Measuring the HIV/AIDS Epidemic: Approaches and Challenges

Ron Brookmeyer*

* Correspondence to Dr. Ron Brookmeyer, Department of Biostatistics, School of Public Health, University of California, Los Angeles, Los Angeles, CA 90095 (e-mail: rbrookmeyer@ucla.edu).

Accepted for publication January 11, 2010.

In this article, the author reviews current approaches and methods for measuring the scope of the human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) epidemic and their strengths and weaknesses. In recent years, various public health agencies have revised statistical estimates of the scope of the HIV/AIDS pandemic. The author considers the reasons underlying these revisions. New sources of data for estimating HIV prevalence have become available, such as nationally representative probability-based surveys. New technologies such as biomarkers that indicate when persons became infected are now used to determine HIV incidence rates. The author summarizes the main sources of errors and problems with these and other approaches and discusses opportunities for improving their reliability. Changing methods and data sources present new challenges, because incidence and prevalence estimates produced at different points in time are not directly comparable with each other, which complicates assessment of time trends. The methodological changes help explain the changes in global statistics. As methods and data sources continue to improve, the development of statistical tools for better assessing the extent to which changes in HIV/AIDS statistics can be attributed to changes in methodology versus real changes in the underlying epidemic is an important challenge.

acquired immunodeficiency syndrome; biological markers; epidemiologic methods; HIV; incidence; prevalence; statistics

INTRODUCTION

Statistical measures of the scope of the human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) pandemic have been revised numerous times by public health agencies. At the end of 2007, the Joint United Nations Program on HIV/AIDS (UNAIDS) decreased its estimate of the numbers of people worldwide living with HIV infection by over 6 million, to 33.2 million (1). UNAIDS also decreased its estimate of the worldwide annual rate of new HIV infections (HIV incidence) by 42%, to 2.5 million. In contrast, the Centers for Disease Control and Prevention (CDC) increased its long-standing estimate of annual HIV incidence in the United States by 40% (2). The changing statistics have been a source of confusion to the scientific community, policy-makers, and the general public. Do the changing statistics mean that the epidemic is growing, waning, or neither?

The confusion has been further fueled by controversial questions raised about the motivation for the statistical revisions. Dr. James Chin suggested in his book *The AIDS Epidemic: The Collision of Epidemiology with Political Correctness* that there has been a tendency to overestimate the scope of the epidemic and that overestimates rather than underestimates have been deliberately chosen for advocacy purposes (3). Sound public health policy relies on accurate statistics about the current status of and trends in the HIV/AIDS pandemic. Local, national, and international statistics aid officials in designing and targeting public health interventions, guide the allocation of resources, and help investigators gauge whether progress is being made in controlling the epidemic.

Statistical approaches for measuring the HIV/AIDS epidemic have evolved as the underlying data sources have grown and become more sophisticated. In the early days of the epidemic, the only sources of data available for
assessing the epidemic’s scope were national AIDS surveillance registries, which evolved in the 1980s, principally in North America, Europe, Australia, and other developed regions. The earliest projections of the epidemic were based on simple extrapolation of trends in the numbers of incident AIDS cases reported to these registries. The accuracy of this approach has had mixed success and serves as a cautionary reminder about the dangers and potential pitfalls of using extrapolation techniques to track public health problems (4). Fortunately, over the past 3 decades, the underlying data sources and methods available for quantifying the HIV/AIDS epidemic have greatly expanded, presenting both challenges and opportunities. My objective in this paper is to review the strengths and weaknesses of current methods and data sources for measuring the HIV/AIDS pandemic and to increase understanding of how changes in these methods help explain the changes in global statistics.

HIV PREVALENCE

HIV prevalence is the percentage of the population living with HIV infection. UNAIDS estimates that approximately one-half of 1% (0.5%) of the world’s population is currently living with HIV infection (1). Several approaches are currently used to obtain national statistics on HIV prevalence, including: national (probability-based) surveys; “sentinel surveillance data” collected among persons being treated at public health facilities, such as pregnant women receiving care at public antenatal clinics; specialized surveys of selected high-risk subpopulations (e.g., sex workers); and back-calculation methods that use data from national HIV/AIDS surveillance registries.

The strengths and weaknesses of each approach depend on the characteristics of the epidemic in a given country. For example, while national surveys may appear to be the most reliable approach, selection biases may operate in some countries to create serious challenges for obtaining representative samples. UNAIDS developed a typology that classifies countries into one of 3 categories in order to guide the choice of approaches for estimating HIV prevalence (3). Generalized epidemics are defined as those in which HIV infection is firmly established in the population and in which sexual networking in the general population is sufficient to sustain the epidemic. Operationally, generalized epidemics are defined as those where the HIV prevalence among pregnant women is consistently greater than 1%. Concentrated epidemics are those in which infection is firmly established in at least 1 subpopulation but not in the general population. Operationally, concentrated epidemics are defined as those in which HIV prevalence is consistently over 5% in 1 subpopulation but is below 1% among pregnant women in urban areas. Low-level epidemics are those in which the epidemic has not spread to a significant degree in any subpopulation. Operationally, low-level epidemics are defined as those in which HIV prevalence both has never been consistently greater than 5% in any subpopulation and is below 1% among pregnant women in urban areas. In point of fact, classification of a country into a single category in the UNAIDS typology is by no means clear and unambiguous because of sampling and nonsampling errors in the HIV subpopulation prevalence rates that are used for classification. Countries may also exhibit features of more than 1 category.

