
Methodology – Understanding the HIV estimates

Produced by the Strategic Information and Monitoring Division

Contents

Chapter 1	3
The data	
Chapter 2	6
From the data to the estimates	
Additional sources of information	10

Chapter 1

The data

1. What data are used to develop the HIV estimates?

The *precise* numbers of people living with HIV, people who have been newly infected or who have died of AIDS are not known. Achieving 100% certainty about the numbers of people living with HIV globally, for example, would require testing every person in the world for HIV regularly—which is logistically impossible and poses ethical problems. But we can estimate those numbers by using other sources of data.

UNAIDS estimates are based on pertinent, available data—including surveys of pregnant women attending antenatal clinics, population-based surveys (conducted at the household level), sentinel surveillance among populations at higher risk of HIV infection, case reporting, as well as other surveillance information.

Different sets of data are used to calculate estimates of HIV prevalence for *generalized* (high-level – where HIV is firmly established in the general population and sexual networking is sufficient to sustain an epidemic independent of sub-populations at higher risk of infection) and *concentrated* (low-level – where HIV is concentrated in groups with behaviours that expose them to a high risk of HIV infection) epidemics.

In countries with *generalized* epidemics, estimates of HIV prevalence are primarily based on surveillance among pregnant women attending sentinel antenatal clinics (ANC). Such data are collected on a routine basis and are currently our primary basis for the assessment of trends. Population-based household surveys that include testing for HIV infection (conducted much less frequently) are used to improve the antenatal clinic data-based estimates. If countries have conducted such a survey, the results are used to calibrate the trend in HIV prevalence. Countries who have not conducted these surveys calibrate their HIV prevalence trends based on a global default adjustments derived from the comparison of HIV prevalence between national surveys and ANC surveillance in other countries.

For countries with low-level or *concentrated* epidemics, HIV estimates are based on studies among key populations who are at higher risk of HIV exposure—such as people who inject drugs, sex workers, or men who have sex with men.

Countries with concentrated epidemics sometimes have additional sources of data which can help refine estimates. In countries with high and consistent coverage of case reporting, these data can add to the estimation process and make estimates more precise.

More and better data from country surveillance and steady improvements in the modelling methodology are enabling countries, with the support of UNAIDS, to develop more accurate estimates.

2. What are the strengths and weaknesses of using antenatal and household surveys in estimating HIV infection levels for generalized epidemics?

Each of these methods has its strengths and weaknesses. Generally, estimates based on antenatal clinic attendees provide a good indication of HIV infection *trends* among the 15-49 year-old population over time.

Studies have shown that high proportions of women in most of the highly-affected countries have access to antenatal clinic services. Where possible, estimates derived from antenatal clinic data have been compared at local level with HIV prevalence data acquired in community-based surveys. Such validation exercises have concluded that estimates based on antenatal clinic sentinel surveillance provide a good approximation of HIV prevalence among adults aged 15-49 (men and women combined) in the local community. However, ANC surveillance is limited in that it only samples pregnant women attending public health services and therefore excludes women who are not pregnant or sexually active and who do not attend public health clinics. The most important limitation is often related to the selection of sentinel antenatal clinics. In general, clinics with larger volumes of pregnant women are included to obtain the minimum required sample size during the few weeks of the annual survey of sentinel clinics. Such clinics are more likely to be in urban areas, and the sample of clinics is often not geographically representative. Remote rural clinics are underrepresented for the most part, although countries are increasingly trying to increase their representation as in-country surveillance efforts are expanding. Finally, by definition, ANC surveillance does not provide information about HIV levels among men.

National population-based household surveys, on the other hand, can reveal important information about the national prevalence level and about the spread of HIV in a country. These surveys are generally geographically representative and can provide estimates for the general population as well as for different subgroups, such as prevalence in urban and rural areas, men and women, different age groups, and different regions. However, household-based surveys by their nature exclude certain high-risk populations (e.g. homeless people, people living in hostels, army recruits, etc.) and might therefore underestimate HIV prevalence. Non-response due to absence from households and refusing to participate in these surveys could also lead to bias in the HIV estimates.

