

THE CURRENT STATUS OF METHODS FOR ESTIMATING THE PREVALENCE OF HUMAN IMMUNODEFICIENCY VIRUS IN THE UNITED STATES OF AMERICA[†]

JOHN M. KARON^{1*}, MEENA KHARE² AND PHILIP S. ROSENBERG³

¹ *Division of HIV/AIDS Prevention – Surveillance and Epidemiology (E-48), National Center for HIV, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, U.S.A.*

² *Division of Health Examination Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland, U.S.A.*

³ *National Cancer Institute, Epidemiology and Biostatistics Program, Epidemiologic Methods Section, Rockville, MD 20852, U.S.A.*

SUMMARY

The prevalence of human immunodeficiency virus (HIV) infection can be estimated by two distinct methods. One method, back-calculation, is a complex statistical procedure that estimates the HIV epidemic curve. The second method is based on data from population-based surveys, which provide estimates of the proportion of persons infected with HIV within subgroups, and on the known or estimated population totals for these subgroups. Estimates from these methods are subject to substantial uncertainty and bias, both of which are difficult to quantify. We review recent use of these procedures to estimate HIV prevalence in the United States of America. We also summarize new data on the uncertainty and the bias in these estimates. Reliable estimates of HIV prevalence can be made only by synthesizing estimates from several procedures and by a comprehensive evaluation of relevant data. Future estimates of HIV prevalence will require modifications of these methods or the development of new methods. © 1998 John Wiley & Sons, Ltd.

Statist. Med., 17, 127–142 (1998)

1. INTRODUCTION

The U.S. Public Health Service recently published new estimates of the prevalence of human immunodeficiency virus (HIV) infection among U.S. residents.¹ These estimates were obtained by synthesizing the results from three different data sources and statistical estimation procedures. One estimate was based on an HIV seroprevalence survey among childbearing women that provided estimates of seroprevalence rates (the proportion of persons infected with HIV) among women of childbearing age within specific demographic strata. The HIV prevalence among women of childbearing age was estimated from these rates and U.S. census population data. Another estimate used data from a national household survey of current health status and

* Correspondence to: John M. Karon, Division of HIV/AIDS Prevention – Surveillance and Epidemiology (E-48), National Center for HIV, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, U.S.A.

[†] W. Meade Morgan served as the editor for this manuscript

Table I. Estimates of HIV prevalence, by data source and sex, United States, 1992*

Data source	Males	Females	Total
Back-calculation [†]	525,000–750,000	120,000–160,000	650,000–900,000
SCBW	550,000–700,000	120,000–160,000	650,000–900,000
NHANES III [‡]	300,000–725,000	60,000–250,000	400,000–900,000

Reproduced from Table 1, *Journal of the American Medical Association*, 276, 128 (1996).

* HIV indicates human immunodeficiency virus, SCBW the Survey in Childbearing Women, and NHANES III the Third National Health and Nutrition Examination Survey. Estimates include persons already diagnosed as having the acquired immunodeficiency syndrome. These estimates include 10,000 children. Estimates for males, females and total are round to the nearest 25,000, the nearest 10,000, and the nearest 50,000, respectively. The totals may differ from the sums of estimates for males and females due to rounding

[†] Based on analyses by sex and race/ethnicity

[‡] Adjusted for persons not covered by the survey design, but not for possible higher HIV prevalence among persons less likely to participate in the survey

statistical procedures which account for the complex multi-stage area probability sampling procedure used in the survey.² The third method, back-calculation, was developed to reconstruct the history of the HIV epidemic.³

Each of these methods provides an estimate of HIV prevalence for a subpopulation of the population of interest. Therefore, each estimate must be adjusted to include the portion of the population not covered. In addition, the estimate from each method is subject to substantial uncertainty and bias, both of which are difficult to quantify.

We summarize these estimation procedures in Sections 1–3. For each procedure, we evaluate the uncertainty and the bias, including the adjustments needed to include the population not covered. Section 4 summarizes the strengths and weaknesses of these methods, and Section 5 summarizes the problems that must be faced in order to make future estimates of HIV prevalence. Detailed results and supporting data are in a technical report.⁴

The prevalence estimates, by estimation procedure and sex, and in Table I. Each estimate is presented as a plausible range. As a result of the sources of uncertainty and adjustments for the population not covered in each of the statistical estimation procedures, formal confidence intervals were not computed for any of these estimates. Each plausible range is the range of one or more formal 95 per cent confidence interval(s) obtained from the statistical procedure applied to the population covered adjusted to include additional persons not covered by the procedure. The adjustment was based on either a point estimate of the prevalence in the population not covered or on an estimate of the proportion of all infected persons who were in the population covered.

2. SEROPREVALENCE SURVEY IN CHILDBEARING WOMEN

The Survey in Childbearing Women (SCBW) was an anonymous, unlinked serosurvey that measured the prevalence of HIV infection among women giving birth to live infants in the United States.⁵ Residual dried blood specimens collected from newborns for routine metabolic screening were tested for HIV antibody by enzyme immunoassay and Western blot. Since a positive test can result from the presence of maternal antibody, a positive test indicates that the mother was infected. Before testing, personal identifiers except for mother's age, race/ethnicity, and county of residence were removed from the sample. In 1993, 44 states, the District of Columbia, and Puerto

Rico participated in the SCBW; for simplicity, we refer to all of these as 'states'. The survey was conducted through the year except for some large states, where it was conducted for at least 3 months.

The SCBW was considered a population-based survey because nearly all infants born during the survey period in each state were screened for metabolic disorders. Therefore, if the HIV prevalence rates are the same among women who do and who do not give birth (after suitable stratification by age, race/ethnicity, and location of residence), data from the SCBW can be used to estimate the proportion of all women of childbearing age who are infected with HIV.

Let p_{syk} and N_{syk} be the proportion of infected women and number of women of childbearing age, respectively, in state s during year y in demographic stratum k . Then the estimated number of HIV-infected women of childbearing age in this state based on data from Y years is

$$P_s = \frac{1}{Y} \sum_y \sum_k N_{syk} p_{syk}. \quad (1)$$

We estimated prevalence by averaging annual prevalence estimates over 1991, 1992 and 1993 in order to obtain a more stable estimate. This procedure is most appropriate if prevalence is not changing rapidly. We describe a procedure for obtaining a confidence interval for (1) below.