In order to estimate HIV prevalence in countries classified as having generalized epidemics, UNAIDS relies on antenatal clinic data and, to the extent that they are available, data from nationally representative HIV prevalence surveys. In order to estimate HIV prevalence in countries classified with concentrated and low-level epidemics, UNAIDS relies on specialized surveys of high-risk groups. The CDC uses a combination of methods to estimate HIV prevalence in the United States, including back-calculation. Below, I consider the strengths and limitations of each approach.

Nationally representative (probability-based) surveys

Nationally representative (probability-based) surveys of HIV prevalence typically employ stratified multistage cluster sampling of the general household population, which requires delineation of the sampling frame and ascertainment of the sampling probabilities of inclusion in a survey. One of the first attempts to conduct a nationally representative HIV prevalence survey was the Third National Health and Nutrition Examination Survey, conducted in the United States in 1988–1994 (6, 7). Nationally representative household-based surveys, such as the Demographic and Health Surveys containing an HIV/AIDS component, have now been performed in over 30 countries throughout the world, including countries in Africa, Asia, Latin America, and Eastern Europe (8, 9). Table 1 summarizes the results from some recently completed nationally representative surveys of HIV prevalence in adult populations (8–12). HIV prevalence ranged from 25.9% in Swaziland to 0.28% in India. UNAIDS now relies on these surveys as the basis for national HIV prevalence statistics in those countries in which they are available.

It is tempting to view nationally representative (probability-based) surveys as the gold standard for estimating national HIV prevalence rates (13); however, such surveys are subject to important potential sources of error. These uncertainties arise from challenges involved in obtaining representative samples. For example, institutionalized populations or populations that do not contain stable households may not be included in the sampling frame of a household-based population survey. Persons at higher risk of HIV infection, such as injecting drug users, men who have sex with men, sex workers, or other highly mobile persons, may be less likely to be present at the time of the survey (14). Even if individuals are selected for sampling and are present at the time of the survey, they may choose not to participate because of concerns about confidentiality and social stigmatization. Nonresponse bias, which occurs when persons selected for inclusion in a survey are either not present or refuse to participate, has the potential to significantly skew results.

Investigators have attempted to address the magnitude of nonresponse bias in surveys of national HIV prevalence rates. Mishra et al. (12, 15) compared nonresponders with responders with respect to some demographic data obtained.
from household questionnaires (of both respondents and nonrespondents) in 14 national Demographic and Health and HIV/AIDS Indicator surveys. They were able to examine some behavioral characteristics for those nonrespondents who completed an interview even though they did not complete an HIV test. These investigators concluded that nonresponders were indeed at higher “risk” of HIV infection based on their demographic and behavioral profiles but that the overall effect of nonresponse on national HIV prevalence estimates was small. A limitation of this study is that Mishra et al. could only partially assess the impact of nonresponse bias, because they did not have data on the actual HIV status of the nonrespondents (12, 15). They assumed that, conditional on some demographic and behavioral variables, the HIV prevalence rates of nonrespondents were the same as those of the respondents, and thus they could only measure bias reflected in imbalances between responders and nonresponders in the demographic and behavioral variables that were measured and on which data were collected.

Residual nonresponse bias refers to nonresponse bias that is not reflected in the measured demographic and behavioral variables. Studies that can assess residual nonresponse bias are relatively rare, because they require knowledge of the HIV status of the nonrespondents. In 1 study, Hull et al. (16) determined the HIV status of nonresponders by examining previously stored sera from patients attending a sexually transmitted disease clinic (testing was blinded so that HIV results could not be linked to individual patients). The researchers found that after controlling for demographic and behavioral characteristics, nonresponders had higher HIV prevalence rates than responders; among men who had sex with men, the HIV prevalence rate was 7.4 times higher among nonresponders than among responders, suggesting considerable residual nonresponse bias. Although it is uncertain how relevant that finding is to national surveys currently being conducted in developing countries, it serves as a cautionary warning about the potential of residual nonresponse bias to distort HIV prevalence estimates. Reniers and Eaton (17) documented “refusal bias,” wherein persons with prior knowledge of their HIV status are less likely to participate in an HIV prevalence survey if they are HIV-positive than if they are HIV-negative; this may explain 1 source of residual nonresponse bias.

How high must the response rate be to exclude the possibility of nonresponse bias significantly skewing the results? The ratio \( \gamma \) of the true HIV prevalence in the population to the prevalence among the responders (4) is

\[
\gamma = \frac{P_{\text{true}}}{P_{\text{responders}}} = 1 + (1 - f)(R - 1),
\]

where \( f \) is the fraction of those surveyed who respond (agree to an HIV test) and \( R \) is the ratio of HIV prevalence rates in nonresponders to those in responders. For example, if the response rate is 90% and \( R = 8 \), then \( \gamma = 1.7 \); that is, the true HIV prevalence would be 70% greater than that estimated from a survey. The take-home message is that significant bias could be present even if the response rate is high, if it

### Table 1. National Adult\(^a\) Prevalences of Human Immunodeficiency Virus Infection From Representative (Probability-Based) Surveys\(^b\)

