, store the population details at this point, and then to use this as a new starting point for projections, resetting from the stored population each time.

Around the central buttons, the worksheet contains a number of tables and matrices containing key parameters.

Several of the matrices reflect assumptions underlying the modelling of sexual behaviour. These include:

- a matrix showing the probability that a woman's partner is from a particular risk group. The relevant table is the topmost table second from the left¹;
- matrices of male-to-female and female-to-male transmission probabilities per sexual contact for various combinations of risk group encounters. The relevant tables are the topmost tables third and fourth from the left;
- a matrix showing the number of new partners per year. The table is the topmost table fifth from the left;
- a matrix showing the number of contacts per male partner from different risk groups of females of the different risk groups, as well as the averages. The table is the rightmost table at the top of the worksheet;
- a matrix of relative frequencies of sex, relative odds of condom use and relative levels of HIV infectiousness, in different stages of HIV disease. This table is the rightmost table, second from the top;
- a matrix specifying the increases in female susceptibility to HIV infection at young ages. This table is the rightmost table, third from the top;
- a matrix of condom usage for each risk group by age. This is the long thin table in the middle of the worksheet; and
- the effectiveness of condoms. This is a single value shown on the worksheet just above the condom usage table.

Other tables reflect further basic assumptions, as follows:

- a table, at top left, showing the total of the starting population;

¹ Note that all of the table locations described here are those for the *lite* version of the model. In the *full* version, some of the tables have slightly different locations.

- a table, second down from the top left, showing the percentage distribution of the male and female adult populations across the four risk groups, as well as relative fertility of females in each risk group;
- a table, third down from the top left, showing the percentage distribution of the male and female immigrant adult population across the four risk groups;
- a table, fourth down from the top left, showing the calculated median times to death, and imported infection for male and female;
- a table, fifth down from the top left, showing proportions of births to HIV-positive mothers that are infected perinatally and by breastfeeding, the proportion of all births that are male, and the correlation between age at sexual debut and riskiness of sexual behaviour.

The matrices described above show the assumptions that apply in the *current* projection year, and are not necessarily the same as the assumptions that apply in the first projection year (when there are assumed not to be any interventions in place). The assumptions about (a) mother-to-child transmission probabilities, (b) HIV transmission probabilities and (c) frequency of condom use are specified for the ‘no intervention’ scenario in the tables in the bottom right-hand corner of the worksheet. The rates of condom usage and probabilities of transmission in the *current* year are calculated from these assumptions and the assumptions in the *Interventions* worksheet.

The *full* version of the model contains a separate *Assumptions* worksheet for each population group, as well as an *Assumptions – All* worksheet. In respect of this and other worksheets, the sheets are labelled with the standard name followed by the name of the population group, namely ‘Asian’, ‘Black’, ‘Coloured’ and ‘White’. The group-specific *Assumptions* worksheets contain all the tables described above in respect of that particular population group.

The *Assumptions – All* worksheet contains all the buttons to drive the model. These include the buttons contained in the Assumptions worksheet on the *lite* version, as well as a number of additional buttons which allow the user to access group-specific information. When the model is first opened, only the aggregate worksheets are seen. These worksheets contain the summary results from the four component population groups. The additional buttons on the *Assumptions – All* worksheet are as follows:

- The ‘Change Assumptions for Asian Group’ reveals all the worksheets specific to the Asian group and places the *Assumptions – Asian* worksheet on the screen. The corresponding Black, Coloured and White buttons reveal all the worksheets for that specific group. Pressing any of these buttons hides all worksheets relating to population groups other than the chosen one.
- The ‘Show All Population Groups’ button reveals all worksheets for all population groups.
- The ‘Hide All Population Groups’ button hides all worksheets except the aggregate ones.