Population-based surveys are costly, complex undertakings, especially if biological testing is included. Therefore, they are done with long intervals in-between survey rounds. Many countries have done only one national survey with HIV testing since 2000; an increasing number of countries have conducted two or more surveys.

Considered together, the various data sources can yield more accurate estimates of HIV infection levels and the demographic impact of AIDS. However, HIV estimates (whether derived from household surveys or sentinel surveillance data) need to be assessed carefully, and the data and assumptions reviewed continually.

3. Are population-based surveys more accurate than antenatal surveys?

For all diseases a sound population-based sample provides better estimates of disease prevalence than a clinic-based sample. National population-based surveys reveal important information about the national prevalence level and about the spread of HIV, particularly among young people, men and residents in rural areas. If response rates are good (e.g. over 75%) and there is no evidence of systematic biases of exclusion of a large proportion of the population with likely different levels of HIV infection, then national estimates that consider data from all sources (surveillance, population-based surveys and if available mortality data) should be close to the true situation.

4. Is sentinel surveillance more accurate than case reporting?

Case reporting generally tends to substantially underestimate the number of people living with HIV. However, in countries that have extensive voluntary counselling and testing programmes, case reports may enable more precise estimates to be developed. Nonetheless, case reporting is unlikely to capture people living with HIV who were recently infected, and who therefore present no symptoms of infection. For these reasons, case reports can only indicate the minimum number of people living with HIV.

On the other hand, reliance on *sentinel surveillance* of at-risk groups at service delivery points can lead to overestimation of HIV prevalence in these groups. This is because such surveillance in some cases detects HIV infection rates among individuals who are at highest risk of HIV infection. For example, sentinel surveillance among sex workers or their clients sometimes is based on those who seek treatment at sexually transmitted infection clinics—and who, by definition, have had unprotected sex. However, other sex workers and clients who *do* practice safe sex—and who therefore tend not to present at these clinics with sexually transmitted infections—generally are not captured in this surveillance.

Chapter 2

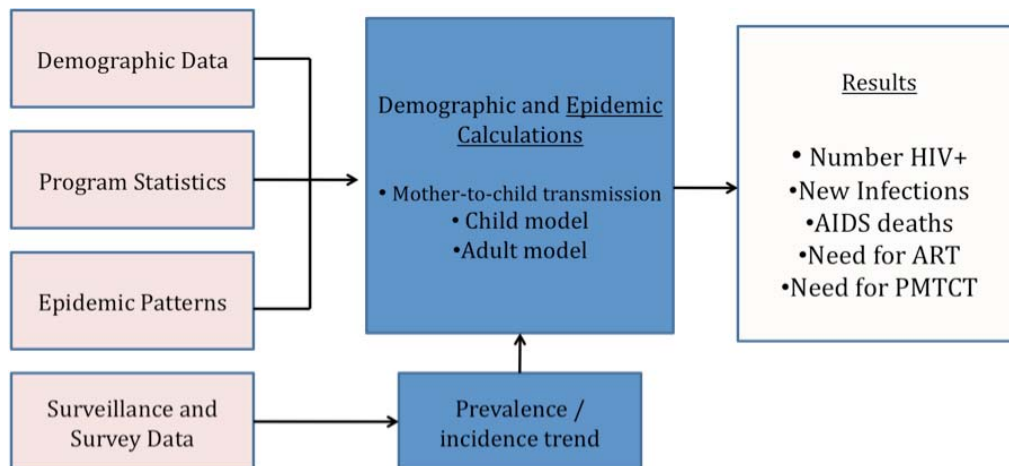
From the data to the estimates

5. How are the HIV estimates arrived at?

Country teams use the software tool Spectrum (www.futuresinstitute.org) and its AIDS Impact Module to estimate the impact of the HIV epidemic. Every two years UNAIDS and partners conduct regional workshops, training national personnel and technicians on the specific tools and methodologies used to produce the national estimates. These teams are then responsible for the HIV estimates and projections. The methods allow for standardization in measurement methods and allow cross-national comparisons and regional aggregation and estimates.