<table>
<thead>
<tr>
<th>Region and Country</th>
<th>Year</th>
<th>Sample Size(^c)</th>
<th>Response Rate(^d), %</th>
<th>HIV Prevalence, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central/West Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benin</td>
<td>2006</td>
<td>12,185</td>
<td>80.3</td>
<td>1.20</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2003</td>
<td>8,559</td>
<td>89.1</td>
<td>1.80</td>
</tr>
<tr>
<td>Cameroon</td>
<td>2004</td>
<td>11,379</td>
<td>91.0</td>
<td>5.50</td>
</tr>
<tr>
<td>Côte d’Ivoire</td>
<td>2005</td>
<td>10,920</td>
<td>77.3</td>
<td>4.70</td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>2007</td>
<td>10,112</td>
<td>89.1</td>
<td>1.30</td>
</tr>
<tr>
<td>Ghana</td>
<td>2003</td>
<td>11,294</td>
<td>84.7</td>
<td>2.20</td>
</tr>
<tr>
<td>Guinea</td>
<td>2005</td>
<td>7,549</td>
<td>90.6</td>
<td>1.50</td>
</tr>
<tr>
<td>Liberia</td>
<td>2007</td>
<td>13,924</td>
<td>84.5</td>
<td>1.50</td>
</tr>
<tr>
<td>Mali</td>
<td>2006</td>
<td>8,618</td>
<td>80.4</td>
<td>0.50</td>
</tr>
<tr>
<td>Niger</td>
<td>2006</td>
<td>8,738</td>
<td>89.0</td>
<td>0.70</td>
</tr>
<tr>
<td>Senegal</td>
<td>2005</td>
<td>9,725</td>
<td>80.4</td>
<td>0.70</td>
</tr>
<tr>
<td>East Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>2005</td>
<td>13,920</td>
<td>79.3</td>
<td>1.40</td>
</tr>
<tr>
<td>Kenya</td>
<td>2003</td>
<td>8,486</td>
<td>73.3</td>
<td>6.70</td>
</tr>
<tr>
<td>Rwanda</td>
<td>2005</td>
<td>10,796</td>
<td>96.5</td>
<td>3.00</td>
</tr>
<tr>
<td>Tanzania</td>
<td>2004</td>
<td>13,350</td>
<td>80.3</td>
<td>7.00</td>
</tr>
<tr>
<td>Uganda</td>
<td>2004</td>
<td>21,359</td>
<td>86.6</td>
<td>6.40</td>
</tr>
<tr>
<td>Southern Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Botswana</td>
<td>2004</td>
<td>24,756</td>
<td>61.0</td>
<td>17.1</td>
</tr>
<tr>
<td>Lesotho</td>
<td>2004</td>
<td>7,063</td>
<td>74.4</td>
<td>23.5</td>
</tr>
<tr>
<td>Malawi</td>
<td>2004</td>
<td>7,868</td>
<td>66.9</td>
<td>11.8</td>
</tr>
<tr>
<td>South Africa</td>
<td>2005</td>
<td>13,215</td>
<td>70.0</td>
<td>16.2</td>
</tr>
<tr>
<td>Swaziland</td>
<td>2007</td>
<td>9,976</td>
<td>82.7</td>
<td>25.9</td>
</tr>
<tr>
<td>Zambia</td>
<td>2002</td>
<td>5,107</td>
<td>76.4</td>
<td>15.6</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>2006</td>
<td>18,631</td>
<td>69.7</td>
<td>18.1</td>
</tr>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambodia</td>
<td>2005</td>
<td>15,867</td>
<td>92.7</td>
<td>0.60</td>
</tr>
<tr>
<td>India</td>
<td>2006</td>
<td>126,357</td>
<td>81.6</td>
<td>0.28</td>
</tr>
<tr>
<td>Caribbean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>2007</td>
<td>26,970</td>
<td>85.0</td>
<td>0.80</td>
</tr>
<tr>
<td>Haiti</td>
<td>2005</td>
<td>10,462</td>
<td>96.6</td>
<td>2.20</td>
</tr>
<tr>
<td>North America</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>2006</td>
<td>12,917</td>
<td>92.0</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Abbreviations: AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus.

a Age ranges for the cited data were 15–49 years for females and 15–59 years for males, with the following exceptions: India, Kenya, and Zimbabwe, 15–54 years for males; Liberia, South Africa, Swaziland, and the United States, 15–49 years for both males and females; Benin, 15–64 years for males; and Botswana, greater than 18 months.

b Surveys included the Demographic and Health Surveys (7–9), the National Health and Nutrition Examination Survey in the United States (5, 6), the South African National HIV Prevalence, HIV Incidence, Behavior and Communication Survey in South Africa (10), and AIDS Impact Survey II in Botswana (11).

c Total number of persons eligible for HIV testing (9) or, for the United States, the total number of persons who completed an interview.

 d Percentage of eligible persons who completed an HIV test.
also turns out that $R$ is large. While the average response rate in the nationally representative surveys that have been conducted is over 80% (Table 1), without quantitative measurements of $R$ we cannot fully assess the impact of nonresponse bias, and thus studies are needed to help quantify the magnitude of $R$.

**Antenatal clinic surveys**

HIV prevalence surveys conducted at public health facilities providing antenatal care for pregnant women have been an important source of information about HIV prevalence, especially in Africa. UNAIDS has essentially relied on antenatal care data for estimating national HIV prevalence in those countries it has classified as having generalized epidemics, if nationally representative survey data are not available (1, 18, 19). The underlying assumption is that HIV levels among pregnant women receiving care at antenatal care clinics are surrogates for HIV levels in the general population.

There are, of course, important potential biases with the use of antenatal care surveys to estimate national HIV prevalence rates. The surveys represent only pregnant women, who by definition have been sexually active and are of reproductive age, and they do not represent men. Even if HIV levels among pregnant women were a surrogate for HIV prevalence in the general population, samples from public antenatal care sentinel surveillance sites are not random samples of pregnant women in a given country. Some pregnant women may receive antenatal care outside of the public sector or not at all. Antenatal care coverage is not representative of all geographic areas within a country, and it is often concentrated in urban areas.