Results

This worksheet is used to store output over time. It thus summarises the results for each year from the start to end date of a projection. From top to bottom it records information on the clear and infected populations, aids sick, deaths, HIV prevalence rates, incidence rates, mortality statistics, demographic indicators, fertility and growth, advocacy output, male and female mortality by age, force of infection, numbers participating in prevention and treatment programmes, and numbers of adults and children in each stage of HIV disease. None of this information is used in the running of the model although some of it is used to produce the graphs. Much of it could be removed if the user had no use for it. However, there is no need to remove or replace any of it in order for the user to produce his or her own output. To add their own output, users can insert rows above the line marked ‘NB: Any additions must be inserted above this line!’ and enter the required extraction formula in column B.

The *full* version of the model contains a separate *Results* worksheet for each population group in addition to the one which is used for aggregate results.

Interventions

Assumptions about the timing and effect of HIV/AIDS prevention and treatment programmes are specified in this worksheet. The top left table lists the five intervention programmes, and the user can enter 'Yes' or 'No' next to each one, depending on whether the intervention is to be introduced over the projection term. The table to the right contains assumptions about the years in which the interventions are introduced and the rates at which they are phased in. In the six tables below, assumptions are entered about the epidemiological and behavioural effects of each of the five interventions: social marketing, improved treatment for sexually transmitted diseases, voluntary counselling and testing, mother-to-child transmission prevention and anti-retroviral treatment.

In the tables at the bottom of the sheet, we model changes in sexual behaviour and knowledge of HIV status due to voluntary counselling and testing. These are the changes over the *previous* projection year. The changes are calculated separately for males and females.

The *full* version of the model contains a separate *Interventions* sheet for each population group, but no aggregate *Interventions* sheet for the population as a whole.

Population

This worksheet summarises population statistics for the start and end date of a projection. Where several projections are done one after the other, the start date is the original start date (unless the 'Store' and 'Reset for Store' functions are used). Working from left to right on the worksheet, we have:

- Details of starting population by age and sex;
- Infected and non-infected population by age and sex as at 1 July of the current year;
- Deaths by age and sex projected for the twelve months beginning 1 July of the current year. In this table, age 'B' represents deaths of babies born during the projected year. Deaths marked '0' represent deaths of those who were under 1 year of age at the start of the projected year;
- Numbers used to construct the population pyramid as at 1 July of the current year;
- A summary table of male and female deaths in the twelve months beginning 1 July of the current year in five-year age groups;

Below these tables, we find:

- A table showing numbers and proportions of new adult infections;
- A table summarising numbers and proportions of AIDS deaths in five year age groups.

The *full* version of the model contains a separate *Population* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

MortTable

This sheet produces the non-HIV mortality rates for each year, by looking up the rates up to 2001 and interpolating between the 2001 rates and ultimate rates for years after 2001. Working from left to right, the worksheet reflects:

- 1-(Non-AIDS survival) (i.e. non-AIDS mortality), by one-year age group for age last birthday and sex, for the current year;
- Mortality improvement factors, by one-year age group and sex;
- Non-AIDS mortality rates by sex and one year age group for exact ages, for the current year;
- Ultimate non-AIDS mortality rates for one-year age group and sex;

- Mortality to date in respect of non-AIDS mortality rates by one-year age group and year, with separate tables for male and female;
- Projected vs recorded deaths for five-year age groups, for each of the years 1996 to 2002, and for male and female separately. These are used to produce the graphs for calibrating the model against the recorded deaths.

The *full* version of the model contains a separate *MortTable* worksheet for each population group, as well as an aggregate worksheet for the population as a whole. The group-specific *MortTable* worksheets do not contain a table of projected vs recorded deaths as described above. The aggregate *MortTable* worksheet contains a table of weights by population group and age for ages 0 to 12 years which is used to compute the aggregate mortality rates.