Estimates for countries and regions are generated using pertinent, available data—including surveys of pregnant women attending antenatal clinics, population-based surveys (conducted at the household level), sentinel surveillance among key populations at higher risk of HIV infection, case reporting, information on antiretroviral therapy and on programmes for preventing mother-to-child infection and demographic data.

Diagram of Spectrum AIDS Impact Module



6. Can new estimates be compared with those from previous years?

The latest estimates cannot be compared directly with estimates published in previous years. Nor should these latest estimates be compared directly with those UNAIDS will publish in the years to come. The assumptions, methodologies and data used to produce the estimates are gradually changing as a result of on-going enhancement of our knowledge of the epidemic. The data that the countries include in the models increases each year providing a more complete understanding of the epidemic curve. Comparing the latest estimates with those published in previous years is liable to yield misleading conclusions. The new trends are based on new assumptions and enhancements and the latest surveillance and programme data

from countries.

In a nutshell, the latest estimates will be more accurate and reliable than those produced in previous years, since they are based on improved methods and more data than earlier estimates. In each round of estimates countries produce estimates for the current year as well as previous years. Trends should be assessed based data produced in the same round since previous rounds will not be comparable.

7. Why are ranges published with the HIV estimates?

The ranges reflect the degree of uncertainty associated with estimates and define the boundaries within which the actual numbers lie.

UNAIDS reports point estimates (for example, fixing HIV prevalence in country X at 12.5%). The ranges of uncertainty around those point estimates are also published. The magnitude of the uncertainty ranges depends on the amount and quality of the data used to create the estimates. All estimates are associated with some level of uncertainty.

Because the quality of data varies from country to country, the ranges of uncertainty surrounding the estimates can be wider or narrower depending on the country. As well, presenting point estimates might encourage a false sense of precision, notwithstanding the fact that ranges of uncertainty are also provided.

Improved methods, enhanced data and new estimation tools are enabling a better understanding of the degrees of uncertainty that surround HIV and AIDS estimates. This is part of an on-going process of improving estimates and developing appropriate ranges—all of which are vital for effective HIV planning and programming at national and regional levels.

UNAIDS is confident that the actual numbers of people living with HIV, people who have been newly infected or who have died of AIDS lie within the reported ranges.

The ranges vary, depending on the quality of HIV data available in different countries. Four factors determine the extent of the ranges around the HIV estimates:

- (i) *The HIV prevalence level* – Ranges tend to be smaller when HIV prevalence is higher. Thus the bounds around the best estimate of adults living with HIV in Zambia would be relatively small compared to a lower prevalence country such as Djibouti where the ranges will be much wider.
- (ii) *The quantity of data* – Countries with more data have smaller ranges than countries with few sites or smaller sample sizes. The ranges for countries with data for a few years and a few surveillance sites (e.g. <5) are wide compared to the ranges for countries with many years of data for many surveillance sites (e.g. >20).

Countries in which a national population based survey has been conducted will generally have smaller ranges around estimates than in countries where such

surveys have not been conducted. National surveys of HIV infection--which are generally more representative of the general population and of specific sub-groups such as men and women, urban and rural areas, or different regions--will generally reduce the level of uncertainty around estimates of HIV.

(iii) *The number of steps or assumptions used to arrive at an estimate* – The more steps and assumptions, the wider the uncertainty range is likely to be (since each step introduces additional uncertainties). For example, ranges around estimates of adult HIV prevalence are smaller than those around estimates of HIV incidence among children, which require additional data on the probability of mother-to-child HIV transmission. The latter are based on prevalence among pregnant women, the probability of mother-to-child HIV transmission, and estimated survival times for HIV-positive children. There is therefore greater uncertainty in these estimates than for adult prevalence alone.

(iv) *The type of epidemic (generalized or low-level/concentrated)* – Ranges tend to be wider in countries with low-level or concentrated epidemics than in countries with generalized epidemics because in low-level or concentrated epidemics, one needs to estimate both the numbers of people in the groups at higher risk of HIV infection *and* HIV prevalence rates in those groups.