2.1. Adjustments for populations not covered by the survey

The estimate in (1) could be extended to all adult and adolescent women if we knew the proportion of infected women who are of childbearing age (assumed to be 15–44 years). Since this proportion is unknown, we approximated it by the corresponding ratio among women recently diagnosed with acquired immunodeficiency syndrome (AIDS). Similarly, we estimated the number of infected adult and adolescent men by multiplying the estimated number of infected women by the male-to-female AIDS incidence ratio among persons recently diagnosed with AIDS. We used this ratio as a heuristic approximation to the corresponding HIV prevalence ratio, with no implication that it describes a causal relation.

Other data (described below) show that birth rates among HIV-infected women decline dramatically after they are diagnosed with an AIDS-defining opportunistic illness (AIDS-OI). As a result, estimates based on the SCBW do not include most infected persons already diagnosed with an AIDS-OI. Therefore, we estimated sex-specific HIV prevalence by adding to these estimates sex-specific estimates of the prevalence of HIV-infected persons already diagnosed with an AIDS-OI. We obtained these latter estimates from AIDS surveillance data.

2.2. Evaluation of bias

The most important potential source of bias in HIV prevalence estimates based on data from the SCBW is an association between HIV infection and the probability that a woman of childbearing age gives birth. Let f be the birth rate among HIV-uninfected women divided by the birth rate among HIV-infected women. If $f < 1$, then uninfected women were underrepresented in the SCBW, so use of the SCBW data would tend to yield a prevalence estimate that is too large. The converse would be true if $f > 1$.

In fact, the relative bias resulting from such an association can be quantified. For a particular time period and demographic stratum, let p and p_B be the proportion of HIV-infected women

among all women and among women giving birth, respectively. Then it can be shown that

$$p = (f + p(1 - f))p_B$$

and hence

$$p = fp_B(1 + p_B(1 - f) + O(p_B^2(1 - f)^2)).$$

Since p_B is almost always less than 0.05, $p \doteq fp_B$. Note that the approximation is especially good if the ratio of birth rates f is approximately 1, and that the leading term in the relative error shows that p is closer to p_B than the approximation $p \doteq fp_B$ suggests.

Data on birth rates among HIV-infected women are available from cohort and surveillance studies. To compare these rates with birth rates among all women of the same race/ethnicity and age group in the same geographic area, we computed directly standardized rates, using as the standard the population distribution of all women among the geographic areas included in a particular study. In each study we considered the adjusted birth rates among HIV-infected women ages 25 to 44 years without AIDS-OIs are similar to birth rates among all women in that geographic area of the same race/ethnicity and age group.⁴ Birth rates tend to be somewhat higher among HIV-infected non-Hispanic black women ages 15 to 24 years than among all black women of that age. However, most infected women are older than 24 years; the proportion infected among women ages 15 to 24 years who are infected is substantially smaller than the corresponding proportion among women ages 25 to 44 years,⁶ and during 1991–1993, the median age at which adult and adolescent women were diagnosed with AIDS-OIs was 35 years (CDC, unpublished data). Thus, an association between birth rates and HIV infection among women without an AIDS-OIs is unlikely to be a serious source of bias in estimates derived from the SCBW. Because birth rates may be lower among infected than among uninfected women ages 25–44 years,⁴ we cannot determine whether our estimation procedure is more likely to underestimate or overestimate prevalence.

In contrast, birth rates are much lower among HIV-infected women who have developed an AIDS-OI than among infected women who have not yet developed an AIDS-OI. In one large study, the pregnancy rate among women with an AIDS-OI was 60 per cent lower than among HIV-infected women without an AIDS-OI (95 per cent confidence interval, 40–80 per cent lower), after adjustment for prognostic factors.⁷ Because extensive data on birth rates among women with AIDS-OIs are not available, we did not attempt to estimate the extent to which these women were represented in the SCBW. Instead, we added an estimate of the prevalence of HIV-infected women with an AIDS-OI already diagnosed to the estimate obtained from the SCBW. To estimate this prevalence, we used time-specific estimates of the number of AIDS-OI diagnoses⁸ and the estimated distribution of survival time after AIDS-OI diagnosis, both based on data from AIDS surveillance.

The other major possible source of bias is the use of AIDS incidence ratios to approximate HIV prevalence ratios. Because the HIV epidemic has different characteristics in different geographic regions, we estimated these ratios for each of five broad regions defined by the U.S. census. We could not stratify on risk group (no risk information is available in the SCBW) nor on age or race (the corresponding data are not available for many states⁴). We used regions instead of individual states in order to obtain stable ratios. Within each region, approximately 80–85 per cent of AIDS cases among adult and adolescent women are diagnosed in women at ages 15 to 44 years, and there has been little change in this proportion during the last 5 years.⁴ Therefore, using this ratio to approximate the corresponding HIV prevalence ratio is unlikely to yield a substantial bias in the HIV prevalence estimate for women. The male-to-female AIDS incidence ratio ranges from

approximately 3 in the Northeast to more than 10 in the West. Within each region except Puerto Rico and the U.S. territories, this ratio has been declining over time. Therefore, using this ratio to approximate the male-to-female HIV prevalence ratio may overestimate the number of infected men and may cause substantial bias in the estimated HIV prevalence among men. However, the plausible range of 550,000–700,000 infected men obtained from this procedure is very similar to the plausible range of 525,000–750,000 obtained from back-calculation (Table I).

Erroneous HIV test results would also cause bias in our estimates. Because the proportion of HIV-infected childbearing women is small (approximately 1.5 per 1000), false-positive test result could result in overestimating HIV prevalence among women. Both the sensitivity and specificity are greater than 0.998 for the enzyme immunoassay and both are 0.996 for the Western blot,⁹ the two steps in the HIV test procedure. As a result, the proportions of all positive tests that are false-positives are likely to be less than 8 per 1000 in most states, and less than 2 per 1000 in the status with the highest HIV prevalence rates.⁴ Therefore, test errors are unlikely to be an important source of bias in our estimates.

2.3. Estimation of uncertainty

It would be very difficult to estimate the uncertainty in the prevalence estimates caused by the adjustments discussed in the previous section. As a result, our plausible range for HIV prevalence estimated from the SCBW includes only the uncertainty in the estimate obtained from equation (1). We estimated the variance V of the estimate from (1) under the assumption that the number of positive tests in the strata defined by year and the demographic factors have independent Poisson distributions. Assuming that the number of positive tests has a binomial or a hypergeometric distribution would give a variance estimate that is only slightly smaller because the birth rates and prevalence rates are generally less than 15 per cent and 5 per cent, respectively.