SexActivity

This sheet calculates the relative amount of sex at each age for males and females. The distribution of male sex activity by age is derived from the distribution for females and the age mix of the partners of the women. The leftmost table on this sheet contains separate indices for male and female sex activity in one-year age groups. (The values for male sexual activity at ages 14 to 59 do not sum exactly to 100% because it is assumed that there is some male sexual activity after age 59.) The next table from the left (above the graph) contains estimates of the proportions of the population aged less than 25 who are sexually active (calculated from the tables below). The third table from the left contains $f(x)$, representing the percentage of total female sexual activity occurring at age x . The fourth table contains the parameters used to generate $f(y|x)$, the percentage of partners of females aged x who are aged y . The next table is the table of $f(y|x)$ values. The table below that contains the values for $g(y)$, the percentage of total male sexual activity occurring at age y . Below that is $g(x|y)$, the percentage of partners of males aged y who are aged x . The table in the bottom left-hand corner of the sheet is used to calculate the rates at which youth become sexually active at each age, and hence the rates at which males and females move from the NOT to PRO, STD and RSK groups. The table also specifies the increase in susceptibility to HIV infection in young females. The *SexActivity* worksheet includes a chart that displays the male and female sexual activity curves, and one, on the far right, that displays the age distribution of male partners of women at selected ages.

The *full* version of the model contains a separate *SexActivity* worksheet for each population group, but no aggregate SexActivity worksheet for the population as a whole.

Male HIVTable and Female HIVTable

These sheets record the survival curves (i.e. the proportion of people surviving by duration since becoming infected) for adults, those born with the virus and those becoming infected through mother's milk. Reading from left to right, the worksheets contain the following tables:

- HIV survival rates by broad age group (14-24, 25-34 and 35 plus) of adults, and duration of infection in one-year intervals;
- Survival of HIV births by duration of infection in one-year intervals;
- Survival of babies infected by mother's milk by duration of infection in one-year intervals.

The columns marked 'Surviving' refer to the proportion of infected people still alive t years after infection.

From the proportions surviving, the model works out death rates, which are recorded in the top set of tables alongside the relevant survival rates. The model also interpolates from these rates to produce a further table of mortality rates by age of the individual and duration of infection in one-year intervals for those aged 14 years and above. This table is found below the chart.

The survival curves are displayed in a chart below the topmost set of tables.

The *full* version of the model contains separate *Male* and *Female HIVTable* worksheets for each population group.

HIVTable

This sheet is used to produce estimates of the prevalence of women attending antenatal clinics by five-year age group and in total. It also contains a table of observed values against which the estimates can be compared for calibration purposes. The top table of this worksheet contains data in five-year age groups. Two columns for pregnant women show the numbers who are HIV positive and negative. A series of columns represents the prevalence of women attending ANC clinics. Finally, there is an adjustment factor which adjusts the prevalence in the population as a whole to allow for the fact that it includes women who attend private clinics, in order to arrive at an estimate of the prevalence of those who attend public antenatal clinics.

The second set of tables when moving from top to bottom contains three sets of data – for ANC, STD clinics and commercial sex workers. The ANC columns record survey and target data. The columns for STD clinics and CSWs record South African and target data for each year. To the right of these tables are two further tables, in which the HIV prevalence rates in each five-year age band, from each antenatal survey, are recorded. The first table represents the raw data, and the second table represents the data after adjusting for the urban bias in the early years of the antenatal surveys.

The *full* version of the model contains separate *HIVTable* worksheets for each population group, as well as an aggregate worksheet for the population as a whole. While most of the other aggregate worksheets are constructed through simple summation of the values in the group-specific worksheets, the values in the aggregate *HIVTable* worksheet have two differences. First the estimated prevalence levels of those attending antenatal clinics in the top table are derived as a weighted average of the same figures in each of the population groups. The weight used is the proportion of pregnant women from each group attending public antenatal clinics. Secondly, the second table down contains the national prevalence figures as published or estimated elsewhere and is not derived from the population-group specific sheets.

Male and Female Adult Survival

These sheets contain the assumptions about the distribution of time male and female adults spend in each stage of HIV disease. The Weibull median and shape parameters for each stage are entered in the top left table. Since all the shape parameters have been set to 1, all waiting times are exponentially distributed. The median parameters are specified separately according to the adult's age at infection: under 25, 25 to 34, or over the age of 34. In the table immediately to the right, the overall median survival (from infection to death) for each of these three age bands is shown.

