Monitoring the Incidence of HIV Infection in the United States

LISA M. LEE, PhD^a
MATTHEW T. MCKENNA, MD,
MPH^b

SYNOPSIS

The Centers for Disease Control and Prevention maintains a national surveillance system that provides data about the HIV/AIDS epidemic for program planning and resource allocation. Until recently, incidence of HIV infection (i.e., the number of individuals recently infected with HIV) has not been directly measured. New serologic testing methods make it possible to distinguish between recent and long-standing HIV-1 infection on a population level. This article describes the new National HIV Incidence Surveillance System.

^aOffice of the Chief Science Officer, Centers for Disease Control and Prevention, Atlanta, GA

^bHIV Incidence and Case Surveillance Branch, Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, GA
Address correspondence to: Lisa M. Lee, PhD, Office of the Chief Science Officer, Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, MS D-50, Atlanta, GA 30333; tel. 404-639-7260; fax 404-639-7341; e-mail: <LMLee@cdc.gov>.

The Centers for Disease Control and Prevention (CDC) is responsible for maintaining a national surveillance system that provides data about the HIV/AIDS epidemic. This information is used for national, state, and local public health HIV/AIDS prevention program planning and resource allocation. Historically, AIDS diagnosis data have been of great value; however, current AIDS data do not represent the entire population affected by the HIV epidemic. Unlike AIDS data, HIV data provide a window into the epidemic at an earlier stage of disease. Until recently, biomedical technology did not discriminate between recent and chronic HIV infection. As a result, HIV surveillance has been limited to monitoring prevalence—the proportion of individuals diagnosed with HIV antibodies regardless of the duration of HIV infection. The incidence of HIV infection in the United States (i.e., the number of individuals recently infected with HIV) has not been directly measured.

The Institute of Medicine, in evaluating the use of HIV data for national resource allocation, recommended in 2001 that CDC “develop an accurate surveillance system focused on new HIV infections that can better predict where the epidemic is headed.”¹ New serologic testing methods make it possible to implement a system that distinguishes between recent and long-standing HIV-1 infection on a population level. The most studied of these methods is the Serologic Testing Algorithm for Recent HIV Seroconversion (STARHS).² These laboratory methods, in conjunction with standard case surveillance procedures and statistical estimation, provide the means to estimate national population-based HIV incidence from the number of recent infections among people who are newly diagnosed with HIV.

ESTIMATING HIV INCIDENCE IN THE UNITED STATES

CDC has been estimating the number of new HIV infections since the late 1980s, initially using a back-calculation method. This method used the current AIDS numbers to estimate the number and timing of the HIV infections that would have had to occur in the past to produce the number of AIDS diagnoses observed in the present.^{3,4} This back-calculation method was appropriate as long as the average time between HIV infection and AIDS diagnosis (the incubation period) was constant and consistent across and within groups. This method became untenable, however, with the introduction of effective therapies that altered the incubation period.

Another method of incidence estimation used data

synthesis—combining information from published cohort studies with estimates of the relationship between prevalence and incidence.⁵ CDC estimates approximately 40,000 new HIV infections in the U.S. per year since 1994. All methods used estimation in the absence of the ability to directly observe and measure new infections or measure new infections only in a small, select population.

A different approach to measuring incidence that has been investigated since the early 1990s has been the “snapshot estimator” approach.^{6,7} Snapshot estimators use information about markers of HIV progression to stratify diagnosed infections into recent vs. long-standing infections. By defining the period of time it takes for HIV infections to transition from one point in the disease to another (i.e., from infection with HIV to antibody positive status), it is possible to calculate the number of new infections that must occur to result in the observed number of people within the recent infection “state.” One of the first markers of progression used for these purposes was the p24 antigen, which appears before HIV antibodies are detectable. However, the “window” of time associated with the p24-positive, antibody-negative state is only a few weeks, and it requires large numbers of observations to derive statistically precise estimates of incidence. In 1998, Janssen, et al.² described a new HIV testing strategy based on the quantitative increase in HIV-specific antibodies that occurs in the first months of infection. The window period associated with this algorithm was approximately six months and is the optimal time period for calculating incidence using the snapshot estimation approach (Figure 1).

