# Estimating Incidence from Prevalence in Generalised HIV Epidemics: Methods and Validation

Timothy B. Hallett[1*], Basia Zaba[2,3], Jim Todd[4], Ben Lopman[1], Wambura Mwita[3], Sam Biraro[4], Simon Gregson[1,5], J. Ties Boerma[6], on behalf of the ALPHA Network

1 Imperial College London, London, United Kingdom, 2 London School of Hygiene and Tropical Medicine, London, United Kingdom, 3 National Institute for Medical Research, Mwanza, Tanzania, 4 Medical Research Council/Uganda Virus Research Institute, Uganda Research Unit on AIDS, Entebbe, Uganda, 5 Biomedical Research and Training Institute, Harare, Zimbabwe, 6 World Health Organization, Geneva, Switzerland

## ABSTRACT

### Background

HIV surveillance of generalised epidemics in Africa primarily relies on prevalence at antenatal clinics, but estimates of incidence in the general population would be more useful. Repeated cross-sectional measures of HIV prevalence are now becoming available for general populations in many countries, and we aim to develop and validate methods that use these data to estimate HIV incidence.

### Methods and Findings

Two methods were developed that decompose observed changes in prevalence between two serosurveys into the contributions of new infections and mortality. Method 1 uses cohort mortality rates, and method 2 uses information on survival after infection. The performance of these two methods was assessed using simulated data from a mathematical model and actual data from three community-based cohort studies in Africa. Comparison with simulated data indicated that these methods can accurately estimates incidence rates and changes in incidence in a variety of epidemic conditions. Method 1 is simple to implement but relies on locally appropriate mortality data, whilst method 2 can make use of the same survival distribution in a wide range of scenarios. The estimates from both methods are within the 95% confidence intervals of almost all actual measurements of HIV incidence in adults and young people, and the patterns of incidence over age are correctly captured.

### Conclusions

It is possible to estimate incidence from cross-sectional prevalence data with sufficient accuracy to monitor the HIV epidemic. Although these methods will theoretically work in any context, we have able to test them only in southern and eastern Africa, where HIV epidemics are mature and generalised. The choice of method will depend on the local availability of HIV mortality data.

*The Editors' Summary of this article follows the references.*

## Introduction

Monitoring the continuing spread of the HIV epidemic is essential for determining public health priorities, assessing the impact of interventions, and making estimates of current and future health care needs [1]. Currently, surveillance systems in generalized epidemics primarily rely on HIV prevalence (fraction of population infected) data collected from women attending selected antenatal clinics [2,3]. Interpretation of these data is complicated by natural epidemiological changes that arise from the long and variable incubation of HIV and AIDS-related mortality [4,5], by biases in the sample due to subfertility associated with bacterial sexually transmitted infections and HIV [6], and by the disproportionate selection of surveillance sites in urban areas [7]. Recently, serological testing has been included in household health surveys, such as the Demographic and Health Surveys (DHS), giving estimates of HIV prevalence in the general adult population based on a standard methodology [8–10]. Measures that relate to the general population are more useful, but there remain several important limitations in using prevalence data to monitor the epidemic. These limitations include the following. (i) Decreases in prevalence do not necessarily indicate a reduction in risk of infection [5]; (ii) changes in prevalence lag behind real changes in risk, particularly at older ages; (iii) comparisons of prevalence between countries can be confounded by different survival times following infection (e.g., if survival following infection is shorter in Asia than in Europe or Africa [11–13], then similar prevalence levels could mask higher incidence rates in Asia); and (iv) the weak association between prevalent infection and risk makes it difficult to identify "high-risk" groups (e.g., higher average prevalence among women does not necessarily mean they are at overall greater risk than men).

A better measure for monitoring the HIV epidemic is incidence (rate of new infections among those at risk). If incidence is known, temporal changes in the epidemic can be better identified and characterised, the pattern of risk over age can be examined to aid the understanding of transmission patterns and highlight targets for interventions, and future health care needs can be predicted. However, whereas prevalence can be measured with independent, anonymised, cross-sectional serosurveys, direct measurement of incidence would require individuals to be identified and followed up from one serosurvey round to the next. In most settings, achieving high follow-up rates is a major logistical challenge and is more expensive than selecting a new panel of participants at each survey round. New types of HIV test that can discriminate recent infections from a single specimen have not so far been able to generate reliable estimates of incidence [14,15]. Most household surveys do not follow up individuals, but it may still be possible to estimate incidence from prevalence estimates in successive surveys, without using complex projection packages. This paper describes and tests two methods for this form of incidence estimation, with the intention that they can be applied in future to sequential prevalence data collected in household surveys, including the DHS.

Several methods have been proposed to estimate incidence from measures of prevalence [16–25]. Some of these work only in early epidemics [20], others only in stable conditions [22]; some need long time-series [19–21] or involve complicated and computationally intensive model-fitting proce-

dures [16–25]. Most of the methods have not been validated by comparing derived estimates with actual measurements, and none to our knowledge has been validated with data from more than one African community or in populations experiencing mature or declining HIV epidemics. Furthermore, these earlier methods have not incorporated recent findings on HIV mortality, and their performance has not been assessed under evolving epidemiological conditions, such as changes in risk and increasing access to antiretroviral therapy (ART).

Both of our methods are simple to use and are based on the idea of demographic accounting [22]. They examine the change in HIV prevalence in a cohort observed at two time periods, allowing for changes due to new infections and mortality among infected and uninfected persons. To gather the mortality data required to estimate incidence exactly would also require cohort follow-up, but we show how mortality information collected from external sources, such as long-running cohort studies, can be used in place of direct measurements. Method 1 uses cohort mortality rates among those infected, whilst method 2 uses the distribution of survival after HIV infection. We test the performance of these methods using data generated by a mathematical simulation model and actual measurements of prevalence and incidence from three community-based cohort studies in Zimbabwe, Uganda, and Tanzania.

## Methods

### Calculating Incidence from Cross-Sectional Prevalence

Consider cross-sectional measures of prevalence separated by an interval of $T$ years in age groups of width $r$ years (Figure 1A). Denote the total number of individuals in age-group $i$ at time $j$ as $N_{i,j}$ and the number of HIV-infected individuals as $H_{i,j}$. Prevalence is defined as $p_{i,j} = H_{i,j}/N_{i,j}$. We construct age cohorts centred on initial ages $a_i$ and size $r$ years, i.e., aged $a_i - \frac{r}{2}$ to $a_i - \frac{r}{2}$ at the start of each interval, and $a_i - \frac{r}{2} + T$ to $a_i - \frac{r}{2} + T$ at the end. We denote the age cohort experience of incidence as $\tilde{\lambda}_i$ (the diagonal parallelogram in Figure 1A). By "age cohort" we refer to the individuals in a particular age group at a particular time. The conventional age-specific incidence rate for ages $a_i - \frac{r}{2}$ to $a_i - \frac{r}{2}$ is defined as (shown as a rectangular area in Figure 1B). We note that $\tilde{\lambda}_i$ and $\lambda_i$ are not t $\lambda_i$ he instantaneous rate of new infections but the average hazard of new infections occurring in the interval. Over short intervals with typical HIV incidence rates, these two ways of describing incidence are very similar. In addition, we define $\tilde{\pi}_i$ as the fraction of infected individuals in the $i^{\text{th}}$ age group at the start of the interval who survive to the second serosurvey.

In a real cohort, we can find an approximation to incidence by considering the change in the number of HIV-infected individuals and finding an approximation to the person-years spent at risk of infection (PYAR) in the cohort [24].

Number of seroconversions

$$\approx H_{i,T} - \tilde{\pi}_i H_{i,0} \qquad (1)$$

Number of person-years spent at risk

$$\approx T\left(\frac{(N_{i,0} - H_{i,0}) + (N_{i,T} - H_{i,T})}{2}\right) \qquad (2)$$

It follows (details in Text S1) that an estimate of cohort incidence in the interval is:

$$\tilde{\lambda}_i \approx \frac{2(Q_i p_{i,T} - \tilde{\pi}_i p_{i,0})}{T(1 - p_{i,0} + Q_i(1 - p_{i,T}))}$$ (3)

$$Q_i \approx 1 - (1 - \tilde{\pi}_i)p_{i,0} - (1 - \exp(-\tilde{\mu}_i T))(1 - p_{i,0})$$

where $Q_i$ approximates the proportionate change in size of the cohort over the time interval $T$ and incorporates the cohort mortality rate for those not infected ($\tilde{\mu}_i$). In practice, if this expression gives a negative value for any $\tilde{\lambda}_i$, it is replaced with zero.