Gouws et al. (19) compared HIV prevalence rates from antenatal care surveillance surveys with rates from nationally representative (probability-based) surveys for 26 countries in which both data sources were available. They concluded that HIV prevalence from antenatal clinic data overestimates that from nationally representative surveys by approximately 20% in both urban and rural areas. An analysis by Montana et al. (20) suggested that the reason antenatal care rates overestimate rates from population-based surveys is because the antenatal care catchment areas are weighted to geographic regions with higher HIV rates; they found that the rates are concordant if the samples from the population-based surveys are restricted to the antenatal care catchment areas. On the basis of these studies, UNAIDS recommended that antenatal care prevalence rates be adjusted downwards by a factor of 0.8 to estimate the rate in the general population in both urban and rural areas in countries with generalized HIV epidemics (1). Prior to 2007, UNAIDS only adjusted downwards antenatal care data from rural areas (1). These new adjustment factors for antenatal care data are the principal reasons why UNAIDS lowered estimates of national HIV prevalence in a number of countries (including Angola, Congo, Eritrea, the Gambia, Guinea-Bissau, Mozambique, Namibia, Nigeria, Somalia, and Sudan).

An important question is whether reliable inferences about time trends in population HIV prevalence can be drawn from trends in antenatal care data. It is plausible, in at least some countries, that antenatal care participation rates have changed, perhaps as a result of public health interventions, including increased availability of antiretroviral therapy. Some countries have documented decreases in HIV prevalence from antenatal care data (21), but it remains unclear to what extent these decreases reflect real decreases in population HIV prevalence or instead result from changes in the populations represented by antenatal care surveillance data.

Several alternatives to antenatal care surveillance have been suggested by some researchers to address weaknesses in antenatal care data. Use of data from routine voluntary HIV counseling and testing centers has been proposed for estimating population HIV prevalence because, unlike antenatal care data, the target population extends beyond child-bearing women (22). Use of data from programs designed to prevent mother-to-child transmission has also been proposed for estimating HIV prevalence; in contrast to antenatal care data, such data are generally neither anonymous nor unlinked, and specimen collection is not restricted to the first clinic visit (23). It remains to be seen whether these alternative surveillance data sources, which are also subject to biases, are any more or less accurate than antenatal care data for estimating point prevalence or time trends in population prevalence.

**Combining surveys in high-risk subpopulations**

In many countries, HIV/AIDS is concentrated in subpopulations that are difficult to adequately study in household-based national surveys (e.g., injecting drug users, sex workers, clients of sex workers, and men who have sex with men)—what UNAIDS terms “concentrated epidemics.” An alternative approach to a national household survey is to stratify the population into risk groups and then estimate the HIV prevalence rate for each risk group (e.g., sex workers) using specialized surveys. If $p_i$ is the HIV prevalence rate in subgroup $i$, then the overall national prevalence rate is

$$P = \sum_{i=1}^{k} w_i p_i,$$

where $w_i$ is the proportion of the national population in subgroup $i$.

The CDC first used this approach to estimate HIV prevalence in the United States (24). UNAIDS calls this “the workbook approach” and recommends its use in countries with low-level and concentrated epidemics (25). There are several sources of uncertainty with this approach. The first concern is identification of the risk groups that adequately subdivide a country’s population into mutually exclusive subpopulations. A second concern is the difficulty involved in obtaining representative samples of high-risk subpopulations (e.g., clients of sex workers), which are needed to estimate $p_i$. A third concern, and perhaps the most challenging issue, centers around the large error resulting from uncertainties in the fractions of the population ($w_i$) in each subgroup population; for example, the numbers of intravenous drug users or sex workers in a country are generally not known.
with a reasonable degree of accuracy. Various approaches, including capture-recapture methods, have been proposed to measure the size of high-risk subpopulations (26).

**Back-calculation methods**

Back-calculation uses HIV/AIDS surveillance data collected by national registries to reconstruct the historical pattern of HIV incidence (4). The basic idea of the method is to use the incubation period distribution to “work backwards” and infer the numbers of persons infected in previous years that would reproduce the observed AIDS surveillance data. The key idea is that AIDS is the tip of the iceberg, because many HIV infections are still incubating.

The original version of back-calculation combined AIDS surveillance data with data on the incubation period in a statistical analysis using the relation

$$\text{AIDS diagnosis date} = \text{HIV infection date} + \text{incubation period}.$$  

Using AIDS diagnosis data (the left-hand side of the equation) together with information about the incubation period (the far-right side of the equation), researchers developed statistical deconvolution methods to infer historical HIV infection dates and thus HIV incidence trends. The method of back-calculation is a lens that allows one to look backwards in time to estimate historical HIV incidence rates. Estimates of HIV incidence in the very recent past are considerably less reliable than estimates of incidence in the distant past, because very few recent infections are reflected in AIDS surveillance data due to the long incubation period.

A number of critical challenges for the back-calculation method emerged in the 1990s: It could only be employed in countries with fairly complete reporting and counting of AIDS cases; the incubation period distribution was changing because of widespread use of antiretroviral drugs; the AIDS surveillance definition changed; and registries began expanding their activities to include persons with asymptomatic HIV infection. Back-calculation methods were extended to address some of these challenges, particularly to use surveillance data on identified persons with HIV infection who may not have an AIDS diagnosis (2, 27–30). These extensions of the original back-calculation methods depend not only on assumptions about the incubation period distribution but also on the trends in and intensity of HIV screening in a population. Such extended back-calculation methods were used recently to assess the scope of the epidemic in the United States (2).

**HIV INCIDENCE**

**HIV incidence** is the annual number of new infections that occur in a population. The ratio of HIV incidence to the size of the uninfected population is the HIV incidence rate, which is the risk of becoming infected per unit of time. While HIV prevalence measures overall disease burden, HIV incidence tracks the leading edge of the epidemic—the growth of new infections.