THE NATIONAL HIV INCIDENCE SURVEILLANCE SYSTEM

With this new technology, CDC began consulting with internal and external experts about the best approaches for implementing the STARHS technology to estimate the number of new HIV infections in the United States each year. Between 2001 and 2005, CDC held five consultations with experts to solicit guidance in epidemiology, surveillance, biostatistics, laboratory science, ethics, policy, and state/local public health practice on a variety of topics associated with developing and implementing HIV incidence surveillance. These topics included surveillance methods, statistical issues, policy and ethical considerations, implementation procedures, and laboratory and specimen transport issues. The recommendations from these consultations helped shape the development and implementation of the National HIV Incidence

Figure 1. Serologic Testing Algorithm for Recent HIV Seroconversion (STARHS)

NOTES: STARHS using the BED HIV-1 Capture Enzyme Immunoassay (EIA), named for the branched peptide sequences for HIV-1 subtypes B, E, and D used in the assay. Mean window for determination of recent infection is 153 days. If a confirmed HIV-1 positive specimen is reactive on the standard sensitive EIA and has a standardized optical density of <0.8 on the BED assay, it is considered recently infected. While misclassification of recent and long-standing infections for individuals is substantial, STARHS can accurately estimate the number of recent infections in groups of people because the number of incorrect "recent infections" approximately equals the number of incorrect "long-standing infections." For example, while STARHS can identify that 10 out of 100 people are recently infected, it cannot precisely identify which 10 are recent.

Surveillance System described here. Currently, 34 state and local moderate-to-high morbidity jurisdictions are funded to conduct HIV incidence surveillance (Figure 2), covering approximately 85% of the epidemic. All funded systems began collecting data in 2005 and are expected to achieve full implementation in both public and private sectors in 2006.

Recommendations from the first expert consultation on surveillance methods were fundamental in creating the incidence surveillance system. The panel considered several methods for estimating HIV incidence, including national household surveys, unlinked serosurveys, cohort record review studies, and expanding the current case surveillance system by applying STARHS to all new HIV diagnoses in the United States. After reviewing each model and considering their individual strengths and weaknesses, the panel

recommended that HIV incidence be estimated by building on the existing national HIV case surveillance system. This system had collected complete, accurate, and timely data on AIDS cases for more than 15 years. Many states had integrated confidential, name-based HIV diagnosis reporting into their AIDS systems using the same validated methods as used for AIDS case ascertainment. Given the strength and reliability of this infrastructure, the panel recommended expanding it to include STARHS testing of all new HIV diagnoses reported to the system.

Expanding existing case surveillance infrastructure

Expanding the existing HIV case surveillance system to incorporate HIV incidence as a sentinel event also enables data use at the state level. CDC provides epidemiologic technical assistance to surveillance areas

Figure 2. 34 U.S. jurisdictions implementing HIV incidence surveillance, August 2005

to meet program obligations for data collection and dissemination, and calculation of incidence will be included in this assistance.

The existing HIV case surveillance infrastructure has several limitations that affect the incorporation of incidence surveillance activities. First, implementation of HIV case surveillance has been staggered over the past two decades, with the last states adopting HIV reporting in 2004. Incidence estimates will not be stable until all states establish HIV case reporting systems that enumerate all HIV cases that were diagnosed prior to the year reporting was implemented. Identifying all diagnosed patients, and differentiating newly diagnosed individuals from prevalent ones, often requires that a reporting system operate for several years.⁸ Second, only a portion of the actual new infections can be observed using this system (i.e., infected individuals who chose to be tested for HIV). Anonymous testers—those not reportable to the confidential system until they seek care and receive their first confidential HIV test—or individuals who are not tested at all cannot be directly ascertained. Estimating the total number of new infections requires the use of statistical methods to account for the bias associated with using a system that directly measures only individuals who test voluntarily.