To estimate incidence in the cohort that is not included in the data at the start of the time interval (but enters the youngest cross-sectional age group during the interval), we assume that incidence is zero at all ages younger than the youngest person in the dataset (i.e., younger than $a_1 - \frac{r}{2}$). If incidence is further assumed to be constant over the interval and mortality is negligibly small at this age, we can describe prevalence in this unseen cohort as: $p_{0,T} = 1 - \exp(-T\tilde{\lambda}_0/2)$, where $p_{0,T}$ is the prevalence in the second serosurvey in the age range $a_1 - \frac{r}{2}$ to $a_1 - \frac{r}{2} + T$, since the average exposure time in that group is $\frac{T}{2}$. From this, we can generate an estimate of incidence:

$$\tilde{\lambda}_0 = \frac{-\ln(1 - p_{0,T})}{T/2}$$ (4)

## Method 1

Method 1 makes use of empirical values of age-specific cohort mortality rates among those infected, which we denote $\tilde{m}_i$ (Figure 1A). In this way, $\tilde{\pi}_i$ can be approximated by $\tilde{\pi}_i = \exp(-\tilde{m}_i T)$. Three sets of values for $\tilde{m}_i$ are given in Table 1 that are appropriate for applications in early epidemics (defined as epidemics that are still expanding), mature stable epidemics (epidemics that have stopped expanding), or mature declining epidemics (epidemics that are in decline) [26].

## Method 2

Alternatively, if cohort mortality rates are not available, $\tilde{\pi}_i$ can be estimated using survival after infection information [22] and an approximation of the current incidence pattern. If we define $s(a,z)$ to be the probability of survival to age $z$ given infection at exact age $a$ and assume that the pattern of incidence (relative levels of incidence at different ages) has remained constant for as long as the oldest person in the data has been at risk of infection, then provided $T \approx r$:

$$\tilde{\pi}_1 \approx \frac{s\left(a_1 - \frac{r}{2}, a_1 + T\right)}{s\left(a_1 - \frac{r}{2}, a_1\right)}$$

$$\tilde{\pi}_1 \approx \frac{s\left(a_1 - \frac{r}{2}, a_1 + T\right)\tilde{\lambda}_0 + \sum_{k=1}^{i-1}\left\{w_k s\left(a_k + \frac{r}{2}, a_1 + T\right)\right\}}{s\left(a_1 - \frac{r}{2}, a_1\right)\tilde{\lambda}_0 + \sum_{k=1}^{i-1}\left\{w_k s\left(a_k + \frac{r}{2}, a_1\right)\right\}} \text{for i>1}$$

$$w_k = \tilde{\lambda}_k \exp\left(-\sum_{q=0}^{k-1}\tilde{\lambda}_q\right)$$

(5)

This expression calculates the chance that infected individuals die during the follow-up period by assuming that the composition of the infected populations in each age group reflects the pattern of incidence at younger ages (Figure 2). The numerator represents weighted survival from infection at earlier ages to the midpoint of the $i^{th}$ age group at the end of the time interval $[0,T]$. The denominator is the weighted survival to the beginning of the interval. Thus, the ratio is the chance of survival from the start to the end of the intersurvey period for individuals at the central cohort age. We approximate the survival of those infected in each age-group by assuming that they were all infected at the midpoint of that age group and this approximation will be more accurate for smaller values of $r$. The weights ($w_k$) are derived from calculating the chance that an individual was infected at each age group, which is the product of incidence in each age group and the probability that they were not infected at younger ages. For simplicity, we do not represent the periods for which individuals were exposed to these rates, but since the same terms are used in the numerator and denominator of Equation 5 this does not have an important effect provided $r$ and $T$ are small.

Since values of $\tilde{\pi}_i$ are calculated using estimates of incidence from younger age groups, which themselves depend on $\tilde{\pi}_i$, the order of calculation matters; it should be $\tilde{\lambda}_0, \tilde{\pi}_1, \tilde{\lambda}_1, \tilde{\pi}_2, \tilde{\lambda}_2$, and so on. These expressions can be evaluated in full automatically using standard spreadsheet software (Text S2).

The probability of survival for $z$ years after infection at age $a$ is modelled as a Weibull distribution:

$$s(a,z) = \exp\left(-\{[z/\beta_a]^k\}\right)$$ (6)

where $k$ and $\beta_a$ are the shape and scale parameters, respectively. The scale parameter (but not the shape parameter) is assumed to depend on the age at infection. Suggested values of the parameters are given in Table 2, which are from a recent meta-analysis of survival rates in southern and eastern Africa [27].

An alternative formula can be used if $T < r$ (see Text S1).

## Methods 1 and 2: Derivation of Cross-Sectional Measures

The final step in calculating incidence by methods 1 and 2 is to convert the cohort incidence rates into incidence rates that relate to conventional cross-sectional age groups (Figure 1B and 1C). If we assume that the cohort seroconversion rates apply uniformly in the follow-up period, we can do this conversion using:

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

The estimate pertains to the interval between two serosurveys and relies on the intersurvey interval being no greater that the width of the age groups ($T \leq r$). If $T > r$ then the same logic holds, but, because more cohorts pass through the cross-sectional age groups in the period $T$, an alternative to Equation 7 for the relative weights of the cohorts is needed (see Text S1).

Both methods assume that the true rate of incidence in the population is low, that incidence has remained stable in the interval between the two surveys, and that the width of the age groups and the interval between the surveys is short. Even
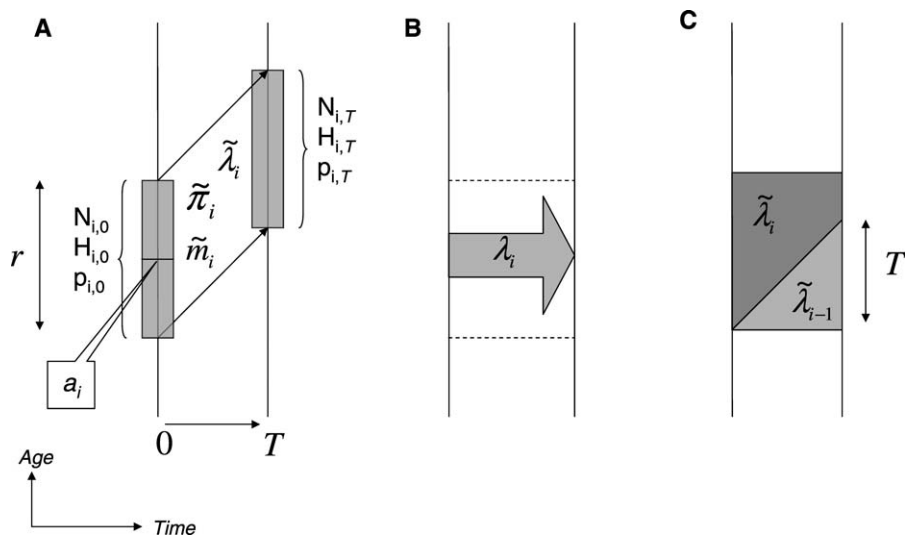
**Figure 1.** Lexis Diagrams Showing Values Used in the Methods

Two serosurveys $T$ years apart quantify prevalence in age groups of width $r$. One age group ($i$) is shown.

(A) At time $j$, $N_{i,j}$ is the total number in survey; $H_{i,j}$ is the number infected; $p_{i,j}$ is HIV; $\tilde{\pi}_i$ is the fraction of infected individuals that survive between surveys; $\tilde{m}_i$ is mortality rate of infected individuals in the cohort; $\tilde{\lambda}_i$ is the rate of HIV incidence in the cohort. $N$ and $H$ are not used required to use the methods but do appear in the mathematical derivation.