8. Why are the data for some countries excluded from the reports?

In concentrated epidemic countries the estimated number of women living with HIV who are pregnant is not easily available. Women living with HIV in these countries are primarily sex workers or partners of men who have sex with men or drug users and thus are likely to have different fertility levels than the general population. UNAIDS restricts the presentation of any estimates that are child-based in concentrated epidemic countries (that means any data on mother to child transmission or estimates related to children infected through mother to child transmission).

If there is not enough data in the Spectrum file to confidently state whether a decline in incidence has occurred UNAIDS will limit the historical presentation of data on incidence for those countries to avoid making inaccurate deductions about trends. Specifically the data are restricted on incidence if there are less than 4 data points for the principal populations, or there has been no data for the last four years.

Many high income countries rely solely on case reporting data for their national information on HIV. Spectrum software is not ideal for fitting data to these data. In the 2012 round of estimates (the round in which the last data input into the model was for year 2012 which were published in September 2013) data from these countries were excluded. However UNAIDS creates Spectrum files that attempt to estimate the epidemics in those countries. Global and regional estimates include the data from these countries that did not produce files.

UNAIDS only publishes estimates for countries with populations of 250,000 or more.

Finally, in a few instances UNAIDS will not publish estimates when additional data or analyses are needed to produce robust estimates.

9. What are the challenges around estimating the number of people who die of AIDS each year?

Estimates of adult AIDS mortality are based on several assumptions and additional sets of data—including estimates of the numbers of adults and children who are HIV-infected, numbers of people receiving antiretroviral therapy, and estimated survival times from infection with HIV to death for adults and children infected with HIV.

Civil registration systems are potentially the best source to obtain an estimate of the mortality due to AIDS. However, in most countries with generalized epidemics, coverage of civil registration is too low to provide useful information on AIDS mortality. Some countries have local demographic surveillance or general information on adult mortality from censuses and surveys that can help estimate mortality levels due to AIDS.

Estimating mortality in countries with low-level or *concentrated* epidemics is even more difficult. Some at-risk groups are likely to have different background mortality, in other words they are more prone to other causes of death (for example, injecting drug users are vulnerable to fatal drug overdoses and other life-threatening hazards). All this can have substantial effects on patterns of mortality. Unfortunately, country-specific data on mortality are seldom available.

10. What is being done to improve national HIV estimates?

New and different sources of data, such as national population-based surveys, are enabling more accurate estimates and more refined understandings of the epidemic's trends. In addition the results from the models are being compared with other sources such as data from epidemiological research studies, demographic surveillance sites, mortality surveys, or early infant diagnosis results to improve the assumptions.

Importantly, the roles of national AIDS programmes have changed significantly since the first set of UNAIDS country specific estimates was produced in 1997. Initially, countries were requested to comment on provisional estimates. Since 2003, the UNAIDS secretariat, along with its partners (including East-West Center, Futures Institute, WHO, UNICEF, the US Census Bureau and the US Centers for Disease Control and Prevention) have carried out a series of regional training workshops in which epidemiologists from over 150 countries were trained in the HIV estimation process. Such efforts have led to much greater involvement by national programmes, national statistics offices and other government and academic organizations in the production of estimates. The result has been better quality estimates, due to the use of additional data and the application of local knowledge.

UNAIDS continues to work with countries, partner organizations and experts to improve data collection. These efforts will ensure that the best possible estimates are available to assist governments, non-governmental organizations and others in gauging the status of the epidemic and monitoring the effectiveness of prevention and care efforts.

Additional sources of information

For a detailed description of the methods, software, quality of data and development of ranges, please see a series of articles published in a supplement in the journal *Sexually Transmitted Infections* in December 2012. The articles can be downloaded for free from the website at <http://sti.bmj.com/>

More information on the advisory group to the estimation process, the UNAIDS Reference Group on Estimates, Modelling and Projections, can be found at www.epidem.org.

For the estimates: please visit www.aidsinfoonline.org.

UNAIDS

20 Avenue Appia
CH-1211 Geneva 27
Switzerland

+41 22 791 3666

unaids.org