American Journal of Epidemiology

Vol. 142, No. 7
Printed in U.S.A.

A BRIEF ORIGINAL CONTRIBUTION

# The AIDS Epidemic in India: A New Method for Estimating Current Human Immunodeficiency Virus (HIV) Incidence Rates

Ron Brookmeyer,[1] Thomas Quinn,[2,3] Mary Shepherd,[2] Sanjay Mehendale,[4] Jeanette Rodrigues,[4] and Robert Bollinger [2]

Human immunodeficiency virus (HIV) incidence rates in India were estimated using a new method that accounts for follow-up bias. Follow-up bias arises in epidemiologic cohort studies when the incidence rate among individuals who do and do not return for follow-up are different. The new method combines data on the prevalence of p24 antigenemia among all those initially screened together with the longitudinal follow-up data on the subset of patients who returned for follow-up. Using these methods, the current HIV incidence rate among patients attending sexually transmitted disease clinics in Pune, India, was 18.6% per year. It was found that follow-up bias can cause significant underestimation in HIV incidence rates, perhaps by as much as 60%. These incidence estimates, together with other HIV seroprevalence studies, suggest the HIV epidemic in India is growing rapidly. *Am J Epidemiol* 1995;142:709–13.

acquired immunodeficiency syndrome; bias (epidemiology); cross-sectional studies; epidemiologic methods; HIV; statistics

Current human immunodeficiency virus (HIV) incidence rates are crucial for estimating the current and future magnitude of the acquired immunodeficiency virus (AIDS) epidemic and for targeting intervention programs. Unfortunately, there is a paucity of data on current HIV incidence rates in developing countries in contrast to the considerable data available on HIV seroprevalence. HIV seroprevalence indicates the proportion of individuals infected sometime in the past, while incidence refers to the current risk (hazard) of infection. Estimates of HIV incidence have typically been based on the longitudinal follow-up of a cohort of uninfected individuals. However, attempts at estimating incidence in developing countries with this approach are often stymied by low follow-up rates, cost, and a form of selection bias called "follow-up bias," which arises if the incidence rates among the subgroup of individuals who return for follow-up are different from the incidence rates among those who do not return for follow-up.

Other approaches for estimating HIV incidence have important limitations. For example, back-calculation, which uses numbers of AIDS cases together with the incubation period to reconstruct past HIV infection rates (1) has very limited utility for estimating current incidence rates because of the long incubation period from infection to AIDS, and because of incomplete reporting of AIDS cases, particularly in developing countries. Another approach, based on trends in HIV seroprevalence derived from serial surveys, can produce misleading estimates because of unknown population changes due to migration, deaths from HIV disease, and changing selection biases (2).

Estimates of the scope of the AIDS epidemic in India have been based principally on HIV seroprevalence surveys. HIV incidence rates are exceedingly difficult to obtain in India as in other developing countries because of the problems of assembling cohorts of high risk individuals, low follow-up rates, and follow-up bias.

## MATERIALS AND METHODS

Human immunodeficiency virus type 1 (HIV-1) incidence rates among patients attending two sexually

Received for publication January 6, 1995, and in final form June 7, 1995.

Abbreviations: AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus; HIV-1, human immunodeficiency virus type 1; HIV-2, human immunodeficiency virus type 2.

[1] Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, MD.
[2] Division of Infectious Diseases, The Johns Hopkins University School of Medicine, Baltimore, MD.
[3] National Institute of Allergy and Infectious Diseases, Bethesda, MD.
[4] The National AIDS Research Institute, Pune, India.
Reprint requests to Dr. Ron Brookmeyer, Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205.

transmitted disease clinics in Pune, India, were computed using three different methods. All patients screened between May 1993 and January 1994 were tested for the presence of HIV-1 and human immunodeficiency virus type 2 (HIV-2) antibody and those that were seronegative were tested for p24 antigen and invited to return for follow-up every 3 months. The first method, called the "cohort" estimate, was based on the cohort of individuals who returned for at least one follow-up visit and was calculated as the ratio of the numbers of seroconverters ($d$) divided by the person-time of follow-up ($T$). The cohort estimate is $I_0 = d/T$. Follow-up bias could cause this estimate to seriously misrepresent the incidence rate among all patients attending the sexually transmitted disease clinic, especially if the follow-up rate is low.