The upper limit of the resulting plausible range is likely to be too small because the estimated variance V does not include the uncertainty in strata with no observed infected women. We calculated an upper bound U for the point estimate by replacing the seroprevalence p_{syk} in each stratum in equation (1) that had no observed infected women with the observed proportion of infected women among all women of the same race/ethnicity in the same state during the Y years used to estimate prevalence. If data on race/ethnicity were not provided, p_{syk} in such a stratum was replaced by the observed proportion infected among all women in the same state during that period. Stratification on race/ethnicity but not age was used in this imputation procedure because relatively few states provided data by age and because, for those states that provided information on age, use of data by race/ethnicity had a much greater effect on prevalence estimates than did use of data by age.⁴

Because the estimate P from (1) is a sum of bounded random variables, the distribution of P is approximately Gaussian. Therefore we computed a plausible range for seroprevalence in each region as

$$P - 1.96\sqrt{V} \text{ to } U + 1.96\sqrt{V}.$$

This is a 95 per cent confidence interval for the point estimate P from (1), with the upper bound shifted upward by $U - P$.

We used data from several states, each of which had detailed demographic data and substantial numbers of positive tests, and two procedures to evaluate the effect of estimating the variance V under the independent Poisson assumption described above. First, we fit Poisson regression

models with a dispersion parameter¹⁰ to annual data by race for selected states. These models assume that the variance for the number of HIV-infected mothers in each stratum in each year is a constant α times the expected number in that stratum in that year. Therefore, if α is estimated to be 2.0, the true standard error is about 40 per cent larger than that obtained from the Poisson assumption; if α is estimated to be 0.5, the standard error is about 30 per cent less. These analyses did not give a consistent pattern of estimates for α , but states with relatively large prevalence rates were more likely to show underdispersion than overdispersion.

Second, we estimated the variance based on a permutation distribution for the prevalence rate within each demographic stratum. This is consistent with our assumption that there was no time trend within each stratum. For many states, data are available by month of delivery, but New York State is the only state with high prevalence rates with data for 12 months per year. We applied the permutation distribution to monthly rates and calculated the variance that would be obtained if we used the observed numbers of births during each month to compute annual prevalence rates for each stratum. For a particular stratum and year, let P be the prevalence, let N be the population total, let n_m be the number tested in month m , and let n_T be the total number tested that year. Let p be the random variable indicating a monthly prevalence rate during the time period corresponding to the permutation distribution. Then

$$\text{var}(P) = N^2 \frac{\sum_{m=1}^{12} n_m^2}{n_T^2} (E[p^2] - E[p]^2)$$

where the expectations are taken over the permutation distribution and hence are averages. Within each race/ethnicity group in New York State, the estimated standard deviation estimated from the permutation distribution was 35–39 per cent smaller than that based on the independent Poisson assumption.

Data that are counts often display overdispersion.¹⁰ Estimates of seroprevalence rates would tend to show underdispersion if a substantial proportion of the women tested in some states represent multiple births by the same women in different years. If estimates of HIV prevalence are made in the future from a survey like the SCBW and the assumption of no trend in prevalence appears valid, the variance of the estimate should be estimated from the permutation distribution, rather than based on a Poisson assumption.

3. NATIONAL HOUSEHOLD SURVEY

The third National Health and Nutrition Examination Survey (NHANES III) evaluated the health and nutritional status of the U.S. household population during 1988–1994. A representative sample of this population was selected using a complex multi-stage area probability sampling procedure. The survey was conducted in 81 randomly selected locations throughout the United States (all 50 states plus the District of Columbia, hereafter referred to as ‘the states’). Of the 39,695 persons ages 2 months and older who were selected for the survey, 31,311 (79 per cent) agreed to the examination component of the survey. An anonymous HIV test was performed on serum samples from persons ages 18 years and older.

Data from this survey were used to obtain a direct estimate of HIV prevalence among persons ages 18–59 years living in households.² Person older than 59 years were eliminated from the analysis because of higher non-response to the examination component resulting from illnesses

that were probably unrelated to HIV infection. The 95 per cent confidence intervals for HIV prevalence obtained directly from NHANES III are 290,000–733,000 for all persons, 250,000–601,000 for males, and 41,000–213,000 for females. The corresponding point estimates are 461,000, 368,000, and 94,000, respectively.²

3.1 Adjustments for populations not covered by the survey

The HIV prevalence estimate from NHANES III does not include persons outside the age range of 18 to 59 years, person living outside the 50 states and the District of Columbia (for example, residents of Puerto Rico), and persons not living in households (including persons who are homeless, in prison, or hospitalized). Therefore, to estimate HIV seroprevalence among all U.S. adults and adolescents, it is necessary to add estimates of HIV seroprevalence among these groups. In addition, it is necessary to estimate the effect of the probable underrepresentation in the survey of three other groups: white men, older black men (ages 40–59 years), and persons previously diagnosed with AIDS-OIs.

We can easily extend the seroprevalence estimates to the first two excluded groups. To account for those not ages 18 to 59, we divided the NHANES III seroprevalence estimate by a proportion computed from AIDS surveillance data, the proportion of the AIDS cases diagnosed at ages older than 12 years that were diagnosed in residents of the 50 states and the District of Columbia at ages 18 to 59 years. This proportion was approximately 0.97 among both men and women during each year, 1992–1994.⁴ To account for residents of Puerto Rico and U.S. territories, we added the HIV prevalence estimates for that region obtained from the SCBW and census data.

Estimating HIV prevalence among persons not living in households is more difficult. Data are available on the number of state and federal prisoners known to be infected with HIV. In addition, the number of prisoners is known, and serosurveys have been conducted in some prison systems. These data suggest that approximately 25,000 persons in state and federal prisons are infected with HIV⁴. We did not have data to estimate HIV prevalence among other incarcerated persons.

Hospitalized persons are included in the sampling frame for NHANES, in that they are eligible to participate if their dwelling unit is sampled. Individuals who reside in households but who were hospitalized at the time of the survey would have been counted as non-respondents. The sample weights correct for non-response but assume that respondents are similar to non-respondents. This assumption is not correct for hospitalized persons because these persons are more likely to be infected with HIV than the general population. Based on estimates of the number of HIV-infected persons hospitalized during 1991, the number of persons discharged from hospital with a diagnosis of HIV, and the average length of hospital stays, we estimated that fewer than 10,000 HIV-infected persons were hospitalized on any given day in 1991.⁴ Therefore, failing to include hospitalized persons is likely to have had relatively little effect on the prevalence estimate obtained from NHANES III.

