Improvements to estimation packages

Report of a meeting of the UNAIDS Reference Group for Estimates, Modelling and Projections held in Glion, Switzerland, July 19-20th 2006

TECHNICAL REPORT AND RECOMMENDATIONS

Joint United Nations Programme on HIV/AIDS

UNAIDS

UNICEF • UNDP • UNFPA • UNDCP
ILO • UNESCO • WHO • WORLD BANK

UNAIDS

The meeting of the UNAIDS Reference Group on Estimates, Modelling and Projections (the 'Epidemiology Reference Group') was organised for UNAIDS by the UK secretariat of the reference group (http://www.epidem.org) based at Imperial College London. Participants of the meeting are listed at the end of this document. The recommendations in this document were arrived at through discussion and review by meeting participants and drafted at the meeting.

Dr Peter White, London, March 2007.

Introduction

The Reference Group

The Joint United Nations Programme on HIV/AIDS (UNAIDS) *Reference Group on Estimates, Modelling and Projections* exists to provide impartial scientific advice to UNAIDS and the World Health Organization (WHO) on global estimates and projections of the prevalence, incidence and impact of HIV/AIDS. The Reference Group acts as an 'open cohort' of epidemiologists, demographers, statisticians, and public health experts. It is able to provide timely advice and also address ongoing concerns through both *ad hoc* and regular meetings. The group is co-ordinated by a secretariat based in the Department of Infectious Disease Epidemiology, Imperial College London (www.epidem.org).

Aim of the meeting

The aim of this meeting was to bring together experts to produce recommendations on a range of topics, including proposals for development of EPP to (i) incorporate changes in risk behaviour, effects of provision of ART, and changes in urbanisation; and (ii) use data from multiple population-based surveys as well as ante-natal clinic surveillance; approaches to plausibility range estimation in Workbook and Spectrum; potentially using a Bayesian approach to uncertainty in EPP; ANC prevalence trends; and insights from national surveys of HIV prevalence.

Approach

The meeting featured both presentations of recent data and group discussions, which focused on specific technical issues. Presentations included examination of the performance of EPP for specific countries, proposals for incorporating effects of ART into EPP and improving implementation in Spectrum, proposals for improved approaches to estimation of plausibility bounds for estimates, and methods to analyse trends in ANC prevalence (see Appendix I for a complete list of presentations).

The meeting was attended by 18 experts from 4 countries (see Appendix II for a list of participants). Each contributed, not only data, insights and analysis, but also worked hard to produce a set of recommendations for UNAIDS and WHO, drafted at the meeting. We would like to thank them for their hard work and attendance at the meeting.

The recommendations drafted at Reference Group meetings give UNAIDS and WHO guidance on how best to produce estimates of HIV/AIDS, an opportunity to review current approaches and also help to identify information needs (earlier reports are published on the Reference Group website www.epidem.org). This transparent process aims to allow the statistics and reports published by UNAIDS and WHO to be informed by impartial, scientific peer review.

IMPROVEMENTS TO EPP

1. Difficulties in fitting EPP to some national HIV surveillance trends: problems and possible solutions

Some epidemic patterns are difficult to fit with EPP. For example, in rapid epidemics (e.g. South Africa), the automated fit may have a peak occurring too early and at too low a level. This is probably due to the binomial likelihood formula giving less weight to the higher-prevalence data-points. Manipulating f_0, which determines the saturation level and hence the peak prevalence, is possible but it requires judgment.

In Kenya, the general pattern of ANC prevalence was to rise to a peak and then decline to a lower, sustained, level. In general, EPP provides a poor fit to these patterns: either it is too flat and fails to reproduce the peak, or it fits the peak but then declines to below the steady-state level. The problem is probably that EPP was designed to reproduce the natural dynamics of an epidemic and so does not capture changes caused by behaviour change. (It was noted that improvements in data quality meant that recent declines were probably real, but that the height of the peaks - which occurred longer ago - may not be reliable. However, it was also noted that there are some rapid declines in ANC prevalence which cannot be explained by natural dynamics, indicating problems with the data.)

The 'true' values of parameters may vary with time, particularly the epidemic growth rate, *r*, reflecting a change in behaviour. However, EPP does not allow this at present. It was recommended that this facility be incorporated, with the user specifying the period over which *r* changes (which may be an instantaneous step-change). The improvement in the fit afforded by this change needs to be examined to determine if allowing *r* to vary is justified in each particular instance. Examples of countries with behaviour change where the fit may be significantly improved are Botswana, Burkina Faso, Kenya, Malawi, Uganda, Zimbabwe. It was notes that users may wish to make projections under an assumption of behaviour change continuing into the future.