Estimation of current HIV incidence is a more challenging undertaking than estimation of HIV prevalence. The main approaches used to estimate current HIV incidence include: cohort studies; cross-sectional surveys that use biomarker assays of recent infection; and statistical estimation from changes in HIV prevalence. UNAIDS has relied on the latter approach, principally by analyzing time trends in HIV prevalence surveys of antenatal clinics. The CDC currently relies principally on the biomarker approach. Below, I review the main ideas and key assumptions of each of the methods.

**Cohort studies**

Cohort studies of uninfected persons have been widely employed to estimate HIV incidence rates; for example, see Karon et al. (31) for a summary of rates from various populations in the United States. While cohort studies are useful for estimating incidence in select subpopulations, public health agencies have not generally relied on them for obtaining national estimates of HIV incidence, because of the costs and logistical difficulties involved in following representative national samples of adequate size.

There are important sources of error in cohort studies that could bias incidence rates, even in selected subpopulations. One error, selection bias, arises if persons who agree to participate and return for follow-up visits are not representative of the target population. Another possible source of error is that follow-up visits could affect the HIV incidence rate through repeated exposure to counseling (such as the promotion of condom use or safe sex or other prevention messages). This phenomenon has been called the “adherence effect,” because it occurs among persons who adhere to the scheduled follow-up visits (32). For example, in the ZVITAMBO study, Piwoz et al. (33) reported increasing HIV knowledge with increasing exposure to an HIV education and counseling program in a cohort of new mothers in Zimbabwe.

**Serial prevalence surveys**

The basic idea of serial prevalence surveys is to infer HIV incidence (annual numbers of new infections in a population) from changes in absolute HIV prevalence (“absolute HIV prevalence” refers here to the numbers of persons living with HIV rather than the percentage infected) at 2 or more points in time. The HIV prevalences $N_1$ and $N_2$ at times $t_1$ and $t_2$, respectively, are related through the balancing equation

$$N_2 = N_1 + m - d + r - s.$$

(2)

Here, $m$ is the number of incident HIV infections that occurred between $t_1$ and $t_2$ among persons in the population at time $t_1$; $d$ is the number of deaths that occurred between $t_1$ and $t_2$ among persons who were HIV-infected in the population at time $t_1$; and $r$ and $s$ are, respectively, the numbers of in-migrations and out-migrations of HIV-infected persons between $t_1$ and $t_2$ (also included in $r$ and $s$ are persons who “age in” or “age out” of the age range of the target
population. In order to estimate HIV incidence from the balancing equation, the investigator requires accurate information about mortality ($d$) and migration changes ($r$ and $s$) in the HIV-infected population.

It is instructive to consider equation 2 in the steady-state situation: Suppose the number of deaths in the HIV-infected population is equal to the number of incident infections ($d = m$) and that the number of in-migrations is equal to the number of out-migrations ($r = s$). Then, under those steady-state conditions, absolute HIV prevalence will appear to be constant ($N_1 = N_2$), even though HIV incidence ($m$) could be high. The cautionary warning here is that stable or even falling HIV prevalence does not imply that HIV incidence rates are zero. Accordingly, in deriving HIV incidence from changes in prevalence (equation 2), investigators must carefully assess the effects of mortality and migration: If mortality ($d$) is overestimated (underestimated), HIV incidence ($m$) will be correspondingly overestimated (underestimated).

Suppose the HIV prevalences in 2 successive cross-sectional surveys conducted $\delta = t_2 - t_1$ years apart are $P_1$ and $P_2$. If we can ignore migration effects, that is, $r = s$, it can be shown that the incidence rate $I$ is

$$I = \frac{P_2 - P_1 R}{(1 - P_1)\delta} \times 100\% \text{ per year,} \quad (3)$$

where $R$, the relative survival rate, is the ratio of the survival probabilities for the HIV-infected population to those for the general population over $\delta$ years. To illustrate the calculations, suppose the HIV prevalence rates were 0.10 in an initial survey and 0.14 in a survey conducted 5 years later. Suppose further that the 5-year survival rates in the HIV-infected and uninfected populations were 0.85 and 0.95, respectively, and thus the relative survival is 0.89 (0.85/0.95 = 0.89). Then the estimated incidence rate from equation 3 is ($0.14 - 0.10 \times 0.89)/100/5(1 - 0.10) = 1.1%$ per year.

The approach of estimating HIV incidence from serial prevalence surveys may find increased utility, because surveys such as the Demographic and Health Surveys are planned to be repeated in some countries every 5 years (8). However, an important drawback of this approach is that it requires critical inputs about mortality and migration, which can be very uncertain. Furthermore, nationally representative surveys realistically can only be conducted once every several years, and as such the approach is unlikely to be able to provide timely information about current trends in HIV incidence.

UNAIDS uses a variation of this approach to derive HIV incidence by analyzing trends in prevalence rates at antenatal clinics (34). The UNAIDS approach consists of fitting a smooth parametric curve to the time series of prevalence rates using a computer software program, the Estimation and Projection Package (35). HIV incidence rates are calculated using additional mortality and demographic assumptions implemented in another software program, SPECTRUM (36). There are several uncertainties in this approach. First, as was discussed above, prevalence trends among pregnant women attending public antenatal clinics may not be a useful surrogate for the general population. Second, mortality and migration assumptions are uncertain. Third, the curve-fitting technique used by UNAIDS to smooth the time series of antenatal care HIV prevalence rates uses a particular mathematical model that allows only a single peak (or mode) in the time trend. Accordingly, the mathematical model currently used by UNAIDS to smooth the noisy time series of antenatal care prevalence rates does not have the flexibility to detect reemerging sharp upturns in HIV incidence. Some Bayesian approaches have been suggested to account for model uncertainty (37). Some related statistical approaches to inferring HIV incidence from trends in HIV prevalence have also been discussed by Williams et al. (38) and Hallett et al. (39).

