Invited Commentary: Should We Estimate Incidence for Undefined Populations?

Victor J. Schoenbach,1 Charles Poole,1 and William C. Miller1,2

When acquired immunodeficiency syndrome (AIDS) and its potential threat to public health were recognized 20 years ago, disease control officials confronted major challenges in tracking the spread of human immunodeficiency virus (HIV) infection (1). Since most persons do not experience characteristic symptoms for many years after infection with HIV, AIDS case reports could not provide timely information about the progress of the epidemic. At the same time, the combination of very low overall prevalence; wide variability in HIV prevalence by geographic, demographic and behavioral subgroups; and the powerful social stigma associated with the behaviors linked to the disease posed major barriers to serosurveys in the general population. Meanwhile, the seriousness of the disease, its potential for rapid dissemination, and the intensity of public concern made effective surveillance essential. Serosurveillance of sentinel populations, notably the Centers for Disease Control and Prevention’s family of serosurveys (1, 2), emerged as a key strategy for obtaining vital surveillance data within the constraints imposed by the nature of HIV and the social and political context.

Although data from sentinel seroprevalence surveys have provided the basis for policy, control measures, and research, these surveys have intrinsic limitations that complicate interpretation of the data they provide (2, 3). One key limitation is that, due to the extended natural history of HIV infection, seropositive persons without AIDS may have a wide distribution of times since infection. The median interval between HIV infection and AIDS diagnosis, previously considered to be 10 years (2), has become even longer with the availability of effective antiretroviral therapy.

The second key limitation of sentinel serosurveys is the difficulty of characterizing the population base that is being sampled. By their very nature, sentinel populations are not sampled from the general population, but are “self-selected” in ways that can make them very unrepresentative of the general population. Serosurveillance has been carried out with hospital inpatients, patients in ambulatory settings, Job Corps participants, military recruits, childbearing women, and sexually transmitted disease (STD) clinic patients. The relation of each of these populations to the general population or to some other clearly defined base population is unknown and can change in unknown ways over time.

The Serologic Testing Algorithm for Recent HIV Seroconversion (STARHS) (4) represents a major advance in distinguishing “new” infections from long-standing ones. The technique is not perfect. Infections can be missed or misclassified if the antibody response has not yet developed or has waned due to weakening of the immune system or highly effective antiretroviral therapy (4). Nevertheless, by avoiding a major problem in interpreting trends in seroprevalence data, STARHS greatly increases the information available from serology. Schwarze et al. (5) have made use of this advance to analyze 10 years of stored serologic specimens from annual serosurveys of patients at the San Francisco sexually transmitted disease clinic. Their analyses disclose a divergence between total seroprevalence (declining) and prevalence of recent infection (stable or rising) among persons attending the clinic.

For most conditions, incidence is superior to prevalence for investigating etiology and evaluating prevention activities. Given the relation prevalence odds = incidence × mean duration (which, for a rare condition such as recent HIV infection, can be approximated by prevalence = incidence × mean duration), the ability to identify recent infections creates a theoretical possibility of estimating incidence from prevalence. This relation is valid only for a stationary population with no net migration (6), however, and does not apply to age-specific prevalence (7).

The “stationarity” assumption requires that “the number of people entering a population is balanced by the number exiting the population in any period of time within levels of age, sex, and other determinants of risk” (6, p. 34). Furthermore, the number of new cases (people entering the prevalence pool) must equal the number exiting the prevalence pool, which implies that the rates of entering and leaving the prevalence pool have remained constant long enough to reach a steady-state equilibrium. For conditions with extended durations (i.e., years), these conditions rarely hold. However, the duration of the state detected by the STARHS technique is only about 4 months (129 days), so that small, gradual changes in incidence will disturb stationarity only slightly. Thus, the stationarity assumption...
may not present a major problem for annual incidence estimates in a mature epidemic.

However, incidence also requires a “population at risk” (base population) in which susceptible persons can, at least in principle, be followed over a period of time. That base population must be validly sampled to estimate incidence by using prevalence and mean duration. Although the STARHS technique permits identification of newly infected persons, it has no effect on the problem of identifying or validly sampling a population base. It is not obvious to what population, if any, the incidence estimates in the paper by Schwarcz et al. apply. Direct interpretation of the seroprevalences in the samples of the clinic population is straightforward. These seroprevalences estimate the proportions of infected persons among those who attended the clinic in each year. Application of the STARHS assay refines these seroprevalence estimates by focusing them on recent infections among the clinic population, but the technique does not identify a population at risk.