(B) Cross-sectional incidence ($\tilde{\lambda}_i$).

(C) Contributions to cross-sectional incidence estimate from incidence in two cohorts ($\tilde{\lambda}_i$ and $\tilde{\lambda}_{i-1}$). For details see main text and Text S1.

doi:10.1371/journal.pmed.0050080.g001

in settings where HIV has reached high endemic levels, incidence rates are typically less than ~4% [1,28], and in household health serosurveys such as DHS, the width of age groups and the interval between surveys is usually 5 y. We therefore believe that these methods are adequate to estimate HIV incidence in most settings with such data. Method 1 further assumes that the pattern of incidence in the populations from which the mortality data are taken is similar to that of the target population. Method 2 instead further assumes that the age pattern of incidence has remained constant for several years and that survival rates with HIV for those infected while in the same age group are approximately equal. The performance of the methods when

these assumptions are violated has been assessed by applying the methods to simulated data (see Text S1 for details of simulation model).

## Validating the Methods in Community-Based Cohort Studies in Africa

To test how well these methods perform, we compare incidence estimates derived using these two methods with actual measurements made in cohort studies. Data were used from three community cohort studies in sub-Saharan Africa; the Manicaland HIV/STD Prevention Project in Zimbabwe [29], the MRC/UVRI Uganda Research Unit on AIDS, which runs a cohort in Masaka district [30] and the NIMR TAZAMA cohort study in Kisesa ward, Tanzania [31]. The Manicaland study furnishes suitable data for only one test time interval (two serosurveys 3 y apart); the Kisesa study provides three test intervals (four surveys, each separated by about 3 y). The Masaka study has conducted 17 annual serosurveys, but 1-y intervals do not generate sufficient incident cases for robust age-specific incidence estimates. To overcome this problem, longitudinal knowledge of HIV status was used to establish HIV prevalence measures at four points in time giving three test intervals in which to measure prevalence, separated by roughly 4 y.

For validation, we used data from "closed cohorts"; that is, we did not include individuals who entered the cohort during a particular interval or were not seen at follow-up and were not known to have died. Seroconversion dates were assigned between the last negative and first positive test results in accordance with analysis procedures developed at each site [32–35]. Poisson-based confidence intervals were calculated for each incidence measurement, which do not take account of any clustering effects in the samples. For method 1, cohort mortality rates typical of "mature,

**Table 1.** Cohort Mortality Rates for HIV-Infected Persons, Tabulated by Age of Cohort at Start of Five-Year Interval and Epidemic Phase [26]

| Age Group, Years | Early | Mature, Stable | Mature, Declining |
|---|---|---|---|
| 15–19 | 0.017 | 0.028 | 0.044 |
| 20–24 | 0.023 | 0.037 | 0.060 |
| 25–29 | 0.029 | 0.048 | 0.077 |
| 30–34 | 0.036 | 0.059 | 0.095 |
| 35–39 | 0.043 | 0.070 | 0.114 |
| 40–44 | 0.051 | 0.083 | 0.134 |
| 45–49 | 0.058 | 0.095 | 0.154 |
| 50–54 | 0.066 | 0.109 | 0.175 |
| 55–59 | 0.075 | 0.122 | 0.197 |
| 60–64 | 0.083 | 0.136 | 0.220 |

Cohort mortality rates typical of mature, declining epidemics were used for Manicaland and Masaka, and rates typical of mature, stable epidemics were used for Kisesa.

doi:10.1371/journal.pmed.0050080.t001

**Table 2.** Parametric Model Estimates of Survival after Infection in Southern and Eastern Africa

| Age Group | Weibull Survival Distribution Parameters | | |
|---|---|---|---|
| | Shape | Scale | Median |
| 15–19 | 2.0 | 16.0 | 13.3 |
| 20–24 | 2.0 | 15.4 | 12.8 |
| 25–29 | 2.0 | 14.1 | 11.7 |
| 30–34 | 2.0 | 12.1 | 10.0 |
| 35–39 | 2.0 | 11.0 | 9.1 |
| 40–44 | 2.0 | 10.1 | 8.4 |
| 45–49 | 2.0 | 7.9 | 6.6 |

These values are model fits to previously published data [27].
doi:10.1371/journal.pmed.0050080.t002

declining" epidemics were used for Manicaland and Masaka, and rates typical of "mature, stable" epidemics were used in Kisesa (Table 1) [26].

## Results

### Testing the Methods on Simulated Data

Simulating data and comparing the known incidence rates with estimates using these methods confirm that incidence can be accurately estimated from serial measures of cross-sectional prevalence. Both methods slightly underestimate incidence when the rate is high and in younger age groups, because infections occur rapidly, and the person-years spent at risk are overestimated by the linear assumptions that underlie the approximations. Although the methods assume that incidence is constant in the period between serosurveys, accurate estimates of average incidence can be obtained for this period even if incidence is increasing or decreasing (Figure 3A and 3B). When incidence falls suddenly, estimates of incidence are immediately reduced and continue to estimate incidence accurately (Figure 3C). In these circumstances, measures of prevalence are slower to respond; after a dramatic fall in incidence of 50% over 5 y, prevalence is expected to decline by only 14%, whereas estimates of incidence based on these methods indicate a 22% reduction (the true average reduction in the 5 y interval is 25%). Immediately after a sudden change in incidence, the estimates are slightly too low because mortality in the older age groups is transiently higher than it would be in a long-term equilibrium with the new pattern of incidence. This bias is smaller when the reduction in incidence is more gradual or when incidence is lower in older age groups.

The methods for estimating incidence depend on using mortality information from an external source. A potential difficulty is that the pattern of mortality actually depends on the pattern of incidence, which is determined by many behavioural factors and could vary between populations and over time. Simulating two alternative scenarios, in which incidence is either highest at young ages (typical for women) or highest at middle ages (typical for men), we can compare the attendant patterns of mortality (Figure 4A). For both scenarios, mortality is low at young ages but much higher at older ages if incidence peaks at young ages. Estimating incidence using method 1 requires age-specific mortality rates for infected individuals and, if the rates used accurately
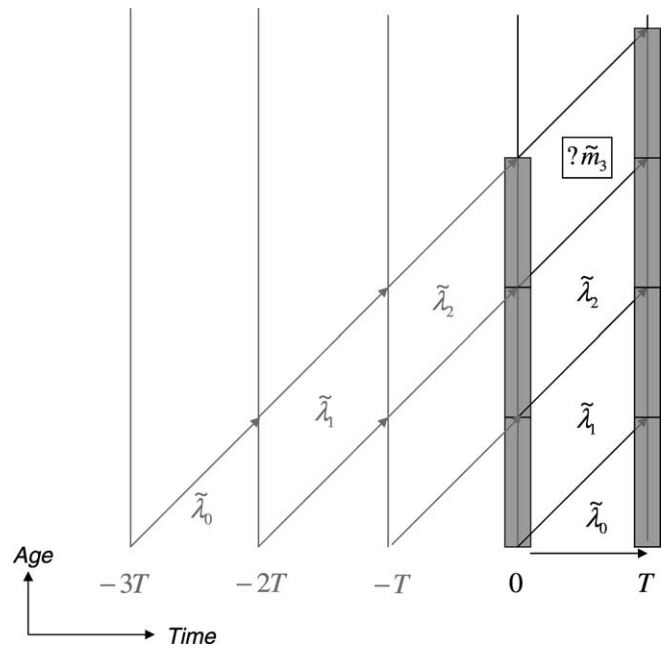


**Figure 2.** Lexis Diagram Showing How Cohort Mortality Rates Can Be Estimated Using Data on Survival after Infection
In order to estimate expected mortality in one cohort (here $\tilde{m}_3$) using data on survival after infection, it is necessary to approximate the composition of the infected population in that cohort with respect to time since infection. Thus, it is necessary to consider the previous exposure of that cohort to incidence (indicated by the grey lines and text). When the intersurvey period is the same as the width of the age groups ($T = r$), and the pattern of incidence is constant, the previous experience of the cohorts will be reflected in the experience of younger cohorts (i.e., cohorts 0, 1, and 2).
doi:10.1371/journal.pmed.0050080.g002

reflect the pattern of mortality in the population, the estimates will be good (solid lines, Figure 4B). However, if the mortality rates are taken from a population with a different distribution of incidence and mortality, the derived incidence estimates will be biased, particularly at older ages (dashed lines, Figure 4B). This bias is small if incidence is concentrated at young ages. Method 2, which does not rely on using age-specific mortality rates, is able to accurately predict incidence in either scenario using the same data on survival after infection and assuming that age patterns of incidence have remained constant in the years preceding the surveys.