NHANES III provided partial coverage of homeless persons. For example, shelters for homeless persons could be sampled, but residents were eligible for the survey only if they regarded the shelter as their permanent address. Therefore, HIV prevalence among homeless persons must be estimated from other data sources. Based on results from the 1990 Decennial U.S. Census¹¹ and on results from a random digit dialling telephone survey conducted in 1990 of 1507 adults ages 18 years or older who were asked about homelessness during the past 5 years,¹² it is plausible that approximately 460,000 to 1,000,000 persons are homeless in the United States on any given

day. Comparing results from an HIV seroprevalence survey conducted in clinics for homeless persons in 14 cities with results from seroprevalence surveys in sexually transmitted disease clinics in the same cities¹³ suggests that 4–6 per cent of homeless adults and adolescents are infected with HIV. Based on this information, we estimated that 20,000 to 60,000 homeless persons are living with HIV infection.

3.2 Effect of lower participation among groups with a relatively high probability of infection

Because the proportion of the population infected with HIV is relatively small, non-response can have a substantial effect on the HIV prevalence estimate obtained from a survey if those who do not participate are substantially more likely to be infected with HIV than those who do participate. Serum samples were obtained from 66 per cent of the men and 73 per cent of the women ages 18 to 59 years selected for NHANES III. Participation rates (as defined by obtaining serum for analysis) were lower among white men ages 18 to 39 years (65 per cent) and ages 40–59 years (66 per cent) and among black men ages 40 to 59 years (64 per cent), than among other subgroups defined by sex, race/ethnicity, and age group. Furthermore, male non-respondents in NHANES III differed from respondents in some important demographic characteristics, including the proportion living in single family households. In contrast, female respondents and non-respondents had very similar demographic characteristics.

Results from a household survey designed to estimate HIV prevalence show that non-respondents may be substantially more likely to have a history of behaviour associated with risk for HIV infection than respondents. This survey was conducted in Dallas, Texas, in 1989.¹⁴ A random sample of 184 persons who initially declined to participate were asked to participate in a special follow-up study after the survey was completed. Data on risk behaviour were obtained from somewhat more than half of these persons. A history of injecting drug use since 1978 was reported by 7.0 per cent of the initial non-respondents, compared with 3.1 per cent of those who participated in the survey (among men, 12.2 per cent and 2.8 per cent, respectively). Among men, the proportions reporting a history of a male-to-male sex since 1978 were 16.8 per cent and 5.1 per cent, respectively.

Brookmeyer and Gail¹⁵ point out that such information can be used to estimate the potential bias in the prevalence estimate obtained from a survey. Let f be the proportion of non-responders, let π_1 be the prevalence rate among the responders, and let ρ be the relative risk of infection among non-responders compared with responders. Then the prevalence rate in the population is

$$\pi = (1 - f)\pi_1 + f\pi_1\rho.$$

We assumed that ρ does not vary among male racial/ethnic groups.

It is possible that the relative risk ρ is smaller among persons surveyed in NHANES III than in the Dallas study. Because the Dallas study was specifically an HIV seroprevalence survey rather than a general health survey, infected men might have been more likely to refuse to participate in that study than in NHANES III. Although data from the Dallas survey do not yield a direct estimate of ρ , these behavioural data suggest that this relative risk was at least 3 among men. Assuming that this relative risk was 1.5 and 2.5 in NHANES III increases the point estimate of HIV prevalence among men by 63,000 and 190,000 persons, respectively. Thus the estimate is quite sensitive to the value of this relative risk.

Refusal or inability to participate by persons with advanced HIV disease would also cause a serious bias in the HIV prevalence estimate obtained from NHANES III. Persons with serious

illnesses are less likely to participate in NHANES than are healthy persons.¹⁶ Of the 29 HIV-infected persons who participated in the first half of NHANES III, three reported taking zidovudine (CDC, unpublished data), so some persons being treated for HIV disease did participate. At the end of 1991, approximately 95,000 living HIV-infected persons had been diagnosed with AIDS-OIs.¹⁷

We conclude that lower participation among person with relatively high probability of HIV infection could have reduced the HIV prevalence estimate obtained from NHANES III by 100,000 to 200,000 persons, mostly men. The plausible range obtained from NHANES III for HIV prevalence among men is 300,000 to 750,000, after adjustment for those not covered by the survey (Table I). Adjustment for the bias associated with lower participation suggests a more likely range of 400,000 to 950,000 infected men, which includes the ranges obtained from back-calculation (525,000 to 750,000) and from the SCBW (550,000 to 700,000).

4. BACK-CALCULATION

Back-calculation can be used to estimate the number of HIV-infected persons, and the distribution of times at which they became infected, required to account for observed AIDS diagnoses.¹⁸ Let $a(t)$ be the number of AIDS cases diagnosed at time t (obtained from AIDS surveillance). Let $f(t)$ be the probability density function of the time from HIV infection to AIDS diagnosis (the incubation period distribution, estimated from cohort studies of HIV-infected persons). Choose $t = 0$ to be an estimated date for the start of the HIV epidemic (for example, January 1977). Let $i(t)$ be the (unknown) number of persons who were infected at time t . Back-calculation is a statistical procedure for finding an estimate of $i(t)$ to solve the integral equation

$$a(t) = \int_0^t f(t-s)i(s) ds.$$

Note that the estimate of $i(t)$ does not include infected persons who are never reported as having AIDS, either because their AIDS diagnosis is never reported or because they die before AIDS is diagnosed. Therefore we estimate HIV prevalence at time t as

$$H(t) = (I(t) - D(t))/R$$

where $I(t)$ is the cumulative number of infections at time t (obtained from the estimate of $i(t)$), $D(t)$ is the cumulative number of deaths among person reported with AIDS (obtained from AIDS surveillance), and R is an estimate of the proportion of HIV-infected persons who will eventually be reported to AIDS surveillance. Note that we assume that R has not changed over time.

Because it is the only method that can be used to reconstruct the history of the HIV epidemic, statisticians have investigated back-calculation very thoroughly. Bacchetti *et al.*³ provide an excellent summary of issues related to back-calculation. We summarize below the important issues concerning use of back-calculation with AIDS surveillance data.