Since STD clinic patients are not representative of all clinic patients, all persons with STD, all persons practicing risk behaviors (8), or all persons in the community, generalizing STD clinic seroprevalence estimates to a base population entails identifying the “individuals in the catchment area who would choose to attend the STD clinic if they needed treatment” (3, p. 448). This is a hypothetical population that cannot be exactly enumerated. Indeed, even this characterization ignores the fact that within the population of people who would attend the STD clinic, the people who do attend are those who are more likely to have an STD at a given time and therefore are more likely to have recently acquired HIV. Since HIV infection is an irreversible state, the population at risk, or study base, consists of all uninfected people who will be counted as cases if they acquire HIV. For the study by Schwarcz et al., therefore, the study base can be defined only as “people who will come to the San Francisco STD clinic.” However, that population is not meaningful as a referent for an incidence estimate.

Given the study’s focus on incidence and prevalence trends, a key question is whether the relation between clinic and population has changed during the period, a possibility that is suggested by the substantial decline (13,829 to 9,429) in patients seen per year during the 10-year period, despite relative stability in the number of those aged 15–49 years in San Francisco. The multiple logistic regression analyses controlling for patient characteristics reduce the influence of changes in the makeup of the patient population. They do not, however, address the question of the relation of the clinic population to the larger community. This relation is a labyrinth of transition probabilities back and forth between the San Francisco population and the subpopulation attending the clinic.

Figure 1 presents a schematic of the linkage between the San Francisco STD clinic and the San Francisco population. The probability of acquiring an STD is affected by the person’s own behavior (e.g., number of partners, partner selection, and condom use), the behavior of his or her partners, and the prevalence of STD in the partner pool. The act of seeking evaluation for an STD is a function of an person’s level of concern, perception of risk, inclination to use health services, and symptom development and recognition or, alternatively, of the probability that a sexual partner becomes infected, is diagnosed, and notifies the individual.

Whether a person visits a particular clinic is a function of such factors as economic resources, clinic referral activities,
presence of symptoms from non-STD causes, community attitudes toward the clinic, clinic hours, availability of transportation, and access to other providers (9, 10). Even if the San Francisco STD clinic evaluates all STDs in the city, many of the STD patients who make up the seroprevalence denominator do not have an STD at the time of the visit. The factors that lead uninfected patients to attend the clinic are therefore important determinants of seroprevalence estimates. For reasons such as these, generalization of results from convenience samples, such as clinic surveys of any kind, is not at all straightforward and requires an in-depth understanding of the local setting (3).

Schwartz et al. recognize the problem in their incidence estimation procedure (e.g., “…the base population is not known and its size and composition may have varied during the study period…” (5, p. 933)). However, that recognition does not deter them from using the “seroincidence” estimates as if they applied to an identifiable population. Their abstract reports that the “pooled seroincidence was 1.6%” (5, p. 925). An unsophisticated reader applying this “incidence” to the San Francisco population aged 15–49 years would estimate approximately 7,800 new HIV infections in 1998, more than one third of the total number reported for that year from the 33 areas of the United States with confidential HIV infection reporting (11).

However, if not to the San Francisco population, to what population does the 1.6 percent apply? Although the authors do not assert that their incidence estimates apply to the population of San Francisco, they nevertheless use these estimates to draw conclusions about the effectiveness of prevention programs in San Francisco. They also suggest that “STARHS testing in sentinel sites such as STD clinics may prove superior to other methods of estimating HIV incidence” (5, p. 932).

The implication of these claims is that any clinical setting can serve to estimate incidence of a condition whose duration is short and is approximately known. Indeed, a recent article estimated “HIV seroincidence” among persons coming to anonymous counseling and testing sites (12). These incidence estimates represent the net effect of the number of new cases that occur during the time period (offset by about 6 months), factors that 1) attract or obstruct attendance by persons engaged in high-risk behavior, 2) lead other people to have themselves tested, and 3) affect the rapidity with which people seek testing. Were this type of outcome-dependent sampling valid, the incidence estimates could be estimated by counting the number of new myocardial infarctions diagnosed in the emergency rooms serving a community and dividing by the product of the number of persons presenting with chest pain or other symptoms of infarction and the mean duration of a new infarction. Whose “incidence” is this?

The objection to estimating incidence without an identified base population is not a criticism of the STARHS technique, which represents a very important advance in HIV serology and can indeed serve to estimate incidence of HIV infection in an identified cohort or other defined population or in a representative sample thereof (4), to the extent that assumptions about stationarity (6) are met. Nor is this objection a criticism of the authors’ data, which are of considerable interest. However, an incidence rate without an identifiable base population is a meaningless number. Virtually any incidence rate applies to some population. Converting the question from “What is the incidence for this population?” to “What is the population for this incidence estimate?” conveys no benefit.

The examination of trends in incidence with the aid of the STARHS method does not, in fact, require the estimation of incidence. To the extent that the clinic population accurately reflects the dynamics in the larger population, trends can be analyzed through examination of the seroprevalence of recent infections (from which the “seroincidence” estimates were derived by multiplying by a constant). Prevalence estimates from sentinel populations also present problems of interpretation and representativeness. However, such estimates do not require imputation of a denominator, and their limitations are well recognized (3, 8). Manufactured incidence estimates add nothing of value but are much more susceptible to being misinterpreted.

REFERENCES