When the age pattern of incidence changes over time, small errors are introduced to the estimates using method 2 (Figure 4C). The errors are highest when the change occurs shortly before the first survey or in the interval between the two surveys.

If ART provision increases from 0% to 30% in the interval between two serosurveys, then the derived estimates of incidence are too high when mortality data from the pre-ART era are used (Figure 5). (ART coverage levels are likely to be much lower than this in most countries in sub-Saharan Africa [36].) This overestimate occurs because individuals are surviving longer with infection leading to increased HIV prevalence, but that prevalence trend is wrongly attributed to new infections by the estimation methods. In these simulations, incidence among adults is only slightly overestimated at first (~4%), and the errors are mostly limited to older ages. The errors in later estimates of incidence, based on
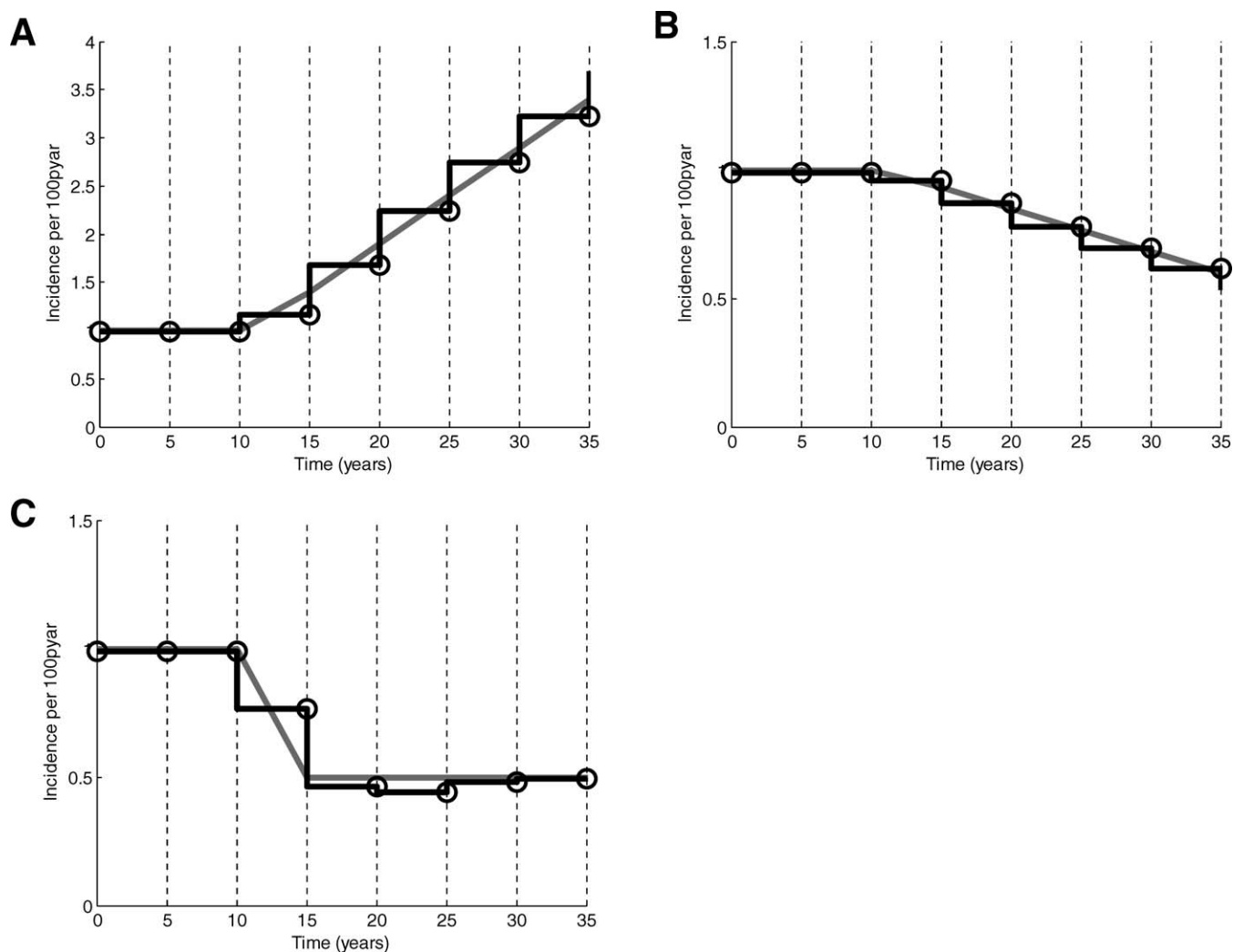
**Figure 3.** Comparison of True Simulated Incidence Rate (Grey Lines) and Estimates Using Either Method (Black Lines; Both Give Same Results) when True Incidence Increases Steadily (A), Decreases Steadily (B), or Decreases Suddenly (C)

Vertical lines indicate when five-yearly serosurveys are done. Estimates of incidence are made at the time of the serosurvey (open circles) but relate to the preceding 5-y period.

doi:10.1371/journal.pmed.0050080.g003

serosurveys taken while ART provision is maintained, become greater as more individuals start ART. Errors at younger ages also increase over time, particularly if incidence is high at young ages.

### Testing the Methods on Data from Cohort Studies

Measurements of incidence in adult and young people, together with the corresponding estimates from both methods are given in Table 3. Almost all estimates fall within the 95% confidence interval of the observed measurements. For method 1, the mean difference between the measurements and estimates for adults is 19%. For method 2, the mean difference is 14% and it is less than 10% in half of the comparisons. Neither method tends to systematically over- or underestimate incidence.

Measured and estimated incidence rates are compared by age group in Figure 6. The overall performance of both methods is good; the pattern and approximate level of incidence is captured successfully in all cases. The estimates

lie within the 95% confidence interval of measurements in 80% (method 1) and 88% (method 2) of comparisons. The average error across all comparisons at each age group for each pair of surveys and at each site is ~35% for both methods. Estimates at older ages are most likely to deviate from measurements substantially. The estimates of incidence among 25- to 34-y-old men and women in the first period in Masaka are the least accurate (Figure 6B).

### Discussion

We developed two methods for estimating HIV incidence in the general population using successive rounds of cross-sectional prevalence data, and tested how well these methods perform using model-simulated data and real data from three African cohort studies. Spreadsheets for implementation of both methods to estimate incidence are provided in Text S2. Both methods provided good estimates of incidence in adults and young people and captured the pattern of incidence with
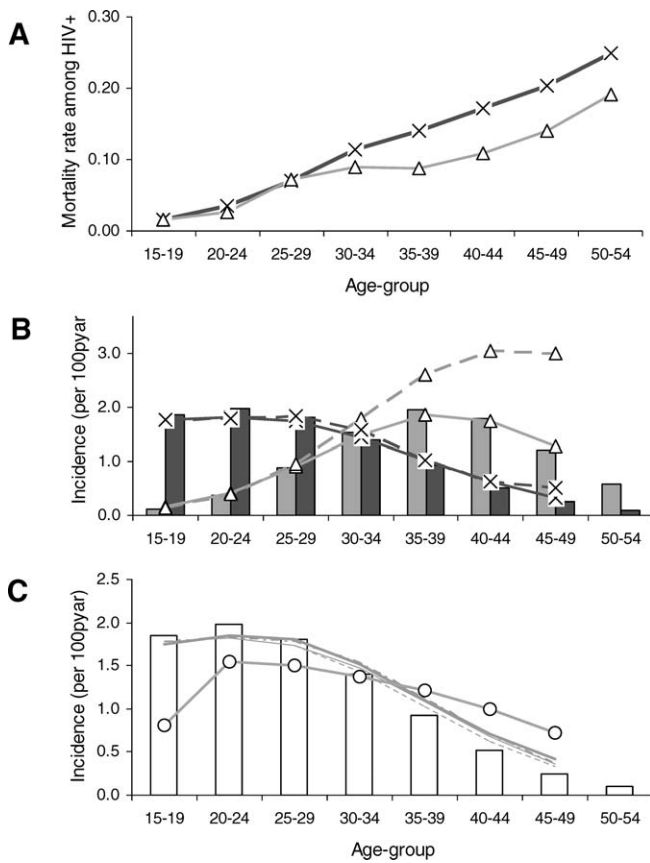
**Figure 4.** Comparison of Method Estimates With Simulated Data Under Range of Conditions Violating Underlying Method Assumptions.