4.1 AIDS surveillance data

Before AIDS incidence data can be used in back-calculation or cumulative deaths can be used in estimating prevalence, AIDS incidence and deaths must be adjusted for reporting delays. Several adjustment techniques have been proposed.¹⁹⁻²² These methods tend to give similar results. It may be desirable to remove seasonal effects in AIDS diagnosis dates.²³

Because the HIV epidemic in the United States is heterogeneous and prevalence estimates are desired for subgroups, back-calculation is usually carried out on data from subgroups defined by demographic and behavioural risk factors. If estimates are made by mode of HIV transmission, it is necessary to consider how to treat cases with no identified risk (NIR). We apportion these cases among the known modes of transmission by using proportions estimated from cases originally reported as NIR for which a risk was determined after further investigation. The greatest effect of this procedure is on AIDS incidence among persons infected with HIV through heterosexual contact,²⁴ the mode of transmission which has shown in the greatest proportional increase in AIDS-OI incidence.²⁵

4.2 Choice of AIDS surveillance data

All 50 states, the District of Columbia, Puerto Rico, and the U.S. territories report AIDS cases to CDC using a uniform surveillance definition. Before 1993, the surveillance definition was based only on AIDS-OIs.²⁶ In January 1993, the definition for adults and adolescents was expanded to include HIV-infected persons with CD4+ T-lymphocyte counts of less than 200 cells/ μ l, a CD4+ percentage of less than 14, or one of three additional AIDS-OIs.²⁷ A large proportion of AIDS cases diagnosed after 1992 were reported based on the CD4+ criteria.²⁸ Later diagnoses of AIDS-OIs are not reported for most persons reported to AIDS surveillance under the CD4+ criteria.

As a result, in the back-calculation analyses we used only those AIDS cases for which an AIDS-OI diagnosis made before 1993 was known. The available estimates of the incubation period distribution are based on the time from HIV infection to AIDS-OI diagnosis. To use all AIDS cases in the back-calculation analysis, we would have to estimate this distribution for the time to AIDS diagnosis under the 1993 definition. It would be very difficult to make such an estimate because HIV and CD4+ testing practices have changed over time and also vary among geographic regions.

Periods when recent AIDS incidence has been affected by changes in the surveillance definition should not be used because HIV incidence estimates can be sensitive to changes in the pattern of recent AIDS incidence. For example, Hay and Wolk²⁹ present estimates of cumulative HIV incidence from back-calculation using AIDS cases diagnosed through two different dates, September 1987 and June 1990 (their Figures 1 and 2). The point estimates of the cumulative number of infections agree very well through 1984. The point estimates based on the short sample continue to increase linearly after 1984, implying constant HIV incidence. In contrast, the point estimates based on the long sample have a sharp decrease in slope at approximately the start of 1985, implying a substantial decrease in HIV incidence at that time. Rosenberg and Gail³⁰ obtained similar results. These results demonstrate the inability of back-calculation to detect the decrease in HIV incidence using a period before the change in HIV incidence was reflected in AIDS incidence.

4.3 Infection time distribution

Because estimates of recent HIV incidence obtained from back-calculation are imprecise, the parametrization of the HIV infection time distribution can substantially affect the resulting HIV prevalence estimates. Most current back-calculation analyses use step functions as a flexible parametric model for the infection time distribution.^{3, 18, 29, 31} Some analyses use a distribution with short (1–3 month) steps and a penalized maximum likelihood estimation procedure to

obtain a smooth infection time distribution.³ These analyses attempt to estimate the recent trend in HIV incidence, but the estimates are sensitive to the pattern of recent AIDS-OI incidence, as discussed above. Rosenberg and his colleagues used fewer steps, typically 5, with the last step covering approximately 5 years. These choices, which are based on the mean squared error from simulation studies,³² result in estimates of recent HIV incidence that are less sensitive to other assumptions in the back-calculation model. Such a model only attempts to estimate the average number of infections per year for the last 5 years, thus providing no information about trends in HIV incidence during this period.

4.4 Choice of the incubation period distribution

The incubation period distribution has been estimated from a number of cohorts. In general, these analyses found that the median time from HIV infection to AIDS-OI diagnosis is 8 to 10 years and that only 3–10 per cent of HIV-infected persons develop an AIDS-OI within 3 years after infection. The incubation period is affected by age, with longer incubation periods among younger persons (see endnote 8 of reference 31 for citations). The incubation period is longer among persons with haemophilia than among other infected persons.³³

The use of prophylactic therapy, which can lengthen the incubation period, makes our knowledge of the incubation period distribution in the population less certain. It is only possible to estimate this distribution among untreated persons for approximately the first 7 years after infection. Prophylactic therapy can reduce the hazard for AIDS-OI diagnosis among infected persons with a CD4+ count less than 200 cells/ μ l by upto 65 per cent,³⁴ but the therapeutic effect is attenuated within approximately 1 year for therapies available during the early 1990s.³⁵ There are no reliable national data on the proportion of HIV-infected persons who used therapy before AIDS-OI diagnosis in the early 1990s.

Uncertainty in the incubation period distribution is especially important because HIV incidence estimates obtained from back-calculation, and hence prevalence estimates, can be quite sensitive to the choice of this distribution. For example, Bacchetti *et al.*³ used four different incubation period distributions to estimate HIV incidence through 1990, based on AIDS cases diagnosed through 1990. The estimates of HIV incidence vary substantially, especially the estimates for 1987–1990 (see their Figure 1). Two analyses estimated that there were more than 100,000 new infections per year in the United States during 1989 and 1990, while the other two estimated that fewer than 15,000 new infections occurred per year during that period. These results demonstrate not only the sensitivity of back-calculation estimates to the incubation period distribution, but also the importance of verifying that back-calculation estimates are consistent with data from seroprevalence studies. CDC's national seroprevalence studies generally show roughly constant seroprevalence rates from 1990 to 1993.^{36–38} Because there were approximately 35,000 to 45,000 deaths per year during that period among persons with AIDS diagnosed, HIV incidence of approximately 15,000 or 100,000 persons per year is not consistent with constant prevalence.

4.5 Adjustments for incomplete reporting

Evaluations of the AIDS surveillance system suggested that approximately 90 per cent of AIDS cases diagnosed during 1987–1989 were reported to CDC.^{39,40} Similar estimates for cases diagnosed during 1990–1992 are not available. It is likely that reporting of AIDS cases was less complete earlier during the 1980s, because the surveillance system was not as well developed.