In the second, third and fourth tables from the top, on the left-hand side of the sheet, the rates of transition out of each stage are specified, according to the duration of the time spent in each stage. These three tables are for the three age bands described above. In the first row of each table, we specify the proportion of people moving out of each stage within the first six months of entering the stage. In the second row, we specify the proportion of people who leave the stage within the next 12 months (as a proportion of the people who remain in the stage for at least 6 months). Similarly in the third row, we specify the proportion of people who leave the stage within the following 12 months (as a proportion of the people who remain in the stage for at least 18 months). Subsequent rows are specified similarly.

In the tables to the right of these three tables, we calculate the proportions of individuals in each stage of disease according to duration of infection, as well as the AIDS mortality rates and survival rates according to duration of infection.

In the top right corner of the worksheet, there are two tables. The first is used to store certain outputs from the previous projection year. These stored outputs are used to calculate the numbers of individuals making transitions between different disease stages over the course of the previous year, and are also used in calculating changes due to voluntary counselling and testing in the previous year. The structure of this table is different in the *Female Adult Survival* worksheet than in the *Male Adult Survival* worksheet. In the *Female Adult Survival* sheet, outputs relating to births are stored in the table.

The calculations of the numbers of transitions between the different disease stages are performed in the second table, to the right.

Below and to the right, there are tables that are used to calculate the proportions of individuals in HIV stages 4, 5 and 6 at the *start* of the year. These calculations are not done for stages 1 to 3, as the rates of progression through the first three stages of disease are assumed not to be affected by the absence or presence of anti-retroviral treatment programmes, and hence do not change from year to year.

The *full* version of the model contains separate *Male* and *Female Adult Survival* sheets for each population group, but no aggregate *Adult Survival* sheet for the population as a whole.

Male and Female Staging

These worksheets calculate the proportion of HIV-positive males and females in each stage of disease, according to current age and current duration of HIV infection. There are six tables – one for each of the six HIV stages. Proportions of individuals in each stage of disease are calculated for each individual age, rather than for each of the three age bands referred to in the *Adult Survival* sheet. These calculations are performed on the assumption that the specified median and shape parameters for the <25, 25 – 34 and >34 age bands apply to people aged 19, 29 and 39 respectively at the time of infection. For people infected at other ages, the proportions in different disease stages are determined through linear interpolation and extrapolation of the proportions determined for 19-, 29- and 39-year olds.

The *full* version of the model contains separate *Male* and *Female Staging* sheets for each population group, but no aggregate *Staging* sheet for the population as a whole.

Paediatric Survival

This sheet is similar in structure to the *Adult Survival* worksheet. The important differences are that (a) there are four disease stages instead of six, and (b) calculations are performed separately for children infected perinatally and children infected through breastmilk.

This sheet contains the assumptions about the distribution of times children spend in each stage of HIV disease. The Weibull median and shape parameters for the AIDS stage are entered in the top left table. Since the shape parameters are each set at 1, the time spent in the AIDS stage is exponentially distributed. The median parameters are specified separately according to the child's mode of infection: perinatal transmission or transmission through breastmilk. In the table

immediately to the right, the overall median survival (from infection to death) for each mode of infection is shown.

In the second and third tables from the top, on the left-hand side of the sheet, the rates of transition out of each stage are specified, according to the duration of the time spent in each stage. These two tables are for the two modes of transmission described above. In the tables to the right of these two tables, we calculate the proportions of children in each stage of disease, according to duration of infection, as well as the AIDS mortality rates and survival rates according to duration of infection. To the right of these are tables that are used to calculate the proportions of children in HIV stages 4, 5 and 6 at the *start* of the year. These calculations are not done for the pre-AIDS stage, as the rate of progression through the pre-AIDS stage of disease is assumed not to be affected by the absence or presence of anti-retroviral treatment programmes, and hence does not change from year to year.