The second method, called the "cross-sectional" estimate, was based on the prevalence of p24 antigenemia among all HIV seronegative patients who were initially screened. HIV p24 antigen is normally detectable in the sera at some point prior to seroconversion (3–7). The relation between the HIV-1 incidence rate, $I$, the proportion of seronegative individuals with p24 antigenemia, $p$, and the mean duration of p24 antigenemia prior to seroconversion, $\mu$, can be derived from an epidemiologic model (8, 9). This gives an equation $p \approx I \times \mu$ that can be solved for $I$ if data on the mean duration of p24 antigenemia prior to seroconversion are available. This well-known epidemiologic relation is usually justified by the assumption of a steady state. However, as described by Brookmeyer and Quinn (8) in fact all that is required is that the number infected per unit time remains approximately constant for at least as long as the maximum plausible duration of p24 antigenemia. This is a reasonable assumption because $\mu$ is relatively short (8). The estimate of incidence derived from this equation is not affected by either low follow-up rates or follow-up bias because it is based only on the cross-sectional survey of the prevalence of p24 antigenemia among all seronegative patients who were initially screened, and does not require any longitudinal follow-up data. However, this "cross-sectional estimate" is sensitive to assumptions about the mean duration of p24 antigenemia prior to seroconversion. This duration has been reported to be on the order of several weeks or a month (4, 8, 10), and one analysis reported $\mu = 22.5$ days (8).

The third estimate of incidence was based on a new approach that combined the cohort and cross-sectional estimates. The advantages of this approach are that it is not affected by any follow-up bias nor does it make assumptions about the mean duration of p24 antigenemia prior to seroconversion, and thus this method is quite different from previous approaches. Our method is based on the following argument. Let the ratio $R$ be the prevalence (proportion) of p24 antigenemia among the seronegatives who do not return for follow-up ($p_1$) divided by the prevalence of p24 antigenemia among those who do return for follow-up ($p_0$). This ratio $R$ also estimates the ratio of incidence rates among those who do not return ($I_1$) and do return for follow-up ($I_0$). This follows from the equation $p = I \times \mu$ and the assumption that the mean durations of p24 antigenemia are the same among infected individuals who do and do not return for follow-up ($R = p_1/p_0 \approx I_1\mu/I_0\mu = I_1/I_0$). Thus, it is possible to estimate the incidence rate among those who did not return for follow-up by $R \times I_0$, where $I_0$ is the cohort estimate. Then, the estimate of the overall HIV-1 incidence rate among all those initially screened, $I$, is a weighted average of the incidence rates of those who did and those who did not return for follow-up:

$$I = w \times I_0 + (1 - w) \times R \times I_0, \qquad (1)$$

where $w$ is the proportion who returned for follow-up, and $I_0 = d/T$. Confidence interval procedures for the incidence rate based on this method as well as the cross-sectional and cohort methods are outlined in the Appendix.

## RESULTS

The overall HIV seroprevalence among 1,885 patients attending sexually transmitted disease clinics in Pune, India, at baseline was 25 percent. Of 1,408 individuals who were HIV-1 and HIV-2 seronegative at baseline, sufficient sera were available on 1,241 individuals who were tested for the presence of p24 antigen, and 15 of these were p24 antigen positive. Of the 1,408 seronegative individuals at baseline, 438 (31.1 percent) agreed to enroll in the cohort study and returned for their first follow-up visit.

The incidence estimates are given in table 1. The cohort estimate of incidence was 11.7 percent per year (95 percent confidence interval (CI) 8.1–16.5) based on 33 observed seroconversions during 282.6 person-years of follow-up. The cross-sectional estimate assuming the mean duration of p24 antigenemia prior to seroconversion was 22.5 days was

$$(15/1{,}241) \times (365/22.5) \times 100$$

$= 19.6$ percent per year (95 percent CI 11.0–32.3).