(A) The age-specific mortality rates for HIV-infected individuals when incidence is highest at young ages (dark line with crosses) or middle ages (grey line with triangles).
(B) Estimates of incidence using method 1 when the "wrong" pattern age-specific mortality rates are used (i.e., from the alternative scenario in [A]: dashed lines) and when the correct rates or method 2 is used (solid lines). Bars show simulated incidence rates.
(C) Estimates of incidence using method 2 when the age pattern of incidence changes (instantaneously between the two scenarios shown in [B]), 5 y (thick line), 10 y (dash-dot line), or 15 y (thin line) before the first survey. The dotted lines show the estimate if the age pattern of incidence does not change; the line with circles shows the estimates if the age pattern changes in the interval between the two surveys. The bars show the average incidence rate in the intersurvey period.
doi:10.1371/journal.pmed.0050080.g004

respect to age. Since incidence is not routinely measured directly at large scales, these methods could be of substantial use in monitoring and comparing the progress of national epidemics, contributing to the interpretation of observed behavioural or epidemiological trends, and refining estimates of disease burden, treatment needs, and the future course of the epidemic. The serial measurements of cross-sectional prevalence that are required for these methods will soon be available from household surveys with HIV testing, such as DHS.

Estimates of incidence in adults and young people were in very close agreement with actual measurements. Greater discrepancies occurred at older ages, because these estimates are most sensitive to the assumptions about mortality and because the comparison incidence measurement itself is more uncertain due to relatively small numbers of seroconversions. Both methods markedly underestimate incidence in
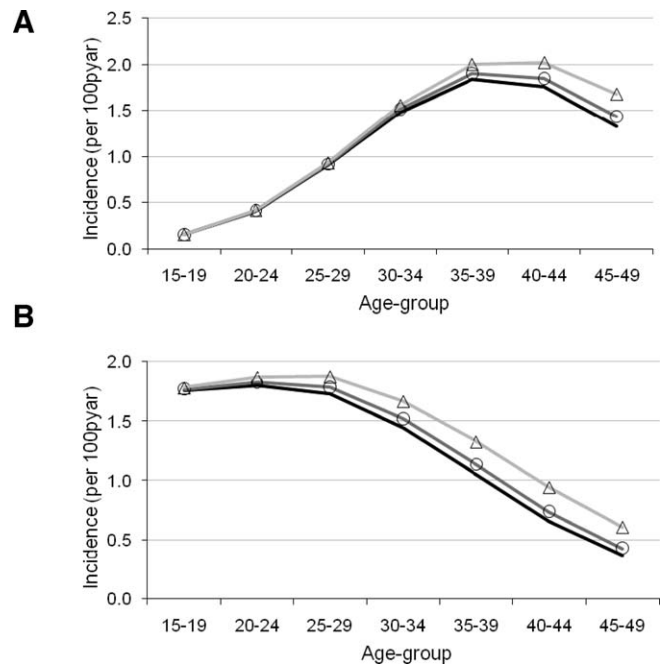


**Figure 5.** Incidence Estimates Using Method 2 Assuming that Provision of Antiretroviral Therapy is Scaled Up from 0% to 30% over Five Years
The grey line with circles shows the estimate of incidence based on surveys before and after the 5-y scale-up; the grey line with triangles shows the estimate of incidence based on surveys after scale-up, while provision is maintained at 30%. The black line shows the estimates if ART is not provided. Graphs show the simulations assuming that the incidence rate is highest at (A) older ages and (B) young ages. Similar results are obtained using method 1.
doi:10.1371/journal.pmed.0050080.g005

the first period from Masaka among 25- to 34-y-old men and women. This underestimate could result from (i) rapid changes in incidence and prevalence at this time (1991–1995), when there was a transition from epidemic growth to decline [35], or alternatively, (ii) the mortality rates used in the calculations being too low because the underlying stability assumptions of both methods were breached by changes in the age pattern of incidence. Such changes are now unlikely in most other African countries where epidemics have matured gradually.

For method 2, the mean error in these estimates of incidence in adults is low: 14% over all comparisons or 10% if the earliest data from Masaka are excluded. For method 1 the mean error is slightly greater (19%), but the mortality rates used were based on regional aggregated data [26] and the accuracy of estimates would be higher if local data were used instead. The uncertainty in estimates made in other settings is hard to quantify and will extend beyond random statistical errors, because the extent to which assumptions and approximations hold will not be known. However, some insight can be derived by using a range of scenarios for the mortality data, which reflect a range of epidemiological conditions (Table 1) [26].

Simulations indicate that these methods could be used to identify and quantify changes in incidence earlier and more accurately than when only trends in prevalence are examined. This is an important application since the monitoring of trends is essential for coordinating an effective response to epidemics [37,38]. However, changes in incidence estimates

**Table 3.** Measurements and Estimates of Incidence in Adults

| Population | Location and Period | Sex | Age Group, Years | Incidence, Observed | 95% CI | Estimated Incidence, Method 1 | Agreement | Estimated Incidence, Method 2 | Agreement |
|---|---|---|---|---|---|---|---|---|---|
| **Adults** | Manicaland 1 | Male | 15–54 | 1.76 | 1.43–2.16 | 2.04 | ** | 1.74 | *** |
| | Manicaland 1 | Female | 15–54 | 1.48 | 1.23–1.79 | 2.07 | — | 1.78 | * |
| | Masaka 1 | Male | 15–44 | 0.44 | 0.29–0.67 | 0.28 | — | 0.28 | — |
| | Masaka 1 | Female | 15–44 | 0.61 | 0.43–0.86 | 0.41 | — | 0.38 | — |
| | Masaka 2 | Male | 15–44 | 0.63 | 0.43–0.91 | 0.60 | *** | 0.52 | ** |
| | Masaka 2 | Female | 15–44 | 0.37 | 0.23–0.61 | 0.42 | ** | 0.40 | *** |
| | Masaka 3 | Male | 15–44 | 0.40 | 0.28–0.57 | 0.48 | * | 0.42 | *** |
| | Masaka 3 | Female | 15–44 | 0.57 | 0.43–0.75 | 0.56 | *** | 0.53 | *** |
| | Kisesa 1 | Male | 15–44 | 0.74 | 0.51–1.09 | 0.72 | *** | 0.74 | *** |
| | Kisesa 1 | Female | 15–44 | 0.87 | 0.63–1.21 | 0.76 | ** | 0.81 | *** |
| | Kisesa 2 | Male | 15–44 | 1.42 | 1.09–1.84 | 1.13 | * | 1.17 | ** |
| | Kisesa 2 | Female | 15–44 | 1.32 | 1.03–1.67 | 1.00 | * | 1.08 | ** |
| | Kisesa 3 | Male | 15–54 | 1.21 | 0.93–1.58 | 0.89 | * | 1.05 | ** |
| | Kisesa 3 | Female | 15–54 | 1.19 | 0.94–1.49 | 0.95 | ** | 1.11 | *** |
| **Young people** | Manicaland 1 | Male | 17–25 | 1.24 | 0.59–2.61 | 1.63 | — | 1.63 | — |
| | Manicaland 1 | Female | 15–25 | 2.00 | 1.33–3.01 | 2.22 | ** | 2.22 | ** |
| | Masaka 1 | Male | 15–24 | 0.31 | 0.16–0.62 | 0.38 | * | 0.34 | *** |
| | Masaka 1 | Female | 15–24 | 0.98 | 0.64–1.50 | 0.91 | *** | 0.80 | ** |
| | Masaka 2 | Male | 15–24 | 0.29 | 0.14–0.61 | 0.41 | — | 0.39 | — |
| | Masaka 2 | Female | 15–24 | 0.52 | 0.28–0.97 | 0.61 | ** | 0.52 | *** |
| | Masaka 3 | Male | 15–24 | 0.27 | 0.15–0.50 | 0.43 | — | 0.41 | — |
| | Masaka 3 | Female | 15–24 | 0.73 | 0.50–1.07 | 0.87 | ** | 0.80 | *** |
| | Kisesa 1 | Male | 15–24 | 0.39 | 0.10–1.56 | 0.45 | ** | 0.44 | ** |
| | Kisesa 1 | Female | 15–24 | 1.09 | 0.45–2.61 | 1.04 | *** | 1.02 | *** |
| | Kisesa 2 | Male | 15–24 | 1.42 | 0.68–2.98 | 1.47 | *** | 1.47 | *** |
| | Kisesa 2 | Female | 15–24 | 1.93 | 1.17–3.21 | 1.49 | * | 1.48 | * |
| | Kisesa 3 | Male | 15–24 | 1.15 | 0.37–3.57 | 1.13 | *** | 1.13 | *** |
| | Kisesa 3 | Female | 15–24 | 1.25 | 0.56–2.79 | 0.94 | * | 0.93 | * |