In the top right corner of the worksheet, there are two tables. The first is used to store numbers of children in each disease stage from the previous projection year. These stored outputs are used in the second table, to the right, to calculate the numbers of children making transitions between different disease stages over the course of the previous year.

The *full* version of the model contains separate *Paediatric Survival* sheets for each population group, but no aggregate *Paediatric Survival* sheet for the population as a whole.

Non-HIV Fertility

This sheet produces age specific fertility rates for HIV negative women in each of the risk groups. Reading from left to right, this contains the following tables:

- Non-HIV fertility rates by the four risk groups and age of the woman in one-year age-groups, from age 14 to 59, for the current year;
- Observed fertility rates by year for the period 1985 to 2006 by one-year age-groups, from age 15 to 49 for all HIV negative women irrespective of risk group. The last column of this table contains Non-HIV fertility rates by one-year age-group;
- A fertility improvement factor and ultimate rates by one-year age groups, from age 15 to 49;
- The ratio of HIV+ fertility to non-HIV fertility by one-year age groups, from age 15 to 49, for the current year (produced by the *HIV+ Fertility* sheet);
- Percentage change in fertility attributable to HIV;
- The proportion of women in each risk group in one-year age groups, from age 15 to 49.

The *full* version of the model contains a separate *Non-HIV Fertility* worksheet for each population group. There is a single aggregate *Fertility* worksheet which combines information from the group-specific *Non-HIV Fertility* and *HIV+ Fertility* worksheets. The *Fertility* worksheet contains the following tables:

- Population group weights for female weights of each age from age 14 to 49 years;
- Percentage change in fertility attributable to HIV. This table, among others, records HIV prevalence and aggregates the HIV+ and HIV- fertility rates from the group-specific worksheets for each age from 14 to 49 years;
- Fertility rates for all women by one-year age-groups, from age 14 to 59.

HIV+ Fertility

This sheet calculates the adjustment ratio for converting the fertility rates of HIV negative women to fertility rates for women infected with the virus. The worksheet contains a single table. The rows record the age of the individual, from 14 to 59 years. A first set of columns records the duration of infection, from one to 29 years. The third last column records the start ratio – the ratio of HIV positive to HIV negative ruling immediately before becoming infected. The second last column

records the initial impact on the fertility ratio of becoming infected. The final column records the reduction factor, which is equal to 1 minus the rate at which fertility drops per year infected.

The *full* version of the model contains a separate *HIV+ Fertility* worksheet for each population group.

Male Migration and Female Migration

These sheets contain the net numbers of migrants into the population at each age for each year. Each of these worksheets contains a single table. The rows of the table reflect the age of the individual in one-year age groups. The columns of the table reflect the year, from 1985 to 2025. The model assumes that beyond 2025 the number of migrants remains the same as those in 2025.

The *full* version of the model contains separate *Male* and *Female Migration* worksheets for each population group, as well as aggregate *Male* and *Female* worksheets for the population as a whole.

FemPRO

This sheet calculates the numbers of females in the PRO risk group at each duration since becoming infected, the number of new infections and the number of infected and uninfected births at each age. The worksheet has a BEFORE table at the top, with an AFTER table below it. The BEFORE position consists of data reflecting the population at the start of the year. The first set of columns contain, for each year of age and each year of duration of being infected, the number of people in that group. The columns to the right of this series contain a range of other parameters that affect the change in position over the projection period. These columns also include calculations of the numbers of people in each stage of HIV disease. The AFTER table again provides a value for each year of age and each year of duration. The AFTER position is calculated from the BEFORE position during projections using the formulas in the cells of the AFTER table to provide the BEFORE table for the next year. The final columns of the AFTER table provide for calculation of necessary adjustments to account for migration.

Below these tables is a STORE table. This is used for storing the population profile when the 'Store Population' button on the *Assumptions* sheet is clicked, so that one can revert to the stored population using the 'Reset Population from store' button.

The *full* version of the model contains a separate *FemPRO* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

FemSTD

As for *FemPRO*.