The estimate of HIV incidence based on the new method of combining cohort and cross-sectional data was 18.6 percent per year (95 percent CI 8.5–40.3). This calculation was based on the ratio of the prevalences of p24 antigenemia, which indicated that the incidence rate among those not returning for follow-up

**TABLE 1. Human immunodeficiency virus (HIV) incidence estimates among sexually transmitted disease clinic patients in Pune, India, May 1993–August 1994***

| Method | Rate (%/year) | | |
|---|---|---|---|
| | Overall | Males | Females |
| Cohort | 11.7 | 9.7 | 20.1 |
| Cross-sectional | 19.6 | 20.8 | 14.2 |
| Combined cross-sectional-cohort | 18.6 | 20.6 | 15.1 |

* Data based on patients first screened between May 1993 and January 1994 and follow-up data through August 1994. HIV-1 antibody testing was based on a second generation enzyme immunoassay, and repeatedly reactive specimens were confirmed by Western blot test (DuPont, Wilmington, Delaware). P24 antigen tests were based on a monoclonal antibody-based antigen capture enzyme-linked immunosorbent assay (ELISA) (Abbott Laboratories, Abbott Park, Illinois). Positive specimens (defined as >15 pg/ml of p24 antigen) were retested with an anti-p24 neutralization procedure to ensure high specificity. Cross-sectional estimates assume $n$ = 22.5 days. All available p24 antigen data from all follow-up visits were used to estimate $R$ in the combined cross-sectional-cohort method.

was $R$ = 1.85 times higher than the HIV incidence rate among those returning for follow-up. The percents that were p24 antigen positive among those returning and not returning for follow-up were 0.874 percent and 1.62 percent, respectively. Equation 1 gave

$$0.31 \times 11.7 + 0.69 \times 1.85 \times 11.7$$

$$= 18.6 \text{ percent per year.}$$

The cohort estimate of incidence, 11.7 percent/year, was lower than that obtained with the other methods. The ratio $R$ = 1.85 suggested the presence of some follow-up bias whereby individuals who did not return for follow-up had a higher HIV-1 incidence rate, although the confidence interval for $R$ was wide (95 percent CI 0.7–5.1). An additional analysis was undertaken to determine if this trend in follow-up bias was corroborated by demographic, behavioral, and other risk factors for HIV incidence collected at baseline. It was found that individuals who returned for follow-up generally reported behaviors in the previous 3 months consistent with lower risk of HIV incidence compared with those who did not return for follow-up (table 2): those who returned for follow-up were more likely to have reported no sexual partners in the previous three months, and to have used a condom and were less likely to have an active genital ulcer.

**TABLE 2. Percent of patients who return for at least one follow-up visit among patients screened in sexually transmitted disease clinics in India by baseline characteristics**

| Characteristic | Overall | | Males | | Females | |
|---|---|---|---|---|---|---|
| | No. | % who return | No. | % who return | No. | % who return |
| No. of sexual partners in previous 3 months | | | | | | |
| 0 | 139 | 37 | 122 | 40 | 17 | 12 |
| 1 | 674 | 27 | 563 | 29 | 111 | 17 |
| ≥2 | 587 | 34 | 446 | 32 | 141 | 40 |
| $p$ value* | | 0.01 | | 0.06 | | <0.001 |
| Condom use in previous 3 months | | | | | | |
| Never | 1,046 | 28 | 899 | 29 | 147 | 22 |
| Sometimes or always | 349 | 40 | 228 | 42 | 121 | 37 |
| $p$ value* | | <0.001 | | <0.001 | | 0.008 |
| Sex with commercial sex workers in previous 3 months | | | | | | |
| No | 597 | 32 | 328 | 35 | 269 | 29 |
| Yes | 806 | 30 | 806 | 30 | 0 | |
| $p$ value* | | >0.25 | | 0.12 | | |
| Currently employed as commercial sex worker | | | | | | |
| No | 1,259 | 30 | 1,132 | 32 | 127 | 16 |
| Yes | 146 | 40 | 4 | 25 | 142 | 40 |
| $p$ value* | | 0.02 | | >0.25 | | >0.001 |
| Active genital lesion | | | | | | |
| No | 755 | 34 | 545 | 36 | 210 | 28 |
| Yes | 653 | 28 | 593 | 28 | 60 | 33 |
| $p$ value* | | 0.03 | | 0.003 | | >0.25 |

* $p$ values are based on chi-squared tests for contingency tables.