It is difficult to estimate the proportion of HIV-infected persons who die before being diagnosed with AIDS-OIs, but data are available on persons in medical care for HIV-infection. Using a competing risks procedure, we estimated that approximately 5 per cent of men who have sex with men (MSWM) and 8 per cent of injecting drug users (IDUs) who are receiving medical care for HIV-infection and have a CD4+ count less than 200 cell/ μ l will die before developing an AIDS-OI.⁸ The risk for death before AIDS-OI diagnosis among all IDUs is greater than 8 per cent, as IDUs have a high risk for death before AIDS-OI diagnosis.⁴¹

An alternative estimate of the proportion of HIV-infected persons whose deaths are reported to AIDS surveillance is based on modelling trends in death rates for deaths due to causes associated with HIV infection. These analyses suggest that, among HIV-infected persons 25–44 years of age, 70–90 per cent of deaths during 1987 among men⁴² and at least 70 per cent of deaths during 1988 among women⁴³ were reported to AIDS surveillance. Thus, there is substantial uncertainty in the proportion R in equation (3), a proportion which has undoubtedly changed over time. This increases our uncertainty in the HIV prevalence estimates based on back-calculation.

5. DISCUSSION

In theory, the best way to estimate the prevalence of HIV infection in the United States would be a national survey designed for this purpose. However, a pilot study conducted in Dallas, Texas, demonstrated that HIV prevalence could not be estimated from such a national household survey as a result of non-response bias.¹⁴

As a result, HIV prevalence can only be estimated from data collected for other purposes. We have summarized the use of three distinct types of data and estimation procedures. Each procedure yields an estimate that must be adjusted to include the estimated prevalence of HIV infection in persons not included in the population to which the estimate applies. Although each method has particular advantages, no individual method can provide an accurate or precise estimate of HIV prevalence.

NHANES III included a representative sample of all adult U.S. household residents from all 50 states and the District of Columbia. However, because the primary purpose of this survey was to estimate the prevalence of conditions substantially more common than that of HIV infection, the estimates of HIV prevalence are relatively imprecise. In addition, in accordance with the confidentiality agreement for this survey, information on risk behaviour leading to HIV infection was not collected, so prevalence cannot be estimated within risk groups. There are three potential sources of substantial bias in this survey. One source is persons not covered by the sampling frame, especially prisoners and some homeless persons. A second source is persons with advanced HIV-related disease whose poor health or hospitalization would result in refusal to participate. Because it is possible to estimate the numbers of HIV-infected person in these two groups, they may be less important sources of bias than the third source, persons who refused to participate for other reasons. Data suggest that there could be substantial bias resulting from the refusal of men to participate. The particular strength of NHANES III, in addition to providing the only data on HIV infection in a representative sample of the U.S. population, is that key data on non-response were collected in the survey. This survey is the only data source that can be used to estimate HIV prevalence in which some of the data needed to adjust for bias were collected.

An alternative method for estimating HIV prevalence from HIV testing data would be to use data on HIV prevalence rates within demographic or behavioural risk groups and estimates of the numbers of persons in these groups. It is not feasible to do this for behavioural risk groups

because it is not possible to make precise estimates of the sizes of the most important groups (MSWM, IDUs, and persons at risk for infection through heterosexual contact). There is also substantial variation in prevalence rates among geographic areas.³⁸ Finally, the prevalence rates observed in specific HIV testing settings may not be representative of all persons in that risk group (for example, MSWM tested at public sexually transmitted disease clinics need not be representative of all MSWM).

The best seroprevalence data for estimating HIV prevalence are those from the SCBW. These data, along with corresponding population counts, yield an estimate of HIV prevalence among women of childbearing age. We extended this estimate to all adult and adolescent women by using data from AIDS surveillance to estimate the proportion of all infected adult and adolescent women who are of childbearing age. The resulting estimate will be biased unless HIV prevalence rates are the same (within demographic strata defined by geographic location, race/ethnicity, and age) among women of childbearing age who do and do not give birth. Data from several studies show that this equality is roughly true, except that birth rates are lower among HIV-infected women who have developed an AIDS-OI. We corrected this bias by adding the estimated prevalence of HIV-infected women with a diagnosed AIDS-OI to the original estimate.

The prevalence estimate from the SCBW is the only such estimate based on a population-based survey in which many HIV-infected persons are detected. However, it is difficult to estimate the uncertainty in this prevalence estimate because we do not know the relative HIV prevalences among women who do and do not give birth (or equivalently, the relative fertility rates in women who are and are not infected with HIV). Data from the SCBW cannot be used to obtain prevalence estimates by risk and by race/ethnicity because data on behavioural risk are not collected, and data on race/ethnicity are not available from all states. Finally, data from the SCBW do not yield an estimate for HIV prevalence among men. To make such an estimate, we used an *ad hoc* procedure to estimate the male-to-female HIV prevalence ratio by the corresponding ratio among recently diagnosed AIDS cases.

Back-calculation is an attractive method for estimating HIV prevalence because it is based on knowledge of the HIV disease process and on a very large number of diagnosed AIDS cases. As a result of the latter, HIV prevalence estimates can be made within groups defined by demographic and behavioural risk characteristics. However, prevalence estimates obtained from back-calculation must be evaluated carefully. These estimates are sensitive to the incubation period distribution used in the model. Because back-calculation uses a sophisticated statistical estimation procedure, the estimates may also be sensitive to whether a number of other underlying assumptions are satisfied. Prevalence estimates derived from back-calculation depend on estimates of recent HIV incidence and the proportion of HIV-infected persons who are never reported as AIDS cases, neither of which can be estimated precisely.

Each of these methods for estimating HIV prevalence has distinct strengths and weaknesses. In particular, for each method, substantial adjustments to an estimate based on a well-defined statistical procedure must be made in order to include HIV-infected persons in populations not covered by the data source on which the estimate is based. For each method, we can estimate the variance of the prevalence estimate derived from the statistical model under the assumptions on which the model is based. However, it is difficult to evaluate the validity of these assumptions for estimates based on the SCBW and on back-calculation. The need to estimate HIV prevalence in the population not covered by the data on which the model is based adds further uncertainty.

Thus, there is substantial uncertainty associated with the HIV prevalence estimate obtained from each of these methods, and this uncertainty cannot be greatly reduced by more sophisticated statistical models or better data.