The *full* version of the model contains a separate *FemSTD* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

FemRSK

As for *FemPRO*.

The *full* version of the model contains a separate *FemRSK* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

FemNOT

This sheet is analogous to the above sheets but without the complication of having to allow for infected people. The worksheet has a BEFORE table at the top, with an AFTER table alongside. The BEFORE position consists of data reflecting the population at the start of the year. The columns contain, for each year of age, the starting position, the number of deaths during the year, and net immigration. The AFTER table provides a value for each year of age reflecting the position at the end of the year. The AFTER position is calculated from the BEFORE position during projections to provide the BEFORE table for the next year. The small NEW table below the AFTER table reflects the number of female YOUNG people joining the FemNOT category over the year.

The *full* version of the model contains a separate *FemNOT* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

MalePRO

As for *FemPRO*.

The *full* version of the model contains a separate *MalePRO* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

MaleSTD

As for *FemPRO*.

The *full* version of the model contains a separate *MaleSTD* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

MaleRSK

As for *FemPRO*.

The *full* version of the model contains a separate *MaleRSK* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

MaleNOT

As for *FemNOT*.

The *full* version of the model contains a separate *MaleNOT* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

MaleOLD

This sheet is analogous to the above sheets without sexual intercourse, i.e. there are no newly infected and no births. The worksheet has a BEFORE table at the top, with an AFTER table below it. The BEFORE position consists of data reflecting the population at the start of the year. The first set of columns contains, for each year of age and each year of duration, the number of people in that group. The columns to the right of this reflect total infected, total clear and infected, and number of AIDS and non-AIDS deaths. The AFTER table again provides a value for each year of age and each year of duration. The final columns of the AFTER table provide for calculation of necessary adjustments to account for migration. The AFTER position is calculated from the BEFORE position during projections using the formulas in the cells of the AFTER table to provide the BEFORE table for the next year.

The *full* version of the model contains a separate *MaleNOT* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

FemOLD

As for *MaleOLD*.

The *full* version of the model contains a separate *FemOLD* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

Young

This sheet calculates the number of survivors from birth at each age up to age 14. The worksheet has a BEFORE table at the top, with an AFTER table below it, and several calculation tables and a STORE table below that. The BEFORE position consists of data reflecting the population at the start of the year. The first set of columns contain, for each year of age and sex, the number of people in that group who are clear, who were born HIV+, and who have become infected through mother's milk. The table includes a row for babies born during the year. The columns to the right of this series reflect the number of female and male deaths predicted for each of the three HIV categories, and estimated female and male net migration. The AFTER table again provides a value for each year of age and sex in respect of each of the three HIV categories. The AFTER position is calculated from the BEFORE position during projections using the formulas in the cells of the AFTER table to provide the BEFORE table for the next year. The table to the right of the AFTER table calculates the numbers of HIV-positive children in each stage of disease. The tables between the AFTER and STORE table contain the formulae for calculation of the number of infected and non-infected babies born during the year, and surviving to the end of the year, as well as the number infected through mother's milk.

The *full* version of the model contains a separate *Young* worksheet for each population group, as well as an aggregate worksheet for the population as a whole.

AIDS Age Profile

The worksheet contains a chart showing the distribution of AIDS cases by age. The data refer to cases up until 1995, as data have not been collected since that date.

HIV Prevalence

The worksheet contains a chart showing HIV prevalence by risk group and sex for each year from the first year of the projection to the final year. The risk groups charted are female PRO, female STD, male STD, female POP and male POP. The POP lines represent the prevalence for the total female and male populations aged 15-49 years.

ANC Age Prevalence

This worksheet contains a chart showing the trends in antenatal prevalence levels observed since 1991, by five-year age group, compared with the prevalence levels projected by the model for each five-year age group.

ANC Age Profile

The worksheet contains a chart showing the projected prevalence of HIV in women attending antenatal clinics, by five-year age groups, for the most recent projection year. The chart also shows

the HIV prevalence levels observed in women attending public antenatal clinics for 1995, 1999, 2000 and 2001.

