Text S1: Derivation of Methods and Further Technical Information

Derivation of estimation of cohort incidence (used in Method 1 and 2)

From equations (1) and (2) on estimates of the number of seroconversions and person-years in the cohort, an estimate of incidence can be derived as follows:

$$\tilde{\lambda}_i = \frac{\text{seroconversions}}{\text{person-years}} \approx \frac{H_{i,T} - \tilde{\pi}_i H_{i,0}}{T\left(\frac{(N_{i,0} - H_{i,0}) + (N_{i,T} - H_{i,T})}{2}\right)} = \frac{2(H_{i,T} - \tilde{\pi}_i H_{i,0})}{T(N_{i,0} - H_{i,0} + N_{i,T} - H_{i,T})} \quad \text{.... (7)}$$

Then replacing values of $H_{i,j}$ with $H_{i,j} = p_{i,j} N_{i,j}$ and dividing by $N_{i,0}$

$$\tilde{\lambda}_i \approx \frac{2\left(p_{i,T}\frac{N_{i,T}}{N_{i,0}} - \tilde{\pi}_i p_{i,0}\right)}{T\left(1 - p_{i,0} + \frac{N_{i,T}}{N_{i,0}} - p_{i,T}\frac{N_{i,T}}{N_{i,0}}\right)} = \frac{2(Q_i p_{i,T} - \tilde{\pi}_i p_{i,0})}{T(1 - p_{i,0} + Q_i(1 - p_{i,T}))} \quad \text{.... (8)}$$

Here $Q = \frac{N_{i,T}}{N_{i,0}}$; If we ignore the possibility of sero-conversion and death occurring in the same interval, the relationship between the size of the cohort at successive sero-surveys can be found in the following way:

$$N_{i,T} \approx N_{i,0} - (1 - \tilde{\pi}_i) H_{i,0} - (1 - \exp(-\tilde{\mu}_i T))(N_{i,0} - H_{i,0}) \quad \text{.... (9)}$$

where $\tilde{\mu}_i$ is the mortality rate for those not infected with HIV. Dividing by $N_{i,0}$:

$$Q = \frac{N_{i,T}}{N_{i,0}} \approx 1 - (1 - \tilde{\pi}_i) p_{i,0} - (1 - \exp(-\tilde{\mu}_i T))(1 - p_{i,0}) \quad \text{.... (10)}$$

For the calculations presented here, we use $\tilde{\mu}_i = 0.01$ for all age-ranges, which approximates the observed mortality rate for uninfected individuals aged 15-44 in sub-Saharan Africa [1,2].

Derivation of cross-sectional measures

Case 1: Inter-survey period not longer than width of age-group $(T \le r)$

In this case, the estimate of incidence in the cross-sectional age-group includes the experience of two cohorts (Figure 1(c)). We calculate an average rate from these two cohort incidence rates, weighted according to the time spent by the cohorts in the fixed age-group. The time spent by each cohort is proportional to the shaded areas in Figure 1(c).

Total person-years spent in fixed age-group: $T.r$

Fraction of person-years spent by cohort $i-1$: $0.5T^2/T.r = T/2r$

Fraction of person-years spent by cohort i: $1 - T/2r$

Hence, equation (7): $\lambda_i = \left(1 - \frac{T}{2r}\right)\tilde{\lambda}_i + \left(\frac{T}{2r}\right)\tilde{\lambda}_{i-1}$.

Case 2: Inter-survey period longer than width of age-groups $(r < T \le 2r)$

The same logic applies when the inter-survey period is longer, but the weighting equation must be adjusted because more cohorts pass through the cross-sectional age-groups in the period. If $r < T \le 2r$, three cohorts pass through the cross-sectional age-group:

Fraction of person-years spent by cohort $i-2$: $0.5(T-r)^2/T.r = T/2r + r/2T - 1$

Fraction of person-years spent by cohort $i-1$:

$1 - \left(0.5r^2 - 0.5(T-r)^2\right)/T.r = 1 - \left(r/2T + T/2r + r/2T - 1\right) = 2 - r/T - T/2r$

Fraction of person-years spent by cohort i: $0.5r^2/T.r = r/2T$

Hence, the new expression to replace equation (7) is:

$$\lambda_i = \left(\frac{r}{2T}\right)\tilde{\lambda}_i + \left(2 - \frac{r}{T} - \frac{T}{2r}\right)\tilde{\lambda}_{i-1} + \left(\frac{T}{2r} + \frac{r}{2T} - 1\right)\tilde{\lambda}_{i-2} \quad \text{...(11)}$$

In theory, further expressions could be found for longer inter-survey periods (i.e. $T > 2r$), but since the linear approximations which underlie these methods work better over shorter intervals, the derived estimates would be less reliable and this is not recommended.