Noisy data for many countries (e.g. Zimbabwe, Ghana) mean that a range of curves have similarly-good fits to the 'cloud' of data points. In these cases the likelihood profile is flat, resulting in poor discrimination between candidate fits. Future consideration of using a Bayesian approach which would enable information from multiple countries rather than fitting to a single country's data was recommended. Implementing a Bayesian approach to uncertainty estimation was also recommended (see below).

Cases have been found where the automated fitting algorithm produces an inferior fit compared with manual fitting (as determined by log-likelihood), indicating that it finds a local maximum likelihood, not the global. Finding global maxima is computationally intensive and so not feasible for individual EPP users. The approach recommended in the short term is for the user to fit by eye and then use the automated curve-fitter to fine-tune the fit. An alternative approach to be investigated is to create a library of runs (possibly region-specific) from which the most appropriate is selected for each country's data.

It was noted that model fits are dominated by the phi parameter, which needs to be constrained to a reasonable range of values. Providing the option to vary both t_0 and f_0 would be beneficial. Testing of the following approaches was recommended: simultaneous fitting of all four parameters with phi constrained; using five starting values of phi in the range (-100, +100) and allowing the user to select the best fit; initially fitting to the median of multiple data points as the starting point for subsequent fitting.

2. ART in EPP, Spectrum, and Workbook

Since ART provision is increasing, it will become necessary to incorporate its effects into EPP. Although ART will have had little impact in most countries on data to date – and will not do so for another few years – countries will want to make projections its impact, including under different scenarios of increasing provision through time.

EPP and Spectrum will require inputs are the number of individuals on ART and their survival. It is important for models to distinguish between untreated individuals who require ART and those who are receiving it. Consideration needs to be given to whether it is necessary to make the rate of drop-out higher in the first year of therapy or not, and how to incorporate the provision of 2nd- and 3rd-line therapy for those countries where it is available. The most likely approach in the short term will be to calculate a survival function reflecting different patterns of ART provision external to EPP and then input summary parameters.

In Spectrum, improvements will be made to how ART need and the impact of ART provision are estimated. It was recommended to use two Weibull functions to describe progression from HIV infection to ART-eligibility and from ART-eligibility to death, and that provision of 2nd-line therapy where available should be included. An output of the estimated number of lives saved by ART should be provided.

It was noted that Workbook will also need ART to be incorporated.

The prevalence of ART-recipients in ANC, GUM clinics, etc needs to be assessed empirically for calibration and validation of models.

3. Modelling demographic changes

Developing countries are still predominantly rural but most are rapidly becoming more urban, with most growth occurring in medium-sized cities. However, migration patterns are typically poorly-characterised, and data often relate only to net flows, which can be comprised of multiple flows in different directions. Mortality data are also often poor, particularly with regard to different causes of death. Fertility data, typically obtained from DHS, tend to be relatively good.

HIV's effects on mortality and fertility can significantly distort population age-profiles, which can in turn affect the dynamics of HIV's transmission. EPP does not take account of this effect at present, but it will become increasingly important in the next decade, as reduced birth rates over the past two decades cause a decline in the young adult population in the present and near future.

A study of the impact of urbanisation on model prevalence estimates is being conducted to determine under what circumstances any effects are important – perhaps when there are large changes in the % urban, or large urban/rural differences in HIV prevalence.

EPP is a generic package, designed to have limited data requirements to enable its widespread use. For countries with the necessary data, increasing the sophistication of EPP would improve its estimates. It would be desirable to have a more-realistic model that can take advantage of detailed epidemiological and demographic information where it is available (e.g. by introducing age-structure or other population stratification), and to make projections under different scenarios. However, this would also increase the complexity of EPP's behaviour, and ensuring the robustness and validity of its fitting whilst ensuring it can also run with a minimum set of data and default assumptions would require substantial work.

Desirable additional features include:

- Integrating the basic epidemiological model into a demographic impact model that projects subpopulations by sex, age and stage of infection.
- Allowing demographic data to vary over time (according to the No-AIDS projection).
- Allowing epidemiological parameters to vary over time.
- Linking the model to a database allowing for processing of many projections and scenarios in a multi-user environment.
- Estimation of HIV-orphanhood.
- Introducing country-specific epidemiological parameters (incubation period, MTCT, etc.).

4. Calibration to data from more sources

Where no population-based prevalence data are available, ANC data often have to used to estimate both trends in, and absolute values of, prevalence. Where a single population-based prevalence estimate is available along with ANC data, the ANC data are used to estimate the trend, with the population-based estimate use to calibrate the absolute prevalence. Increasingly, data from more than one population-based survey are becoming available, providing more information on both prevalence level and trend. Additionally, improved mortality data offer another source of information. Hence, more-sophisticated methods of calibration are required.