Calibration

The worksheet charts actual, 'target' and modelled prevalence among women attending ANC and STD clinics as well as modelled CSW and RSK prevalence and CSW 'target' prevalence from the first year of projection to the end year. The 'target' data reflect patterns found in other African countries that have more experience of the epidemic than South Africa.

Cumulative Deaths

The worksheet contains a chart showing the cumulative number of people who have died from AIDS in each year from the start of the projection to the last year of the projection.

Deaths

The worksheet contains a chart showing the numbers of people becoming infected with HIV, dying from AIDS, and dying from other causes in each year.

Pyramid

The worksheet contains a chart of the population pyramid for each sex in the last year of projection in five-year age groups.

Mortality

The worksheet contains a chart of the projected mortality curves for each sex in the last year of projection with and without AIDS. The label q_x on the Y-axis refers to the probability of a male or female individual aged exactly x dying within the next year of age.

Reported deaths – Female and Male

These worksheets chart reported deaths, after adjustment for under-recording, by five-year age group for 1996 (calendar year) and 1997/98, 1998/99, 1999/2000, 2000/2001, 2001/2002 and 2002/2003 against total projected deaths for the last year of the projection.

Copyright

The copyright worksheet opens automatically when the user opens the model workbook with macros enabled. (If macros are not enabled, the user will be unable to run any projections.)

Appendix C: Description of worksheets in provincial version workbook

The AssumptionsProv workbook is used to paste starting assumptions into the ASSA2008 *full* model to create a model of a chosen province. In order to run it the user must first ensure that the ASSA2001 *full* version of the model is open.

Initializer

This worksheet allows the user to specify:

- the province for which the data should be copied; and
- the name of the file into which the assumptions are to be pasted (the source or destination file).

By clicking the square button under the specification table, the user causes data to be copied from this workbook to the specified destination file.

EC – Common

This worksheet contains the following tables:

- a table giving the formulae used to calculate the modelled ANC prevalence, as well as the observed prevalence for 1995, 1999 and 2000 by 5-year age groups;
- a table giving observed and target rates of infection for ANC, STD clinics and CSW for each of the years from 1985 to 2008
- two tables showing the number of deaths by 5-year age group for years 1996-2007. One table shown male deaths and the other shows female deaths.

The values in these tables are copied into the corresponding cells of the *HIVTable* worksheet of the destination workbook.

EC – Asian

This worksheet contains the following tables:

- a single-row table giving the initial population. This is copied to the *Assumptions – Asian* worksheet;
- a table showing the percentage distribution of the male and female adult populations across the PRO, RSK and STD risk groups, as well as relative fertility of females in each risk group. The values are copied to the *Assumptions – Asian* worksheet, where the percentage in the NOT group is derived;
- a table showing the percentage distribution of the male and female immigrant adult population across the PRO, RSK and STD risk groups. The values are copied to the *Assumptions – Asian* worksheet, where the percentage in the NOT group is derived;
- a table showing the equation used to compute imported infectivity of PROs for male and female. The values are copied to the *Assumptions – Asians* worksheet;
- A table showing the relative rates of condom usage in the PRO, STD and RSK groups;
- moving to the right, a table showing, for each year of age, condom usage for the RSK group, age distribution of the starting population, the mortality improvement index and ultimate mortality for male and female, non-AIDS mortality to date for each year for male and female, Non-HIV fertility, the fertility improvement factor and ultimate rate, the start ratio, initial factor and reduction factor for the impact of HIV duration on fertility, and net male and female migration for each year from 1985 to 1996. The RSK condom usage data is copied to the *Assumptions – Asians* worksheet. The age distribution is copied to the *Population - Asians* worksheet. The mortality data is copied to the *MortTable – Asians* worksheet. The fertility data is copied to the *Non-HIV Fertility – Asians* and *HIV+ Fertility – Asians* worksheets. The

migration data are copied to the *Male Migration – Asians* and *Female Migration – Asians* worksheets.