However, separate analyses for males (81 percent of those screened) and females (19 percent of those screened) indicated quite different trends. The incidence rate among males who did not return for follow-up was estimated to be $R = 2.66$ times the rate among males who did return for follow-up. Generally, males who returned for follow-up presented with characteristics consistent with lower risk of incidence of HIV infection compared with males who did not return for follow-up (table 2). In contrast, the incidence rate among females who did not return for follow-up was estimated to be only $R = 0.65$ times the incidence rate among females who did return for follow-up. This was generally corroborated by the baseline data (table 2) that showed females who returned for follow-up presented with characteristics consistent with a higher risk of incidence of HIV infection compared with females who did not return for follow-up. Females who returned for follow-up were more likely to be commercial sex workers, and to report at least two sexual partners in the previous 3 months.

The cohort estimate of the female HIV incidence rate was more than two times the male incidence rate (table 1). However, after adjusting for follow-up bias by combining the cross-sectional p24 antigen data with the cohort data, the female incidence rate was lower than the male incidence rate (table 1). A main explanation for this is that female commercial sex workers who have relatively high HIV incidence rates were 2.5 times more likely to return for follow-up than other females.

## DISCUSSION

The traditional cohort method of estimating HIV incidence is problematic if the follow-up rate is low or if there is follow-up bias. Our data indicate that the HIV incidence rate in a screened population may be nearly 60 percent larger than that based on the prospective follow-up of a cohort because of follow-up bias. Advantages of the cross-sectional estimate are that follow-up data are not required, and the estimate is not affected by follow-up bias. However, it is sensitive to assumptions about the duration of p24 antigenemia. Advantages of the method that combines cross-sectional data on p24 antigenemia with cohort data are that it accounts for follow-up bias and makes no assumptions about the duration of p24 antigenemia. Both the cross-sectional and combined methods gave estimates broadly consistent with each other. Our data show the feasibility of gleaning estimates of HIV incidence from a large cross-sectional survey of p24 antigenemia in an area with a growing epidemic. However, large sample sizes will be required because the duration of p24 antigenemia is relatively short (8), and thus the approach may be most useful in some developing countries where the incidence rate is high and where it is possible to screen large numbers of individuals.

Our HIV incidence estimate of 18.6 percent per year may not be representative of all sexually transmitted disease patients in India. Approximately 53 percent of our female patients were commercial sex workers, and 52 percent of our male patients presented with active genital ulcer disease. Nevertheless, the high incidence rates are broadly consistent with a number of surveys among sexually transmitted disease clinics in India that show a rapid rise in seroprevalence rates in the 1990s (11–14). The confidence interval based on the incidence estimate of 18.6 percent per year was wide. Nevertheless, even the lower confidence limit of over 8 percent per year was alarmingly high. Larger sample sizes are required to produce narrower confidence intervals. The development and application of other diagnostic tests for other markers of recent infection with longer durations, such as polymerase chain reaction, may be useful for improving the precision of incidence with these methods (8).

The HIV incidence estimate of 18.6 percent per year together with a seroprevalence rate of 25 percent indicates an epidemic doubling time of less than 2 years. This doubling time is similar to doubling times reported in the 1980s (15) in selected populations in a number of countries in Africa, such as Malawi, Kenya, and Uganda. The high seroprevalences of HIV infection that have been reported among commercial sex workers as well as other sexually transmitted disease patients are evidence of the extent of the past growth of the epidemic in India. The current high incidence rates indicate that the HIV epidemic in India continues to grow rapidly.

## ACKNOWLEDGMENTS

Supported by the National Institute of Allergy and Infectious Diseases (AI 33879-02), the National Cancer Institute (CA 48723), the NIH-NCPR OPD-GCRC (5M01RR00722), the NIH-Fogarty International Center Program of International Training Grants in Epidemiology Related to AIDS (D43TW0000), and Family Health International (FHI).