More research is required, but the currently-preferred approach is to include national-survey prevalence estimates as data points and fit an ANC-bias parameter. It is easy to include these data in the likelihood function and enables national surveys to inform trend estimates where there is more than one of them. This approach has the advantage of not having to assume that ANC data are unbiased sample from 'true' distribution. It was noted that DHS prevalence estimates may also be biased; corrections need to be made before the data are input into EPP.

The magnitude of ANC bias tends to vary geographically, but DHS data are not usually representative at a sufficiently fine scale for local adjustments to be made. Also, it is important to note that introduction of prevention-of-mother-to-child-transmission (PMTCT) programmes may have created a new source of bias for ANC sites where it is offered, by attracting HIV-positive women to them.

For concentrated epidemics, it is important to be aware of the common limitations of prevalence estimates. Importantly, high-risk groups may be excluded from household samples, resulting in under-estimation of prevalence. Also, when prevalence is low, confidence intervals on estimates may be large, limiting the scope for model-validation, unless the sample size is large. However, enlarging general population surveys is not a cost-effective way to better-characterise those who are at most risk of having HIV; a more targeted approach is required.

In the future, a method for calibration to data on incidence as well as prevalence will be required, but reliable incidence estimates for most countries will be years away so this is not a pressing need currently.

5. Improved estimation of plausibility bounds

It was recommended that Bayesian melding (BM) should be the statistically-robust approach used to estimate plausibility bounds. This involves sampling from ranges of parameter values ('priors') to generate a set of 'candidate' curves from which there is likelihood-weighted resampling to produce a range of plausible model fits that indicate the amount of uncertainty in the estimates. Initially it should be applied only to EPP, because the computational demands of its inclusion in Spectrum could be very heavy. It was noted that the Monte Carlo bootstrapping approach that is currently advocated has yet to be automated in Spectrum in a user-friendly manner, although work is on-going. Incorporation of an automated bootstrapping or BM approach into Workbook would be a major undertaking and is not recommended at present.

BM needs to be developed in parallel to the other changes recommended at the meeting, and EPP's existing features should be preserved, in case completing the implementation is not possible in time for the start of the next round of training in EPP. Also, users will require training in the use of BM.

There are a number of detailed technical issues to be addressed regarding BM implementation in EPP, including testing on a number of prototype countries. A small progress meeting is planned for October, followed by a report to the November Reference Group meeting.

6. Analysis of ANC clinic data as a proxy for trends in HIV prevalence in the most-affected countries

In 2001 the UN General Assembly made a Declaration of Commitment to reduce HIV prevalence among young people (aged 15 to 24) by 25% in the most affected countries by 2005, and to reduce it by 25% globally by 2010. To assess whether these goals have been reached there is a need to assess prevalence trends in countries. There is a need to provide guidance on appropriate robust methods for determining whether apparent trends in HIV prevalence are statistically significant in different situations. Complications include missing data, inconsistent sampling, variation between countries in the amount of rural coverage, heterogeneity in trends within a country – e.g. different patterns in urban and rural areas. Another problem is that different countries were in different phases of epidemics during the period that is to be assessed, so it is possible that in some cases prevalence may have fallen in the absence of an effective intervention whilst in others prevalence may have risen or remained constant despite an effective intervention.

In the past, trends in the median prevalence of multiple sites have been used, but these are not ideal since they disregard most of the information available, particularly between-site variation. If this method is used then it is essential that only consistently-sampled sites are used to calculate each median, because otherwise changes in sampled sites could create spurious trends – e.g. an expansion of rural ANC sampling would probably result in a spuriously declining prevalence estimate since rural prevalence is usually lower.

The appropriate method of analysis depends upon the sampling process, including its precision; the number of clinics sampled (all / most / few); the number of women sampled per clinic; the between-clinic variability in prevalence; sources of bias; the sampling design (random or convenience sample); and how representative are the pregnant women sampled of the population of interest. Ideally, multiple tests would be performed to examine if their results are consistent.

The sampling method needs to be consistent through time and across sites, and only sites that have been sampled at all time-points used for the analysis should be used, to avoid problems caused by missing data. At least three time-points are required for regressions, whilst paired non-parametric tests require two time-points.

Work is on-going, but suggestions made at the meeting are summarised here. Non-parametric methods such as the Paired Sign test or Wilcoxon Sign-Rank Test require fewer assumptions, and so are more robust, but they have less power than parametric methods. The chi-squared linear trend test may be appropriate: first test to see whether there is a linear trend; then test whether it is significantly up or down. Another approach is to use linear regression: first test to see whether all sites have the same trend (SITE x YEAR interaction); then test whether that trend is up or down.