- Below this there are two tables which contain assumptions for HIV interventions: IEC, STDs, VCT, PMTCT and ART;
- below those tables there is a table of empirical prevalence values (if any) which are pasted into the *HIV Table – Asians* sheet
- to the right of this table showing there is a table showing ART effectiveness assumptions. The values are pasted to the *Paediatric Survival – Asian* sheet.
- below this are two tables with the assumptions for rolling out of interventions which are pasted into the *Interventions – Asians* sheet
- below that is a table showing on the left, the ratio of prevalence of those attending antenatal clinics to that of all pregnant women, and on the right empirical prevalence by age for two periods (if the data are available). These tables are pasted into the *HIV Table – Asians* sheet.

Other worksheets

The remaining worksheets names take the form *Prov – Common* or *Prov – Population Group* and contain the same tables as the *EC – Common* and *EC – Asian* worksheets respectively.

Appendix D: Summary of worksheets

In theory, the user can change virtually any value or formula in the workbook. In practice, there are a limited number of values and formulae which can usefully and sensibly be changed. The third column of the table below indicates which worksheets contain cells with assumptions that can sensibly be changed. The fourth column indicates which worksheets are altered by the model during each projection.

Worksheet name	Worksheet function	Can be changed	Changed by model during projections
Assumptions	Sets overall assumptions	YES	YES
Results	Displays numeric results		YES
Interventions	Sets assumptions	YES	YES
Population	Sets assumptions	YES	YES
MortTable	Sets assumptions	YES	YES
SexActivity	Sets assumptions	YES	NO
Male HIVTable	Sets assumptions	YES	YES
Female HIVTable	Sets assumptions	YES	YES
HIVTable	Used for calibration projection	NO	YES
Male Adult Survival	Sets assumptions	YES	YES
Female Adult Survival	Sets assumptions	YES	YES
Male Staging	Calculates proportions of adults in each HIV stage	NO	YES
Female Staging	Sets assumptions	YES	YES
Paediatric Survival	Sets assumptions	YES	YES
Non-HIV Fertility	Sets assumptions	YES	YES
HIV+ Fertility	Sets assumptions	YES	NO
Male Migration	Sets assumptions	YES	NO
Female Migration	Sets assumptions	YES	NO
FemPRO	Calculates steps of projections	NO	YES
FemSTD	Calculates steps of projections	NO	YES
FemRSK	Calculates steps of projections	NO	YES
FemNOT	Calculates steps of projections	NO	YES
MalePRO	Calculates steps of projections	NO	YES
MaleSTD	Calculates steps of projections	NO	YES
MaleRSK	Calculates steps of projections	NO	YES
MaleNOT	Calculates steps of projections	NO	YES
MaleOLD	Calculates steps of projections	NO	YES
FemOLD	Calculates steps of projections	NO	YES
Young	Calculates steps of projections	NO	YES
AIDS Age Profile	Graphs results	NO	YES
HIV Prevalence	Graphs results	NO	YES
ANC Age Prevalence	Tests results against reality	NO	YES
ANC Age Profile	Graphs results	NO	YES
Calibration	Tests results against reality	NO	YES
Cumulative Deaths	Graphs results	NO	YES
Deaths	Graphs results	NO	YES
Pyramid	Graphs results	NO	YES
Mortality	Graphs results	NO	YES
Reported deaths – Female	Tests results against reality	NO	YES
Reported deaths – Male	Tests results against reality	NO	YES
Copyright			

Appendix E: Acronyms

ANC	Antenatal clinic
ART	Anti-retroviral therapy
ASSA	Actuarial Association of South Africa
CSW	Commercial sex worker
HIV	Human immuno-deficiency virus
IEC	Information and education campaigns
IMR	Infant mortality rate
MRC	Medical Research Council
MTCTP	Mother-to-child-transmission prevention
SADHS	South African demographic and health survey
SALT	South African Life Tables
STD	Sexually transmitted disease
VCT	Voluntary counselling and testing
WHO	World Health Organisation