## REFERENCES

1. Brookmeyer R. Reconstruction and future trends of the AIDS epidemic in the United States. Science 1991;253:37–42.
2. Brookmeyer R, Gail MH. AIDS epidemiology: a quantitative approach. New York: Oxford University Press, 1994:

80–1.
3. Allain JP, Laurian Y, Paul DA, et al. Serological markers in early stages of human immunodeficiency virus infection in hemophiliacs. Lancet 1986;2:1233–6.
4. Haseltine WA. Silent HIV infections. N Engl J Med 1990; 323:1308–12.
5. Clark SJ, Saag MS, Decker WD, et al. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. N Engl J Med 1991;324:954–60.
6. Daar ES, Moudgil T, Meyer R, et al. Transient high levels of uremia in patients with primary human immunodeficiency virus type 1 infection. N Engl J Med 1991;324:961–4.
7. Clark SJ, Kelen GD, Henrard DR, et al. Unsuspected primary human immunodeficiency virus type 1 infection in seronegative emergency department patients. J Infect Dis 1994;170: 194–7.
8. Brookmeyer R, Quinn TC. Estimation of current human immunodeficiency virus incidence rates from a cross-sectional survey using early diagnostic tests. Am J Epidemiol 1995;141: 166–72.
9. Freeman J, Hutchison G. Prevalence, incidence, and duration. Am J Epidemiol 1980;112:707–23.
10. Bowen PA, Lobel SA, Caruana RJ. Transmission of human immunodeficiency virus by transplantation. Clinical aspects of time course of viral antigenemia and antibody production. Ann Intern Med 1988;108:46–8.
11. Tripathy S, Bannerjee K, Rodrigues J, et al. Increasing HIV infection in Western India. IX Int Conf AIDS, Berlin 1993; abstract PO-C08-2764.
12. Hasmukh S, Gilada IS, Sharma R, et al. Progression of seroprevalence among attendees of STD clinics in Central Bombay—a study over 6 years. Int VIII Conf AIDS, Amsterdam 1992;8(2):C296 abstract PO-C4306.
13. US Bureau of Census, Center for International Research. Overview of HIV trends and patterns in India. HIV/AIDS Surveillance Data Base, June 1993.
14. Rodrigues JJ, Mehendale SM, Shepherd ME, et al. The biological and behavioral risk factors for prevalent HIV infection in sexually transmitted disease clinics in India. Br Med J 1995;311:283–6.
15. Anderson RM, May RM, Boily MC, et al. The spread of HIV-1 in Africa: sexual contacts, patterns and the predicted demographic impact of AIDS. Nature 1991;352:581–9.
16. Breslow NE, Day NE. Statistical methods in cancer research. Vol II. Design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, 1987.

## APPENDIX

In this Appendix, the confidence interval procedures for the incidence rate are outlined. Confidence intervals for incidence based on the cohort method used the Poisson distribution (see reference 16 for additional detail) by relating the number of seroconverters, $d$, to a Poisson distribution with mean given by the product of the true incidence among returners and $T$, where $T$ is the total person-time of follow-up. Similarly, confidence intervals for incidence based on the cross-sectional method were based on referring the number who were p24 antigen positive at baseline to a Poisson distribution with mean given by the product of the number who were HIV seronegative and tested for p24 antigen, the true incidence, and $\mu$. Both of these confidence intervals are exact and do not rely on asymptotic approximations.

An approximate asymptotic confidence interval for incidence based on the combined cross-sectional-cohort method was derived. Let $X_1$ and $X_2$ be the number who were p24 antigen positive among the returners and nonreturners, respectively. Let $X_3$ be the number of individuals who seroconverted on follow-up but did not previously test antigen positive. Then, it can be shown by application of the delta method (with $d = X_1 + X_3$) and assuming $X_1$, $X_2$, and $X_3$ have Poisson distributions, that the variance of $\ln(I)$ is approximately

$$\frac{\left[w - (1-w)R\dfrac{X_3}{X_1}\right]^2 X_1}{c^2(X_1+X_3)^2} + \frac{(1-w)^2R^2}{c^2X_2} + \frac{X_3}{(X_1+X_3)^2},$$

where $c = w + (1 - w)R$. An approximate 95 percent confidence interval for $I$ is $\exp(\ln I - 1.96V)$, $\exp(\ln I + 1.96V)$, where $V$ is the square root of the variance of $\ln(I)$. This variance formula for $\ln I$ accounts for uncertainty both in $R$ and $I_0$ as well as the correlation between them. However, it is only an approximate large sample procedure.