All incidence values given per 100 pyar. Estimated value are standardised to same age-distribution as measurement.
Agreement between estimate and measurement: absolute error values *** < 10%; ** < 20%; * < 30%.
CI, confidence interval.

can also be generated by shifts in the age pattern of incidence, so data from parallel behavioural surveillance must be used to assess this possibility when such changes are found [37].

In principle, these methods should work just as well in other settings, provided that locally applicable mortality data are used. The data we have used in the validation exercise were considered appropriate because they came from the same region and from a period during which ART was not widely available. The choice of method (1 or 2) elsewhere will depend on which assumptions seem more likely to be true for the population in question and on the availability of locally collected data. The investigations using simulated data alert us to the danger of using the wrong age-specific mortality rates in method 1, which can vary according to the phase of the epidemic and the prevailing pattern of incidence. We are reassured that in some settings the age-specific mortality rates follow a similar pattern [26], but this does not necessarily mean that they will do so elsewhere [39] (e.g., in other types of communities or other parts of Africa). Method 2, on the other hand, may be applied to settings with different incidence patterns using the same data on survival after infection, because it effectively calculates its own mortality rates by assuming that incidence patterns have remained constant in the recent past. The agreement between the results from the two methods indicates how successfully the procedure for estimating mortality rates in method 2 works.

Applying cohort mortality rates or data on survival post-infection from populations without treatment to populations in which treatment has been available for several years could lead to inflated estimates of incidence. We recommend that method 1 be used where recent age-specific cohort mortality data that pertain to the population in question are available; but where they are not, method 2 may be used with a regional estimate of survival after infection instead.

In the tests against actual measurement of incidence, the mortality rates used were partly informed by data from the same cohorts (50% of the person-years used to derive the age-specific estimates in [26]). This may overstate how well method 1 will perform at a national level, because there would not necessarily be such agreement between the actual mortality in the population in question and in the external mortality rates that are used. We emphasise that in this study we sought to validate the methods rather than a particular set of mortality data. Also, whilst in these tests we have been able to use real cohorts to measure incidence, in the application to the household surveys with HIV testing, cross-sectional prevalence measures for birth cohorts (with different individuals appearing in the two serosurveys) would have to be relied upon. As with other analyses of trends in prevalence, as long as any differences in participation rate according to serostatus remain constant, no additional biases will be introduced.
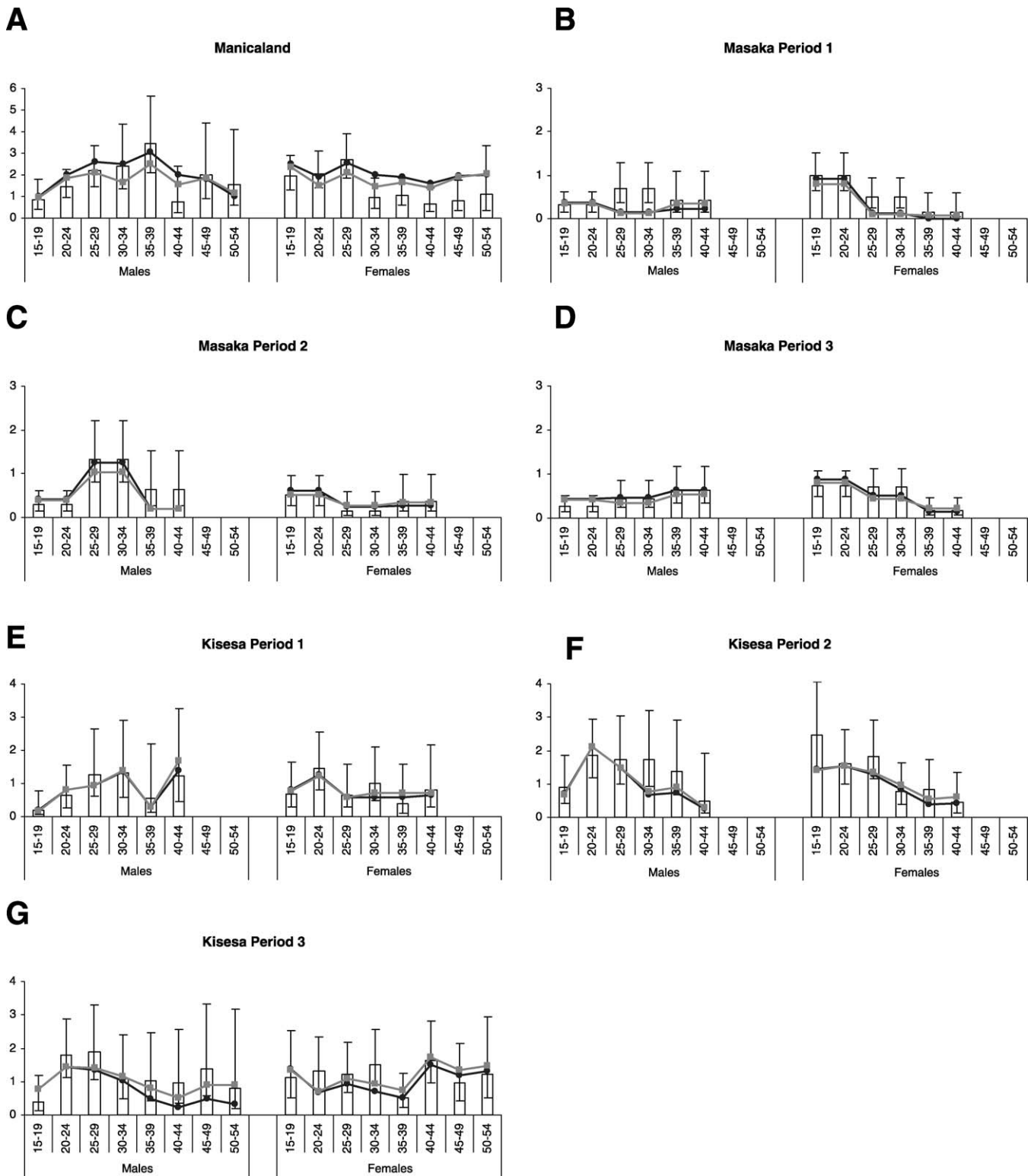
**Figure 6.** Result from the Cohort Studies Analysis

For each graph, the bars show incidence measured in the closed cohort with 95% confidence intervals and the lines show derived incidence estimates using method 1 (dark grey line) and method 2 (light grey line). (A) Manicaland; (B) Masaka, period 1; (C) Masaka, period 2; (D) Masaka, period 3; (E) Kisesa, period 1; (F) Kisesa, period 2; (G) Kisesa, period 3.