Biomarkers in cross-sectional surveys

The basic idea of the biomarker approach to estimating HIV incidence rates is to use a biomarker to identify persons who were recently infected (40). This approach requires a cross-sectional survey of a representative sample of persons in the population from whom serum specimens were collected at a single point in time, which is in contrast to the cohort approach, which requires that multiple specimens be collected longitudinally over time. The biomarker is used to identify persons who are in the “window” period—a period of time shortly after incident infection occurs.

The biomarker approach relies on the epidemiologic relation that the prevalence of a condition is equal to the incidence multiplied by the mean duration of the condition. Here, the “condition” refers to the window period. Then the incidence rate $I$ is estimated from the equation

$$I \approx \frac{P}{\mu} \times 100, \quad (4)$$

where $P$ is the proportion of persons in the window period among all persons who are either uninfected or in the window period and $\mu$ is the mean window period. If the mean is expressed in years (days), then $I$ is expressed as the percentage per year (per day) of the uninfected population that becomes infected. The duration of the window period is not fixed but rather is random and has a probability distribution. Individuals may have window periods either above or below the mean window period.

To illustrate the calculations with equation 4, suppose that in a cross-sectional survey sample of 10,000 persons, we find that 2,274 are HIV-positive and 7,726 are HIV-negative on the standard HIV antibody test. The 2,274 HIV-positive persons are then evaluated with a second assay (such as the BED assay) for a biomarker to determine which persons are in the window period. Suppose 201 persons are found to be in the window period. If the mean window period based on these assays is 158 days (0.43 years), the HIV incidence rate is $201/(201 + 7,726) \times (1/0.43) \times 100 = 5.9%$ per year.

The biomarker approach was first suggested using an assay for P24 antigenemia (40). That biomarker necessitates large sample sizes to obtain statistical precision because the durations of time (window periods) in which HIV-infected persons are P24 antigenemic are relatively short.
Subsequently, the Serologic Testing Algorithm for Recent HIV Seroconversion (STARHS) was developed (41). This approach uses a dual-antibody testing system whereby persons are tested with a standard HIV antibody test and those who test positive are tested with a second assay in an attempt to distinguish recent infections. Persons who are positive on the first assay and are identified by the second assay as having a “recently” occurring infection are said to be in the window period. An advantage of the STARHS approach is that only persons who are HIV-positive need to be tested with the second assay. Several classes of assays for the second assay have been proposed, including those based on the quantity of HIV antibodies (the detuned assay (41)), the proportion of HIV antibodies (the BED assay (42)), and the avidity of HIV antibodies (43, 44). The “detuned assay” is based on the absolute titer of HIV-1-specific antibody and was one of the first antibody-based assays developed for the purpose of estimating HIV incidence (41). The BED HIV-1 immunoglobulin G capture enzyme immunoassay was subsequently developed and is based on the proportion of immunoglobulin G that is directed against HIV immunoglobulin G antigens B, E, and D. The BED assay is thought to be more robust and less sensitive to HIV-1 subtype variation than the original detuned assay. Individuals are defined to be in the window period if they are positive on the standard enzyme immunoassay (the first assay) and negative on the second assay (e.g., either the BED assay or the detuned assay). A positive BED assay result has been defined as one that is above the optical density cutoff of 0.8 (2, 42).

The mean window period depends on the biomarker assays that are used, and it may also depend on HIV subtype and the immunologic characteristics of the population. The mean window period based on the BED assay for subtypes B and E is approximately 158 days (2). An analysis of subtype C suggested a larger mean (45).

One important issue with antibody-based assays for recent infection, such as the BED assay, is that persons with advanced HIV disease (AIDS) will tend to be classified as having recent infections because of associated declines in anti-HIV antibody levels. Furthermore, persons receiving antiretroviral therapy may also be classified as having “recent” infection by some of these assays (44, 46). This has led to the recommendation that it is necessary to exclude persons with AIDS or persons on antiretroviral therapy from being counted in the window period (46).

If the HIV incidence rate has remained approximately constant for a period of time and that period is big enough to include all possible window periods, the biomarker approach produces an unbiased estimate of the HIV incidence rate. If, however, the HIV incidence rate has not remained constant, the biomarker approach is not estimating the current instantaneous HIV incidence rate but rather a time-weighted average of HIV incidence over the recent past. The time-weighted average extends back in time over a period that depends on the distribution of window periods. Specifically, it has been shown that the weighting function in the time-weighted average is the backward recurrence time density of the window period distribution (47). Accordingly there is a statistical tradeoff between assays that have large window periods and those that have small window periods. Assays with large window periods generally will have smaller standard errors. However, assays with smaller window periods will produce a more current (unbiased) estimate of the HIV incidence rate. It is the classic tradeoff of bias versus variance.