6. PROSPECTS FOR THE FUTURE

Modifications of the methods we have used, or the development of new methods, will likely be required to make future estimates of HIV prevalence. It should be possible to use data from the next NHANES, which is currently scheduled to begin in 1998 and is expected to test participants for HIV infection. However, it is likely that our methods for estimating prevalence from SCBW data and from back-calculation would have to be modified.

The SCBW was suspended in May 1995 in order to evaluate how HIV testing should be conducted in order to help prevent HIV transmission from infected mothers to their infants. The Survey may be resumed in selected areas. If it is resumed, there will be other challenges in using these data to estimate HIV prevalence, in addition to the reduced geographic coverage. The prevalence rate decreased among childbearing women in New York City from 1989 to 1994.⁴⁴ It is not known what caused this decrease; possible explanations are mortality, an increase in the proportion of HIV-infected women who have advanced disease (an AIDS-OI), and a change in the proportion of infected women who become pregnant or terminate a pregnancy. However, with a maturing epidemic, similar trends are likely to be seen elsewhere and would need to be incorporated into the estimation procedure. In addition, the proportion of HIV-infected women of childbearing age who know that they are infected (before or during pregnancy) is likely to increase. There already is widespread use of medical therapy that reduces the probability that an infected woman will transmit her infection to her baby. Both of these factors may affect an HIV-infected woman's decision to become pregnant, or, if pregnant, to terminate the pregnancy. Such a decision affects the relative birth rate among uninfected women compared to that among infected women and hence also affects the bias in the estimate derived from SCBW data.

As a result of the change in the AIDS surveillance definition in 1993, the current surveillance definition is inconsistent with available estimates of the incubation period distribution. It is natural to consider back-calculation based on deaths among persons with AIDS diagnosed³ or on the estimated incidence of AIDS-OIs.⁸ There are problems with either alternative approach. The expansion of the surveillance definition in 1993 caused two artifacts in AIDS surveillance data: recorded deaths now include some deaths before AIDS-OIs are diagnosed, and estimated AIDS-OI incidence increased relatively rapidly during 1991–1992.⁸ In addition, the number of deaths among persons with AIDS diagnosed began to fall dramatically during 1996, most likely as a result of therapeutic advances,¹⁷ and widespread use of protease inhibitors beginning in 1996 is likely to affect both AIDS incidence and deaths. The effects of recent therapeutic advances would need to be modelled in order to use recent surveillance data in back-calculation, but the required estimates are not likely to be available soon.

As a result of these problems, estimating HIV prevalence will be even more challenging in the future. Obtaining reliable estimates will require that data of high quality continue to be available from epidemiologic studies and surveillance systems. Because each method for estimating HIV prevalence produces an estimate that has substantial uncertainty and that must be corrected for bias, reliable estimates can be obtained only by synthesizing the estimates from several procedures and by considering relevant epidemiologic data.

ACKNOWLEDGEMENTS

We thank Dan Rosen, M. P. H., for computational assistance and James W. Curran, M. D., M. P. H., Geraldine McQuillan, Ph.D., and an anonymous referee for helpful suggestions.

REFERENCES

1. Karon, J. M., Rosenberg, P. S., McQuillan, G., Khare, M., Gwinn, M. and Petersen, L. R. 'Prevalence of HIV infection in the United States, 1985–1992', *Journal of the American Medical Association*, **276**, 126–131 (1996).
2. McQuillan, G. M., Khare, M., Karon, J. M., Schable, C. A. and Vlahov, D. 'Update on the seroepidemiology of human immunodeficiency virus in the United States household population: NHANES III, 1988–1994', *Journal of Acquired Immune Deficiency Syndromes*, **14**, 355–360 (1997).
3. Bacchetti, P., Segal, M. R. and Jewell, N. P. 'Backcalculation of HIV infection rates (with discussion)', *Statistical Science*, **8**, 82–119 (1993).
4. Karon, J. M. *Methods for Estimating HIV Prevalence in the United States*, Publication D821, National AIDS Clearinghouse, U.S. Department of Health and Human Services, Washington D.C., 1996.
5. Gwinn, M., Pappaioanou, M., George, J. R., Hannon, W. H. Wasser, S. C., Redus, M. A., Hoff, R., Grady, G. F., Willoughby, A., Novello, A. C., Petersen, L. R., Dondero, T. J. and Curran, J. W. 'Prevalence of HIV infection in childbearing women in the United States: surveillance using newborn blood samples', *Journal of the American Medical Association*, **265**, 1704–1708 (1991).
6. St. Louis, M. E., Petersen, L. R., Anderson, J. E., Schable, C. A. and Dondero, T. J. 'Seroprevalence rates of Human Immunodeficiency Virus infection at sentinel hospitals in the United States', *New England Journal of Medicine*, **323**, 213–218 (1990).
7. Chu, S. Y., Hanson, D. L. and Jones, J. L. 'Pregnancy rates among women infected with human immunodeficiency virus', *Obstetrics and Gynecology*, **87**, 195–198 (1996).
8. Karon, J. M., Green, T. A., Hanson, D. L. and Ward, J. W. 'Estimating the number of AIDS-defining opportunistic illness diagnoses from data collected under the 1993 AIDS surveillance definition', *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **16**, in press (1997).
9. Schochetman, G. and George, J. R. *AIDS Testing. A Comprehensive Guide to Technical, Medical, Social, Legal, and Management Issues*, Springer-Verlag, New York, 1994.
10. McCullagh, P. and Nelder, J. A. *Generalized Linear Models*, Chapman and Hall, London, 1989.
11. Bureau of the Census. 'Fact sheet for 1990 decennial census counts of persons in selected locations where homeless persons are found', Report CPH-L-87 (no date given).
12. Link, B. G., Susser, E., Stueve, A., Phelan, J., Moore, R. E. and Struening, E. 'Lifetime and five-year prevalence of homelessness in the United States', *American Journal of Public Health*, **84**, 1907–1912 (1994).
13. Allen, D. M., Lehman, J. S., Green, T. A., Lindgren, M. L., Onorato, I. M. and Forrester, W. 'HIV infection among homeless adults and runaway youth, United States, 1989–1992', *AIDS*, **8**, 1593–1598 (1994).
14. Centers for Disease Control. 'Pilot study of a household survey to determine HIV seroprevalence', *Morbidity and Mortality Weekly Report*, **40**, 1–5 (1991).
15. Brookmeyer, R. and Gail, M. H. *AIDS Epidemiology: a Quantitative Approach*, Oxford University Press, New York, 1994.
16. Ezzati, T. and Khare, M. 'Nonresponse adjustments in a national health survey', 1992 Proceedings of the Survey Research Methods Section, American Statistical Association, 339–344 (1993).
17. Centers for Disease Control and Prevention. 'Update: trends in AIDS incidence, deaths, and prevalence – United States, 1996', *Morbidity and Mortality Weekly Report*, **46**, 165–173 (1997).
18. Brookmeyer, R. 'Reconstruction and future trends of the AIDS epidemic in the United States', *Science*, **253**, 37–42 (1991).
19. Karon, J. M., Devine, O. J. and Morgan, W. M. 'Predicting AIDS incidence by extrapolating from recent trends', in Castillo-Chavez, C., *Mathematical and Statistical Approaches to AIDS Epidemiology*, Lecture Notes in Biomathematics, vol. 83, Springer-Verlag, Berlin, 1989, pp. 58–88.
20. Rosenberg, P. S. 'A simple correction of AIDS surveillance data for reporting delays', *Journal of Acquired Immune Deficiency Syndromes*, **3**, 49–54 (1990).
21. Brookmeyer, R. and Liao, J. 'The analysis of delays in disease reporting: methods and results for the acquired immunodeficiency syndrome', *American Journal of Epidemiology*, **132**, 355–365 (1990).