doi:10.1371/journal.pmed.0050080.g006

The methods do not correct for differential migration between infected and uninfected individuals and, by using closed populations, we have removed the effect of migration from the test data, which may improve their performance. However, we intend these methods to be applied to national-level data, and since rates of international migration are much lower than internal migration (e.g., between rural and urban areas and migration in and out of field sites in cohort studies), international migration should not introduce serious errors. Moreover, incidence estimates will only be misleading if migration rates vary by HIV status; for example, incidence will be underestimated if infected individuals are more likely to leave the country than others, or overestimated if individuals from a higher-prevalence country immigrate. Although regional migration has been associated with HIV infection in spreading epidemics [40], it is thought that this pattern is mostly driven by increased vulnerability to infection rather than more frequent movement among high-risk groups [41,42]. Most cross-sectional studies cannot distinguish these possibilities, but recent cohort data from eastern Zimbabwe show that rural–urban migrants currently have socioeconomic characteristics and vulnerability to HIV before migrating similar to those of nonmigrants [43]. In most settings it is not possible to quantify rates of migration with respect to HIV status, but if the same association is true for international migrants it is unlikely that migration will lead to large errors in incidence estimates. Therefore, as with other analyses of prevalence time series [7], the estimates should be interpreted with care if international migration rates are high, have recently changed, or are dominated by movement from a country with a very different HIV prevalence.

If treatment with ART becomes more common, the mortality rate of infected people will change, and use of the mortality data provided here would lead to overestimates of incidence, particularly at older ages, for which the effect of ART will be greatest. However, our simulations indicate that, in the next few years, this error may be small because not all of those in need will start treatment [36], and the proportion of infected individuals who are on treatment will increase gradually over time. If survival on ART is longer than we have assumed, then we would expect the errors to be greater. The errors in the estimates would increase over time if mortality data from the pre-ART era continued to be used, but new data from cohort studies incorporating the impact of ART will soon become available. The utility of these new measurements will depend on the criteria for treatment access and uptake levels being broadly similar in the cohort studies and national populations. Estimates of incidence among young people (under age 25 y) using existing mortality data from the pre-ART era are expected to remain reliable even when ART is common, as we have found that HIV mortality and ART eligibility are very low in this age range [44].

## Supporting Information

**Text S1.** Derivation of Methods and Further Technical Information
Found at doi:10.1371/journal.pmed.0050080.sd001 (89 KB PDF).

**Text S2.** Spreadsheets for Implementation of Methods 1 and 2 to Estimate Incidence
Found at doi:10.1371/journal.pmed.0050080.sd002 (116 KB XLS).

### References

1. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2006) 2006 Report on the global AIDS epidemic. Geneva: United Nations. Available: http://www.unaids.org/en/HIV__data/2006GlobalReport/default.asp. Accessed 8 March 2008.
2. UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2000) Guidelines for second generation HIV surveillance for HIV: The next decade (2000). Geneva: WHO. WHO/CDS/EDC/2000.05. Available: http://www.who.int/hiv/pub/surveillance/pub3/en/index.html. Accessed 8 March 2008.
3. Chin J (1990) Public health surveillance of AIDS and HIV infections. Bull World Health Organ 68: 529–536.
4. UNAIDS, Wellcome Trust Centre for the Epidemiology of Infectious Diseases (1999) Trends in HIV incidence and prevalence: Natural course of the epidemic or results of behavioural change? UNAIDS Best Practice Collection. Geneva: UNAIDS. UNAIDS 99.12e. Available: http://www.who.int/hiv/strategic/surveillance/pubincprev/en/index.html. Accessed 8 March 2008.
5. Hallett TB, Aberle-Grasse J, Bello G, Boulos LM, Cayemittes MPA, et al. (2006) Declines in HIV prevalence can be associated with changing sexual behaviour in Uganda, urban Kenya, Zimbabwe, and urban Haiti. Sex Transm Infect 82: i1–8.
6. Zaba B, Gregson S (1998) Measuring the impact of HIV on fertility in Africa. AIDS 12: S41–50.
7. Ghys PD, Kufa E, George MVfor the Unaids Reference Group on Estimates Modelling and Projections' ad hoc Working Group on interpreting trends in prevalence and incidence of HIV infection in countries with generalised epidemics (2006) Measuring trends in prevalence and incidence of HIV infection in countries with generalised epidemics. Sex Transm Infect 82: i52–56.
8. [No author listed] (2008) Measure DHS: HIV prevalence. Calverton (Maryland): Demographic and Health Surveys, Macro International. Available: http://www.measuredhs.com/topics/hivprev/start.cfm. Accessed 8 March 2008.
9. Boerma JT, Ghys PD, Walker N (2003) Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard. Lancet 362: 1929–1931.
10. Garcia-Calleja JM, Gouws E, Ghys PD (2006) National population based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates. Sex Transm Infect 82: iii64–70.
11. Rangsin R, Chiu J, Khamboonruang C, Sirisopana N, Eiumtrakul S, et al. (2004) The natural history of HIV-1 infection in young Thai men after seroconversion. J Acquir Immune Defic Syndr 36: 622–629.
12. [No authors listed] (2000) Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: A collaborative re-analysis. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Concerted Action on SeroConversion to AIDS and Death in Europe. Lancet 355: 1131–1137.
13. Morgan D, Mahe C, Mayanja B, Okongo JM, Lubega R, et al. (2002) HIV-1 infection in rural Africa: Is there a difference in median time to AIDS and survival compared with that in industrialized countries? AIDS 16: 597–603.
14. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, et al. (2002) Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. AIDS Res Hum Retroviruses 18: 295–307.
15. Karita E, Price M, Hunter E, Chomba E, Allen S, et al. (2007) Investigating the utility of the HIV-1 BED capture enzyme immunoassay using cross-sectional and longitudinal seroconverter specimens from Africa. AIDS 21: 403–408.
16. Podgor MJ, Leske MC (1986) Estimating incidence from age-specific prevalence for irreversible diseases with differential mortality. Stat Med 5: 573–578.
17. Ades AE, Medley GF (1994) Estimates of disease incidence in women based on antenatal or neonatal seroprevalence data: HIV in New York City. Stat Med 13: 1881–1894.
18. Ades AE (1995) Serial HIV seroprevalence surveys: interpretation, design, and role in HIV/AIDS prediction. J Acquir Immune Defic Syndr Hum Retrovirol 9: 490–499.
19. Sakarovitch C, Alioum A, Ekouevi DK, Msellati P, Leroy V, et al. (2007) Estimating incidence of HIV infection in childbearing age African women using serial prevalence data from antenatal clinics. Stat Med 26: 320–335.
20. Williams B, Gouws E, Wilkinson D, Karim SA (2001) Estimating HIV

incidence rates from age prevalence data in epidemic situations. Stat Med 20: 2003–2016.

21. White RG, Vynnycky E, Glynn JR, Crampin AC, Jahn A, et al. (2007) HIV epidemic trend and antiretroviral treatment need in Karonga District, Malawi. Epidemiol Infect: 1–11.

22. Gregson S, Donnelly CA, Parker CG, Anderson RM (1996) Demographic approaches to the estimation of incidence of HIV-1 infection among adults from age-specific prevalence data in stable endemic conditions. AIDS 10: 1689–1697.

23. Saidel T, Sokal D, Rice J, Buzingo T, Hassig S (1996) Validation of a method to estimate age-specific human immunodeficiency virus (HIV) incidence rates in developing countries using population-based seroprevalence data. Am J Epidemiol 144: 214–223.

24. Batter V, Matela B, Nsuami M, Manzila T, Kamenga M, et al. (1994) High HIV-1 incidence in young women masked by stable overall seroprevalence among childbearing women in Kinshasa, Zaire: estimating incidence from serial seroprevalence data. AIDS 8: 811–817.

25. UNAIDS Reference Group on Estimates Modelling and Projections (2002) Improved methods and assumptions for estimation of the HIV/AIDS epidemic and its impact: Recommendations of the UNAIDS Reference Group on Estimates, Modelling and Projections. AIDS 16: W1–14.

26. Zaba B, Marston M, Crampin M, Isingo R, Biraro S, et al. (2007) Age-specific mortality patterns of HIV infected persons: a comparative analysis of community study data from five African countries. AIDS 21: S87–S96.

27. Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, et al. (2007) Time from HIV seroconversion to death: A collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy. AIDS 21: S55–S63.

28. UNAIDS/WHO (2007) AIDS epidemic update. Geneva: UNAIDS. Available: http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2007/default.asp. Accessed 8 March 2008.

29. Gregson S, Garnett GP, Nyamukapa CA, Hallett TB, Lewis JJ, et al. (2006) HIV decline associated with behavior change in eastern Zimbabwe. Science 311: 664–666.

30. Mulder DW, Nunn AJ, Wagner HU, Kamali A, Kengeya-Kayondo JF (1994) HIV-1 incidence and HIV-1-associated mortality in a rural Ugandan population cohort. AIDS 8: 87–92.

31. Boerma JT, Urassa M, Senkoro K, Klokke A, Ngweshemi JZ (1999) Spread of HIV infection in a rural area of Tanzania. AIDS 13: 1233–1240.

32. Lopman B, Nyamukapa C, Mushati P, Mupambireyi Z, Mason P, et al. (2008) HIV incidence in 3 years of follow-up of a Zimbabwe cohort—1998–2000 to 2001–03: contributions of proximate and underlying determinants to transmission. Int J Epidemiol 37: 88–105.

33. Wambura M, Urassa M, Isingo R, Ndege M, Marston M, et al. (2007) HIV prevalence and incidence in rural Tanzania: results from 10 years of follow-up in an open-cohort study. J Acquir Immune Defic Syndr 46: 616–623.

34. Shafer LA, Biraro S, Kamali A, Grosskurth H, Kirungi W, et al. (2006) HIV prevalence and incidence are no longer falling in Uganda—a case for renewed prevention efforts: evidence from a rural population cohort 1989–2005, and from ANC surveillance [Abstract THLB0108]. 16th International AIDS Conference. 2006 August; Toronto, Canada. International AIDS Society. Available: http://www.iasociety.org/Search.aspx. Accessed 8 March 2008.

35. Kamali A, Carpenter LM, Whitworth JA, Pool R, Ruberantwari A, et al. (2000) Seven-year trends in HIV-1 infection rates, and changes in sexual behaviour, among adults in rural Uganda. AIDS 14: 427–434.

36. Boerma JT, Stanecki KA, Newell ML, Luo C, Beusenberg M, et al. (2006) Monitoring the scale-up of antiretroviral therapy programmes: Methods to estimate coverage. Bull World Health Organ 84: 145–150.

37. Garnett GP, Garcia-Calleja JM, Rehle T, Gregson S (2006) Behavioural data as an adjunct to HIV surveillance data. Sex Transm Infect 82: i57–62.

38. Hallett TB, White PJ, Garnett GP (2007) Appropriate evaluation of HIV prevention interventions: From experiment to full-scale implementation. Sex Transm Infect 83: i55–60.

39. Porter K, Zaba B (2004) The empirical evidence for the impact of HIV on adult mortality in the developing world: data from serological studies. AIDS 18: S9–S17.

40. Duckett M (2000) Migrants and HIV / AIDS. Dev Bull: 18–20.

41. UNAIDS, International Organisation for Migration (2001) Population mobility and AIDS: UNAIDS technical update. Geneva: UNAIDS. Available: http://data.unaids.org/Publications/IRC-pub02/JC513-PopMob-TU__en.pdf. Accessed 8 March 2008.

42. International Organisation for Migration (2004) Population mobility and HIV/AIDS. Available: http://www.old.iom.int/documents/publication/en/iom__hiv__brochure__july__2004.pdf. Accessed 8 March 2008.

43. Mundandi C, Vissers D, Voeten H, Habbema D, Gregson S (2006) No difference in HIV incidence and sexual behaviour between out-migrants and residents in rural Manicaland, Zimbabwe. Trop Med Int Health 11: 705–711.

44. World Health Organization, UNAIDS, UNICEF (2007) Towards universal access: Scaling up priority HIV/AIDS interventions in the health sector. Available: http://www.who.int/hiv/mediacentre/. Accessed 30 October 07.

## Editors' Summary

**Background.** More than 25 million people have died from AIDS and about 33 million people are currently infected with human immunodeficiency virus (HIV, the virus that causes AIDS). Faced with this threat to human health, governments and international agencies are working together to halt the AIDS epidemic. An important part of this effort is HIV surveillance. The spread of HIV needs to be monitored to assess the impact of interventions (for example, the provision of antiretroviral drugs) and to plan for current and future health care needs. HIV surveillance in countries where the epidemic has spread beyond specific groups into the whole population (a generalized epidemic) has mainly relied on determining the prevalence of HIV infection (the fraction of the population that is infected) among women attending antenatal clinics. Recently, however, household health surveys (for example, the Demographic and Health Surveys) have begun to use blood testing for antibodies to the AIDS virus (serological testing) to provide more accurate estimates of HIV prevalence in the general adult population.

**Why Was This Study Done?** Although prevalence estimates provide useful information about the HIV epidemic, another important indicator is incidence—the number of new infections occurring during a specific time period. Incidence measurements provide more information about temporal changes in the epidemic and transmission patterns and allow public-health experts to make better predictions of future health care needs. But, whereas prevalence can be measured with anonymized serological surveys, individuals would have to be identified and followed up in repeat serological surveys to provide a direct measurement of incidence. This is expensive and hard to achieve in many settings. In this study, therefore, the researchers develop and validate two mathematical methods to estimate HIV incidence in generalized HIV epidemics from prevalence data.

**What Did the Researchers Do and Find?** Changes in the fraction of the population living with HIV (prevalence) can occur not only because of changes in the rate of new infections (incidence), but also because mortality rates are much higher for infected individuals than others. The researchers' methods disentangle the contributions to HIV prevalence (as measured in serological surveys) made by new infections from those due to deaths from AIDS and other causes. Their first method incorporates information on death rates collected in cohort studies of HIV infection (cohort studies investigate outcomes in groups of people); their second method uses information on survival after HIV infection, also collected in long-running cohort studies. The accuracy of both methods was assessed using computer-simulated data and actual data on HIV prevalence and incidence collected in three community-based cohort studies in Zimbabwe and Uganda (countries with generalized but

declining HIV epidemics) and Tanzania (a country with a generalized, stable epidemic). Both methods provided accurate estimates of HIV incidence from the simulated data. Using the data collected in Africa, the mean difference between actual measurements of incidence and the estimate provided by method 1 was 19%; for method 2 it was 14%. In addition, the measured and estimated incidences were in good agreement for all age groups.

**What Do These Findings Mean?** These findings suggest HIV incidence rates can be estimated from repeat surveys of prevalence with sufficient accuracy to monitor the HIV epidemic. The accuracy of the estimates across all age groups is particularly important because knowledge of the age-related risk pattern provides the information on transmission patterns that is needed to design effective intervention programs. Because these methods were tested using data only from southern and eastern Africa where the HIV epidemic is mature and generalized, they may not work as well in regions where the epidemic is restricted to subsets of the population. Other factors that might affect their accuracy include the amount of international migration and the uptake of antiretroviral therapies. Nevertheless, with the increased availability of serial measurements of serological prevalence, these new methods for estimating HIV incidence from HIV prevalence could prove extremely useful for monitoring the progress of national HIV epidemics and for guiding HIV control programs. The authors include spreadsheets that can be used to calculate incidence by either method from consecutive survey data.

**Additional Information.** Please access these Web sites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.0050080.

- The US National Institute of Allergy and Infectious Diseases provides information on HIV infection and AIDS
- The US Centers for Disease Control and Prevention provides information on global HIV/AIDS topics (in English and Spanish)
- The HIV InSite provides comprehensive and up-to-date information on all aspects of HIV/AIDS from the University of California San Francisco, including country reports on the AIDS epidemic in 195 countries, including Uganda, Zimbabwe, and Tanzania
- Avert, an international AIDS charity, provides information on all aspects of HIV/AIDS, including fact sheets on understanding HIV and AIDS statistics, and on HIV and AIDS in Africa
- The Demographic and Health Surveys program collects, analyzes, and disseminates information on health and population trends in countries around the world