Alternative Formula for Calculating Cohort Survival Rates

Equation (5) holds when $T \approx r$. Experimentation in situations when this is not the case suggests that small departures from this (i.e. $T < r$) will not introduce large errors. However, we have derived an alternative formula that should give a better approximation when $r/2 \le T < r$.

$$\tilde{\pi}_1 \approx s(a_1, a_1 + T)$$

$$\tilde{\pi}_2 \approx \frac{s(a_1, a_2 + T)\lambda_1 + s(a_2 - r/4, a_2 + T)\tilde{\lambda}_{i-1}\exp(-\lambda_1)}{s(a_1, a_2)\lambda_1 + s(a_2 - r/4, a_2)\tilde{\lambda}_{i-1}\exp(-\lambda_1)}$$

$$\tilde{\pi}_i \approx \frac{\sum_{k=1}^{i-1}\{w_k s(a_k, a_i + T)\} + s(a_i - r/4, a_i + T)\tilde{\lambda}_{i-1}\exp\left(-\sum_{k=1}^{i-1}\lambda_i\right)}{\sum_{k=1}^{i-1}\{w_k s(a_k, a_i)\} + s(a_i - r/4, a_i)\tilde{\lambda}_{i-1}\exp\left(-\sum_{k=1}^{i-1}\lambda_i\right)} \quad \text{for i} > 2 \qquad \text{....(12)}$$

$$w_k = \lambda_k \exp\left(-\sum_{q=1}^{k-1}\lambda_q\right)$$

Details of simulation model

The simulation model generates prevalence and mortality measurement as a function of time for a given pattern of incidence. Here a is age and t in the time since the epidemic stared (both in years); $U(a,t)$ is the number of uninfected individuals with that many years of sexual activity and $I(a,w,t)$ is the number individuals infected w years ago with that many years of sexual activity. The rate of change of $U(a,t)$ and $I(a,w,t)$ is described by the following partial differential equations.

$$\begin{aligned}\frac{\partial}{\partial a}U(a,t)+\frac{\partial}{\partial t}U(a,t) &= -U(a,t)\left(\lambda(a,t)+\mu(a)\right)\\ \frac{\partial}{\partial a}I(a,w,t)+\frac{\partial}{\partial w}I(a,w,t)+\frac{\partial}{\partial t}I(a,w,t) &= \lambda(a,t)U(a,t)-I(a,w,t)\left(h(a-w,w)+\mu(a)\right)\end{aligned} \quad \ldots(13)$$

$$h(x,z)=\left(\frac{k}{\beta_x}\right)\left(\frac{z}{\beta_x}\right)^{k-1} \quad \ldots(14)$$

$\lambda(a,t)$ is the instantaneous incidence rate; $\mu(a)$ is the rate of mortality from causes other than AIDS; $h(x,z)$ is the hazard of AIDS-death for those infected at age x who have survived z years. In these simulations, survival after infection is parameterised as in Table 1 and the background mortality rates are based on World Bank estimates for Africa in the pre-AIDS era [3].

The boundary conditions are:

$$\begin{aligned}I(0,0,0) &= 0\\ U(a,0) &= \phi(a)N_0\\ U(1,t) &= \phi(1)N_0\end{aligned}$$

where $\phi(a)$ is the fraction of the population at age a, in the absence of AIDS-mortality.

Three scenarios for the age pattern of incidence are simulated – constant incidence over age, incidence highest at older ages (reflecting patterns expected early in epidemics) and incidence highest at younger age (reflecting patterns observed in mature epidemics) (see Figure 4(b)). To generate changes over time, these age-specific rates are multiplied by a scaling variable, which is set to unity and then increases or decreases (or remains the same) at a specified time (see Figure 3 for illustration of temporal changes).

The impact of ART is simulated by manipulating the survival distribution with HIV so that it instead reflects survival until "ART is started". Models indicate that in the current mode of ART delivery (Hallett *et al.*, submitted), individuals that are started on ART will typically have CD4 cell counts below 200 and be within two years of death without therapy. Therefore, here it is assumed that individuals that can access ART are initiated two years before they would otherwise die of AIDS. Survival for those individuals that do not receive ART is distributed exponentially with mean 2 years.

A fraction of individuals that newly become in need of ART, *g(t)*, receive ART. This fraction can change over time to represent no provision in the early part of the epidemic, a linear scale-up of provision over five-years to 30% and then a maintained level thereafter.

Long-term survival of individuals on ART is not yet known, but following recommendations from the UNAIDS Reference Group on Modelling, Estimates and Projections [4], we assume that survival on ART does not vary with age and is exponentially distributed with mean 10 years.

References

1. Zaba B, Todd J (2006) Survival following HIV infection in adults in developing countries prior to initiation of Anti-Retroviral Therapy: Technical Report to UNAIDS. Genva: UNAIDS.
2. Department of Economic and Social Affairs- Population Division http://wwwunorg/esa/population/unpophtma (accessed 16th April, 2007).
3. Anderson RM, May RM, Ng TW, Rowley JT (1992) Age-dependent choice of sexual partners and the transmission dynamics of HIV in Sub-Saharan Africa. Philos Trans R Soc Lond B Biol Sci 336: 135-155.
4. UNAIDS reference Group on Estimates Modelling and Projections (2006) Improviing estimates of national HIV burdens and ART need, and modelling the impact of prevention programmes. Report of a meeting held in Athens, Greece, December 13-15th. [available from http://www.epidem.org/Publications/Athens2005report.pdf].