The HIV incidence surveillance system, developed as an extension of national HIV case reporting, has received a determination of non-research by the human subjects protection process at CDC and is exempt from review by the Institutional Review Board (IRB). Informed consent is not required to collect and analyze cases of notifiable diseases, so consent for STARHS is not required and most case surveillance data are derived from clinical records. Because HIV testing is a medical procedure, the privacy of HIV information that is maintained in clinical records is protected by federal and state laws,⁹ which specifically allow public health data to be collected from these clinical records. Therefore, data contained in the HIV incidence and case surveillance system are maintained according to the U.S. Department of Health and Human Services manual on security and confidentiality.¹⁰ Policies and procedures based on these guidelines and local laws are in place at state and local health departments. Information maintained in hard copy and electronic formats are secured and protected using these rules to assure the privacy and confidentiality of individuals reported as having HIV infection. These measures extend to protect the HIV incidence surveillance information held locally. Access by HIV incidence

surveillance staff to information in the case record, HIV testing history, and STARHS data is governed by the same security and confidentiality requirements. Under these guidelines, information that could identify an individual (i.e., name, address, zip code) is not included in the dataset that is transmitted from local surveillance areas to CDC.

Laboratory technology for HIV incidence surveillance

The assay currently used in STARHS is the BED HIV-1 Capture Enzyme Immunoassay (EIA), named for the branched peptide sequences for HIV-1 subtypes B, E, and D used in the assay.¹¹ CDC developed the assay for the express purpose of estimating population-based HIV incidence, not for clinical or diagnostic purposes. Because its individual-level clinical and diagnostic usefulness has not been determined or approved by the Food and Drug Administration, it is at this time considered a tool for public health surveillance only. Therefore, CDC and its surveillance partners are prohibited from providing individual test results to patients or their health-care providers. CDC has recommended to state and local surveillance programs that health-care and partner counseling and referral providers use other readily available epidemiologic, laboratory, and clinical information (such as history of a negative HIV test, history of primary HIV infection illness, CD4 count, and HIV viral load results) to choose the best course of action for treatment and partner notification.

The principle of the BED HIV-1 Capture EIA is based on the observation that the ratio of HIV-specific immunoglobulin G (IgG) to total IgG increases with time after HIV infection. The time lapse between an individual's HIV-positive test on the standard EIA to the time when the individual's serum—if tested with the BED HIV-1 Capture EIA—reaches an optical density (OD) level predetermined to distinguish recent from long-standing infections, is defined as the mean STARHS window period (Figure 1). Although the mean STARHS window period may vary slightly by HIV subtype, the mean window period used to calculate population-based incidence estimates is 153 days.

In 2006, results from an incidence estimation using the BED HIV-1 Capture EIA in South Africa indicated concern that the assay, along with the statistical estimator, had overestimated the incidence of HIV infection compared with incidence estimated using other methods, including modeling and prospective studies.¹² The inability to exclude individuals with AIDS in serosurveys, the use in the developing world of diagnostic algorithms with lower specificity, and the high levels of chronic co-infections (and thus IgG) among people

living in the developing world may contribute to overestimation. In the U.S., however, linked case information is available to exclude individuals with AIDS who may falsely test “recent” on STARHS. Additionally, STARHS is done in the U.S. only in cases in which the specimen is confirmed HIV-1 antibody positive using a Western blot or immunofluorescence assay (IFA) confirmatory test. Finally, levels of chronic co-infection are low in the U.S. As such, false-recent results due to high IgG do not impact the assay's performance. A key factor for successful use of the BED assay in the U.S. is incorporating its use into the existing case-based surveillance system so that STARHS results can be interpreted in the context of all case information, as opposed to a single, cross-sectional assay result.

HIV case surveillance

National HIV case surveillance on which the incidence system is based is described in detail elsewhere.^{8,13} Briefly, all states and territories have laws or regulations requiring that HIV cases be reported to state or local public health agencies. As of July 2006, 50 states and territories use confidential, name-based reporting for HIV cases that they use for AIDS case reporting, resulting in an integrated HIV and AIDS disease reporting system. Information is reported on a standard case report form and includes data on patient demographics, HIV risk behaviors, laboratory and clinical events, and virologic and immunologic status. These data are submitted to the state or local public health authority and entered into the HIV/AIDS Reporting System, a standard software data management system. The reports are then forwarded without personal identifiers to CDC, where duplicate cases reported from more than one state are taken out through a process that identifies cases as potential duplicates (i.e., have the same alphanumeric soundex code,¹⁴ date of birth, and sex) of cases reported by other states. Through interstate communication procedures endorsed by the Council of State and Territorial Epidemiologists (CSTE),¹⁵ case identification numbers are sent back to state surveillance coordinators, who then use the identifying and epidemiologic information they retain to resolve whether the potential duplicates are the same or different people. This in turn is reported back to the CDC via the case identification numbers, and potential duplicates are resolved in the national dataset. This de-duplicated national dataset is used to develop the national epidemiologic profile of HIV/AIDS in the United States.