The CDC has used the biomarker approach to determine HIV incidence in the United States (2, 48). Specimens from persons who were voluntarily tested and found to be infected within a 22-state surveillance catchment area were tested with the BED assay. The (absolute) HIV incidence in the population was calculated as the ratio of the (absolute) number in the window period to the mean duration of the window period. The resulting figure was then adjusted for the following concerns: 1) the surveillance catchment area covered only 22 of the 50 states; 2) only about 17% of the HIV-positive specimens collected in the surveillance system were tested with the BED assay; and 3) not all persons who become HIV-infected submit to an HIV test. In the analysis by Hall et al. (2), the investigators’ conclusion was that 56,300 persons were infected in the United States in 2006. An important source of bias associated with biomarker-based estimation of HIV incidence within a national surveillance system is that the samples that are tested may not be representative of the population. For example, Remis and Palmer (49) quantified the bias associated with HIV test-seeking behaviors and found that a biomarker-based estimate of incidence can be biased because of increased early testing among recently infected persons. In such cases, weighting of the sample and associated statistical adjustments may help investigators adjust the rates to increase representativeness to more accurately reflect the target population (for an example of such adjustments, see Karon et al. (48)).

The biomarker approach has been used in both developed and developing countries to estimate HIV incidence. However, there have been reports, especially from Africa, that cohort estimates of incidence are lower than biomarker (BED) estimates, raising the concern that the BED approach systematically overestimates HIV incidence (50). In fact, UNAIDS issued a cautionary statement about the use of the BED assay to estimate HIV incidence (51). These reports raise questions as to why cohort estimates do not agree with the biomarker estimates.

There are a number of potential sources of error with the biomarker approach that could help explain discrepancies with cohort studies (32). The cross-sectional samples may not be representative of the target population if, for example, specimens are drawn from antenatal care clinics, sexually transmitted disease clinics, or voluntary testing and counseling centers. Another important source of bias in a biomarker estimate of incidence arises if the mean window period used in equation 4 is incorrect. The mean window may be sensitive to the tail of the distribution of window period lengths, and there have been reports of long window periods. Confidence interval procedures that account for uncertainty in the mean window period have been developed, including an analytic approach (52) and a Monte Carlo approach (53). The mean window period may vary by HIV subtype or the immune status of the population (chronic infection can distort BED results). Persons with
advanced HIV disease (AIDS) or persons who have taken antiretroviral agents should be excluded from the cross-sectional samples prior to calculation of incidence rates. Laboratory quality control problems may also be a source of error.

Several statistical adjustment procedures for the biomarker approach have been proposed to account for so-called misclassification error with respect to the timing of infection (45, 54). The original motivation for the development of these adjustments was to lower biomarker estimates to bring them in line with cohort estimates, particularly in Africa. These adjustment procedures attempt to account for so-called misclassification errors that arise if, for example, persons in the window period in fact had “long-standing infections.” However, these procedures have not been universally adopted (2). Investigations into the validity of these statistical adjustment procedures suggest the following conclusions. If patients eventually progress through the window period (i.e., they do not remain permanently in the window), no further adjustment to equation 4 is necessary to account for “misclassification” (32, 55–58). If, however, a fraction of patients are “assay nonprogressors” and remain in the window period permanently and indefinitely, additional statistical adjustment is required (56–58). Some studies have suggested that there is a subpopulation of elite HIV suppressors who remain in the window period without ever progressing to AIDS; this can then distort HIV incidence estimates from cross-sectional studies if such persons are not adequately taken into account (59). The proportion of elite suppressors who remain in the window period permanently without ever progressing to AIDS or receiving antiretroviral agents is an unsettled question at this time. In any case, the reliability of biomarker-based estimates of HIV incidence could be improved by 1) excluding elite suppressors through RNA polymerase chain reaction testing (60), 2) better estimating the mean window period across HIV subtypes and diverse populations and settings, 3) excluding persons who have AIDS or are using antiretroviral agents, and 4) improving the representativeness of the cross-sectional samples of the target population. Algorithms based on multiple assays that include BED, CD4-cell, and RNA polymerase chain reaction testing appear to offer a promising approach to improving the accuracy of cross-sectional estimates of HIV incidence (59, 60). Because there are important sources of error with both cohort and biomarker-based HIV incidence estimates, neither cohort nor biomarker estimates should be blindly viewed as the gold standard for assessing the validity of the other.

**PERSPECTIVES**

There have been significant advances in approaches for estimating HIV prevalence (61), but there still remains considerable uncertainty. Table 2 summarizes those sources of uncertainty and issues associated with the estimation of HIV prevalence discussed above. Estimating current HIV incidence rates presents even greater challenges than estimating prevalence. Table 3 summarizes issues associated with estimation of current HIV incidence.

UNAIDS reduced its estimate of worldwide HIV prevalence by over 6 million persons in 2007 (1). An important reason for that revision was changes in the methods used to derive the prevalence statistics (18). Those changes include reliance on newly available nationally representative HIV prevalence surveys instead of surveys of pregnant women attending public antenatal clinics in many countries classified as having generalized epidemics. Nationally representative HIV prevalence surveys were nearly nonexistent in developing countries prior to 2000, while now over 30 countries have them. Reliance on national prevalence surveys rather than antenatal care data accounts for over 70% of the decline in worldwide HIV prevalence from 2006 to
Infection is a powerful opportunity for tracking the epidemic. First is the United States. An increase in the rate of occurrence of new infections in the changes rather than any direct evidence of a recent real increase in reported US HIV incidence results from methodological approaches that had not been used previously. The 40% increase in reported HIV incidence—an approach that had been used previously. Those adjustments are based on a comparison of antenatal clinics and national surveys in the subset of countries in which both are available, but the extent to which those adjustments are applicable to other countries in which national surveys have not yet been conducted is unknown.

Why is HIV prevalence so hard to measure? In developing countries that are most affected by the epidemic, the infrastructure for disease surveillance is inadequate, and only recently have a large number of nationally representative HIV prevalence surveys been undertaken. New statistical adjustments to antenatal clinic prevalence rates are now used. Those adjustments are based on a comparison of antenatal clinics and national surveys in the subset of countries in which both are available, but the extent to which those adjustments are applicable to other countries in which national surveys have not yet been conducted is unknown.