Appendix I: Meeting Agenda

Wednesday 19th July		
0900 (15)	Opening remarks	Geoff Garnett / Peter Ghys
Improvements to EPP		
0915 (30)	Difficulties in fitting EPP to some national HIV surveillance trends: problems and possible solutions	Josh Salomon Tim Brown John Stover
0945 (15)	*Discussion: brief, whole-group, for clarification of details*	-
1000 (15)	Comparison of EPP & ASSA fits to Botswana, South Africa & Zimbabwe	Rob Dorrington
1015 (35)	Potential use of time-varying parameters in EPP to model changes in risk behaviour and epidemiology	Tim Brown & Nick Grassly
1050 (30)	Incorporating ART's effects on survival	Geoff Garnett Josh Salomon Tim Brown
1120 (10)	*Discussion: brief, whole-group, for clarification of details*	-
1130 (20)	*Break*	-
1150 (20)	Urban growth and modelling its effects.	Thomas Buettner
1210 (10)	*Discussion: brief, whole-group, for clarification of details*	-
1220 (20)	Fitting EPP to more than one national population based survey	Nick Grassly & Josh Salomon
1240 (20)	*Discussion: brief, whole-group, for clarification of details*	-
Lunch		
1300 (60)	*Lunch*	-
Discussion		
1400 (2h40)	*Discussion (in work groups)* *Including coffee break*	-
1640 (50)	*Discussion: reporting back*	Peter Ghys / Geoff Garnett
1730	Close	-

Thursday 20th July		
0900 (15)	Opening remarks	Geoff Garnett / Peter Ghys
Plausibility range estimation		
0915 (30)	Plausibility range estimation – implementation in EPP & Workbook	Meade Morgan / Josh Salomon
0945 (15)	Plausibility range estimation – implementation in Spectrum	John Stover
1000 (30)	Assessing uncertainty with the EPP model using Bayesian melding	Adrian Raftery
1030 (25)	Discussion	-
1055 (35)	*Break*	-
1130 (25)	Discussion	-
ANC prevalence trends		
1155 (35)	Evaluation of ANC prevalence trends: quantifying the decrease and assessing its significance (1)	Rob Lyerla
1230 (20)	Evaluation of ANC prevalence trends: quantifying the decrease and assessing its significance (2)	Meade Morgan
1250 (10)	Discussion	-
Lunch		
1300 (60)	*Lunch*	-
National surveys: insights for generalized epidemics & role in concentrated epidemics		
1400 (30)	Vietnam: comparison of AEM/EPP estimates with DHS survey estimates	Tim Brown
1430 (25)	Senegal & Madagascar: Comparison of EPP adjusted with survey	Karen Stanecki
1455 (20)	Urban ANC vs Urban survey results in generalised epidemics: Should urban ANC be adjusted?	Eleanor Gouws
1505 (45)	Discussion	-
1600	Close	-

Appendix II: List of Participants

Ties Boerma
WHO, Geneva, Switzerland

Tim Brown
Senior Fellow, East-West Center,
Honolulu, USA

Thomas Buettner
Chief, Estimates and Projection Section
UN Population Division, USA

Rob Dorrington
Director, Centre for Actuarial Research
University of Cape Town, South Africa

Tim Fowler
Chief, Health Studies Branch
International Programs Center
U.S. Census Bureau, USA

Geoff Garnett
UNAIDS Epidemiology Ref. Group Secretariat
Department of Infectious Disease Epidemiology
Faculty of Medicine, Imperial College London, UK

Peter Ghys
Manager, Epidemic and Impact Monitoring,
Policy, Evidence and Partnerships Department,
UNAIDS, Geneva, Switzerland

Nicholas Grassly
Department of Infectious Disease Epidemiology
Faculty of Medicine, Imperial College London, UK

Eleanor Gouws
Epidemic and Impact Monitoring,
Policy, Evidence and Partnerships Department,
UNAIDS, Geneva, Switzerland

Rob Lyerla
Epidemic and Impact Monitoring,
Policy, Evidence and Partnerships Department,
UNAIDS, Geneva, Switzerland

Meade Morgan
Health Scientist, Global AIDS Program,
Centers for Disease Control and Prevention
Atlanta, GA 30333, USA

Adrian Raftery
Departments of Statistics and Sociology,
University of Washington, Seattle, USA

Joshua Salomon
Department of Population & International
Health Harvard School of Public Health, USA

Karen Stanecki
Epidemic and Impact Monitoring,
Policy, Evidence and Partnerships Department,
UNAIDS, Geneva, Switzerland

John Stover
Futures Institute
Glastonbury CT 06033, USA

Neff Walker
UNICEF, New York, USA

Peter Way
Chief, International Programs Center
U.S. Census Bureau
Washington, DC 20233-8800, USA

Peter White
UNAIDS Epidemiology Ref. Group Secretariat
Department of Infectious Disease Epidemiology
Faculty of Medicine, Imperial College London, UK