Data for HIV incidence

Two additional components are necessary for HIV incidence estimation: testing history information, including the number and timing of previous tests and a remnant blood sample from all newly diagnosed cases for STARHS testing. The acquisition of testing history information occurs as part of routine case surveillance through an expanded or supplemental case report form. The acquisition of the remnant blood sample occurs after an HIV test is confirmed positive. Remnant blood is aliquoted by the diagnostic laboratory and sent to the public health laboratory at the same time that the confirmation test is reported to the surveillance unit in the health department. The HIV incidence surveillance coordinator in each state then determines whether the report represents a new HIV diagnosis and, if so, notifies the public health lab to send the specimen for STARHS testing. STARHS results are returned to the reporting state's HIV incidence surveillance coordinator, where they are combined with the other case data.

The source of cases categorized as recent by STARHS consists of the population of people who volunteer for HIV testing and are diagnosed with HIV infection. However, the population of interest is all newly infected people, including those who were not tested for HIV soon after their infection. To extrapolate to the population as a whole from the observed sample, each observed recent infection must be statistically weighted based on the probability of testing shortly after infection. The details of the calculations used to estimate population-based incidence are presented in greater detail elsewhere.¹⁶

Statistical estimation

To calculate population-based incidence, all newly diagnosed cases classified as recently infected by STARHS or other information (i.e., a recent negative HIV test) during a calendar year are regarded as a sample of all newly infected individuals. The weights assigned to each identified case are based on an estimate of the probability of detecting a recent HIV infection in such a cross-sectional sample. This probability of detection is the probability that the diagnostic HIV test is undertaken during the window period of the BED assay. Thus, this probability is dependent on an individual's frequency of testing. The weight for each case is the inverse of the probability of detection. To estimate the probability of selection and assign a corresponding weight, members of the sample are assigned to mutually exclusive groups based on information about previous testing behaviors. These groups are used to assign weights based on whether the person was tested

before the first positive result, type of HIV diagnosis (HIV diagnosed with or without AIDS), result of the BED test, and availability of necessary information.

Individuals in the observed sample are assigned one of three types of weight: zero weight, calculated weight based on detection probability, or derived weight using a proportional method for cases with incomplete information (Figure 3).

Zero weight. A zero weight is assigned to all cases diagnosed as AIDS at the time of the first HIV-positive test. The incubation period for untreated HIV infection (i.e., time from HIV infection to AIDS symptoms) is approximately 10 years.¹⁷ By definition, an individual with AIDS at the time of HIV diagnosis has been infected for greater than the six-month STARHS window period. Cases diagnosed with AIDS when first diagnosed with HIV can be incorrectly categorized as a recent infection by STARHS, however, because of a fall in HIV-antibody levels. Cases with concomitant HIV and AIDS diagnoses are assigned a zero weight regardless of the STARHS result. A zero weight is also assigned to individuals who did not have AIDS at HIV diagnosis, did not take antiretroviral medications within the six months before testing positive, and who were deemed "long-standing" by STARHS.

Calculated weights. Based on detection probabilities, calculated weights are assigned to cases for which it is known whether or not there was an HIV test before the first diagnostic test; that did not have concomitant HIV and AIDS diagnoses; that did not take antiretroviral medications within the six months before testing positive; and that were deemed "recent" by STARHS. Cases assigned a calculated weight include individuals whose previous testing history is known, including both those who report having had no HIV test before their first positive test and those who report having had a negative test prior to their first positive test.

Derived weights. Derived weights are assigned using proportional allocation to people for whom the existence of a test prior to the first positive HIV test is unknown.