The UNAIDS approach to estimating HIV incidence has relied principally on indirect estimation using trends in serial prevalence surveys from antenatal clinics. However, those estimates depend on assumptions about the survival rates of the HIV population. If longer survival assumptions are used, lower estimates of HIV incidence will be produced. In 2007, UNAIDS changed its assumption regarding the median survival of the HIV-infected population from 9 years to 11 years in many countries, which was the principal change driving the reduction in the UNAIDS estimate of worldwide HIV incidence from 4.1 million in 2006 to 2.5 million in 2007 (1). The US government now uses the biomarker (BED) approach to estimate HIV incidence—an approach that had not been used previously. The 40% increase in reported US HIV incidence results from methodological changes rather than any direct evidence of a recent real increase in the rate of occurrence of new infections in the United States.

There are on the horizon exciting opportunities to improve our ability to measure the HIV/AIDS epidemic. Two developments, if combined together, potentially offer a powerful opportunity for tracking the epidemic. First is the availability of nationally representative probability-based HIV prevalence surveys. Surveys such as the Demographic and Health Surveys were nearly nonexistent prior to 2000 but have now been performed in over 30 countries. Second is the development of assays for biomarkers to identify persons recently infected. The US government estimated HIV incidence (2) using biomarkers (the BED assay) in serum samples collected during routine surveillance, but special statistical adjustments were required to address issues of representativeness because the samples were not derived from a nationally representative survey. We now have an opportunity to obtain direct population-based estimates of HIV incidence in many countries by incorporating assays for biomarkers for recent infection in nationally representative HIV prevalence surveys such as the Demographic and Health Surveys. The approach is cost-effective because it requires assaying only those HIV-positive samples that are already being collected and identified in ongoing HIV prevalence surveys. Furthermore, incorporation of biomarkers for recent infection into 2 or more serial national HIV prevalence surveys (e.g., Demographic and Health Surveys) would be an opportunity to track changes in HIV incidence. Such studies could produce the first direct population assessments of changes in national HIV incidence rates.

The current BED assay is not perfect, and development of improved assays and assay properties, including precise quantification of the detectable window periods of recent infection for various populations, settings, and viral strains, should be priorities. Development of improved assays for recent infection (perhaps based on a combination of different assays) would be invaluable. Nationally representative HIV prevalence surveys that incorporate assays for biomarkers of recent infection offer a promising cost-effective opportunity to significantly improve our ability to measure the HIV epidemic in many countries and to answer the central but remarkably difficult question: Are we winning the HIV prevention war?

Another opportunity to improve HIV incidence estimation is to embed a follow-up mortality substudy within a national HIV prevalence survey such as the Demographic and Health Surveys. Such an approach would provide the mortality and

**Table 3.** Major Sources of Error In and Problems With Approaches for Estimating Current Incidence Rates of Human Immunodeficiency Virus Infection

<table>
<thead>
<tr>
<th>Cohort Studies</th>
<th>Serial HIV Prevalence Surveys</th>
<th>Biomarkers in Cross-Sectional Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representativeness (selection bias)</td>
<td>Representativeness</td>
<td>Representativeness</td>
</tr>
<tr>
<td>Low follow-up rates</td>
<td>Errors in assumed mortality rates</td>
<td>Mean window period (may depend on strain, population)</td>
</tr>
<tr>
<td>Adherence effects</td>
<td>Unaccounted-for migration effects</td>
<td>Laboratory errors</td>
</tr>
<tr>
<td>Expense</td>
<td>Model for prevalence curve&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Impact of advanced HIV disease and antiretroviral therapy</td>
</tr>
<tr>
<td></td>
<td>Time interval between surveys&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Expense</td>
<td></td>
<td>Window period distribution and assay nonprogressors</td>
</tr>
</tbody>
</table>

Abbreviation: HIV, human immunodeficiency virus.

<sup>a</sup> Especially problematic if based on a time series of serial prevalence surveys from sentinel populations, such as pregnant women attending antenatal care clinics.

<sup>b</sup> Especially problematic if based on large national prevalence surveys with at least several years between surveys.
migration information necessary to use equation 3 to estimate HIV incidence rates from 2 or more representative prevalence surveys. Specifically, samples of HIV-infected and uninfected persons would be followed for mortality, allowing estimation of the relative survival rate, a key input needed to estimate HIV incidence from equation 3. Such an approach could complement the biomarker-based estimates and provide an opportunity to corroborate findings.

We now have an impressive toolbox of statistical methods and data sources for measuring the HIV epidemic, which presents challenges for comparing statistics over time. Incidence and prevalence estimates produced at different time points using different methods are not directly comparable with each other, which complicates assessment of time trends. Tracking relative changes in key measures of the epidemic over time may be as important as or more important than any single estimate of prevalence or incidence at 1 point in time. Ultimately, development of effective disease control strategies depends not on a single statistic calculated at 1 time point but on an understanding of the complex interaction over time of the linked network of heterogeneous subepidemics around the world. As methods and data sources improve, it will be increasingly important for public health officials to effectively communicate to the public the extent to which changes in HIV/AIDS statistics reflect significant changes in the dynamics of the epidemic versus simply changes in the yardstick used to measure the epidemic.

ACKNOWLEDGMENTS

Author affiliations: Department of Biostatistics, School of Public Health, University of California, Los Angeles, Los Angeles, California; and Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland.

The author gratefully acknowledges the support of the Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey.

Conflict of interest: none declared.

REFERENCES

5. Macro International, Inc. HIV Notes From Measure DHS.