22. Bacchetti, P. 'The impact of lengthening AIDS reporting delays and uncertainty about underreporting on incidence trends and projections', *Journal of Acquired Immune Deficiency Syndromes*, **7**, 860–865 (1994).
23. Bacchetti, P. 'Seasonal and other short-term influences on United States AIDS incidence', *Statistics in Medicine*, **13**, 1921–1931 (1994).
24. Green, T. A. 'Using surveillance data to monitor trends in the AIDS epidemic', *Statistics in Medicine*, **17**, 143–154 (1998).
25. Centers for Disease Control and Prevention. 'HIV/AIDS Surveillance Report', 8 (no. 2), 1–39 (1996).
26. Centers for Disease Control. 'Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome', *Morbidity and Mortality Weekly Report*, **36**, 1S–15S (1987).
27. Centers for Disease Control and Prevention. '1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults', *Morbidity and Mortality Weekly Report*, **41**, 1–19 (1992).
28. Centers for Disease Control and Prevention. 'Update: Impact of the expanded AIDS surveillance case definition for adolescents and adults on case reporting – United States, 1993'. *Morbidity and Mortality Weekly Report*, **43**, 160–161, 167–170 (1994).
29. Hay, J. W. and Wolak, F. A. 'A procedure for estimating the unconditional cumulative incidence curve and its variability for the human immunodeficiency virus', *Applied Statistics*, **43**, 599–624 (1994).
30. Rosenberg, P. S. and Gail, M. H. 'Uncertainty in estimates of HIV prevalence derived by backcalculation', *Annals of Epidemiology*, **1**, 105–115 (1990).
31. Rosenberg, P. S. 'Scope of the AIDS epidemic in the United States', *Science*, **270**, 1372–1375 (1995).
32. Rosenberg, P. S., Gail, M. H. and Pee, D. 'Mean square error of estimates of HIV prevalence and short-term AIDS projections derived by back-calculation', *Statistics in Medicine*, **10**, 1167–1180 (1991).
33. Rosenberg, P. S., Goedert, J. J. and Biggar, R. J. 'Effect of age at seroconversion on the natural AIDS incubation distribution', *AIDS*, **8**, 803–810 (1994).
34. Longini, I. M., Clark, W. S. and Karon, J. M. 'The effect of routine use of therapy in slowing the clinical course of human immunodeficiency virus (HIV) infection in a population-based cohort', *American Journal of Epidemiology*, **137**, 1229–1240 (1993).
35. Hoover, D. R. 'The effects of long term zidovudine therapy and Pneumocystis carinii prophylaxis on HIV disease. A review of the literature', *Drugs*, **49**, 20–36 (1995).
36. Weinstock, H. S., Sindhu, J., Gwinn, M., Karon, J. and Petersen, L. R. 'Trends in HIV seroprevalence among persons attending sexually transmitted disease clinics in the United States, 1988–1992', *Journal of Acquired Immune Deficiency Syndromes*, **9**, 514–522 (1995).
37. Prevots, D. R., Allen, D. M., Lehman, J. S., Green, T. A., Petersen, L. R. and Gwinn, M. 'Trends in HIV seroprevalence among injection drug users entering drug treatment centers, United States, 1988–1993', *American Journal of Epidemiology*, **143**, 733–742 (1996).
38. Centers for Disease Control and Prevention. 'National HIV Serosurveillance Summary: Results through 1992, Vol. 3', U.S. Department of Health and Human Services, Atlanta, GA, Report HIV/NCID/11-93/036, 1994.
39. Buehler, J. W., Berkelman, R. L. and Stehr-Green, J. K. 'The completeness of AIDS surveillance', *Journal of Acquired Immune Deficiency Syndromes*, **5**, 257–264 (1992).
40. Rosenblum, L., Buehler, J. W., Morgan, W. M., Costa, S., Hidalgo, J., Holmes, R., Lieb, L., Shields, A. and Whyte, B. M. 'The completeness of AIDS case reporting, 1988: a multisite collaborative surveillance project', *American Journal of Public Health*, **82**, 1495–1499 (1992).
41. Stoneburner, R. L., Des Jarlais, D. C., Benezra, D., Gorelkin, L., Sotheran, J. L., Friedman, S. R., Schultz, S., Marmor, M., Mildvan, D. and Maslansky, R. 'A larger spectrum of severe HIV-1 related disease in intravenous drug users in New York City', *Science*, **242**, 916–919 (1988).
42. Buehler, J. W., Devine, O. J., Berkelman, R. L. and Chevarley, F. M. 'Impact of the human immunodeficiency virus epidemic on mortality trends in young men, United States', *American Journal of Public Health*, **80**, 1080–1086 (1990).
43. Buehler, J. W., Hanson, D. L. and Chu, S. Y. 'The reporting of HIV/AIDS deaths in women', *American Journal of Public Health*, **82**, 1500–1505 (1992).
44. Smith, P., Glebatis, D., Pulver, W. and DeBuono, B. 'Trends in HIV seroprevalence among childbearing women in New York State', Abstract 410, 3rd Conference on Retroviruses and Opportunistic Infections, Washington, D. C., 1996.