Calculated weights

There are two types of calculated weights: one for people who report having no previous test before the first positive test and the other for those who report having tested negative before the first positive test. The probability of selection and the corresponding weight for cases with no previous test can be obtained using survival functions that account for the distribution of the AIDS incubation period and the distribution of the

Figure 3. Clinical and testing history information for people who are newly diagnosed with HIV used to assign weights to calculate the number of incident HIV infections in the entire population

Information on testing behaviors or clinical stage at diagnosis	Weight
Cases of AIDS diagnosed at the time of HIV diagnosis and deemed "long-standing" by STARHS	0
A negative HIV test is reported prior to the diagnostic test ^a	The interval in time between the last negative test and the diagnostic test divided by the average length of time spent in the window period = calculated weight^b
No HIV test prior to the diagnostic test	Inverse of a probability derived from a gamma function that represents the distribution of window periods. The function is parameterized using the proportion of HIV cases diagnosed with AIDS at the time of their HIV diagnosis = calculated weight^{b,c}
Incomplete information on testing history or STARHS results	Proportional allocation based on data from people with STARHS results and testing history information who share similar demographic and/or clinical (i.e., CD4 results) characteristics to people with incomplete information = derived weight

^aThese weights are only calculated for infected individuals who are classified as recently infected by STARHS. People classified as "long-standing" by STARHS are given a weight of 0.

^bGlynn MK, Lee LM, McKenna MT. The status of national HIV case surveillance, United States 2006. Public Health Rep 2007;122(Suppl 1):63-71.

^cMortimer JY, Salathiel JA. "Soundex" codes of surnames provide confidentiality and accuracy in a national HIV database. Commun Dis Rep CDR Rev 1995;5:R183-6.

STARHS = Serologic Testing Algorithm for Recent HIV Seroconversion

BED window period. This probability is parameterized based on the proportion of individuals diagnosed with AIDS at the time of their HIV diagnosis.¹⁸

The calculated weight for people who report having had a negative test before the first positive test is given by the time interval (T) from the last negative HIV test to the first positive HIV test divided by the average time spent in the window period during the interval T among those who seroconverted at time zero. The average time in the window period is proportional to the size of T , because people who test very frequently will have less opportunity to spend long periods of time in the window period. However, for values of T greater than 24 months, this period essentially equals the overall mean window period, $\mu = 153$ days, for the BED assay.

For all other cases for which the existence of an HIV test prior to the first diagnostic test is unknown, weights are derived from the calculated weights using proportional allocation—where the proportion of such cases among all cases with complete information is extrapolated to the population without STARHS data based on the entire number of newly diagnosed cases in the population.

Finally, in the incidence estimation step, cases are assigned an incidence weight, and the incidence count for a specific population group is estimated by summing the weights of respondents in the population considered:

$$I = \sum_i W_i$$

where I = incidence, W = statistical weight, and i = case.

Computation of the associated variance is complex and has been outlined elsewhere.^{16,17}

CONCLUSION

New laboratory technology and a strong national HIV case surveillance infrastructure provide the opportunity to directly measure for the first time the number of new HIV infections occurring in the U.S. population for a given period. These data will provide public health officials with a more accurate picture of the front end of the HIV epidemic and help guide prevention science. Each part of the spectrum of the epidemiology of HIV disease, from risk behaviors through death from AIDS, sheds more light onto the complex task of developing and evaluating HIV prevention and treatment programs at the local and national levels. The ability to monitor HIV incidence as part of that spectrum enhances our ability to prevent the spread of HIV in our communities.

The authors would like to acknowledge the Incidence and Viral Resistance Team of the HIV Incidence and Case Surveillance Branch at CDC and the 34 state and local health department

partners for contributing to the development and implementation of the national HIV incidence surveillance system, as well as Drs. Maria Rangel and Ruiguang Song for their astute review of the statistical estimation section.

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the CDC.

REFERENCES

1. Institute of Medicine. No time to lose: getting more from HIV prevention. Washington: National Academy of Sciences; 2000.
2. Janssen RS, Satten GA, Stramer SL, Rawal BD, O'Brien TR, Weiblen BJ, et al. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *JAMA* 1998;280:42-8.
3. Rosenberg PS. Scope of the AIDS epidemic in the United States. *Science* 1995;270:1372-5.
4. Karon JM, Khare M, Rosenberg PS. The current status of methods for estimating the prevalence of human immunodeficiency virus in the United States of America. *Stat Med* 1998;17:127-42.
5. Vu MQ, Steketee RW, Valleroy L, Weinstock H, Karon J, Janssen R. HIV incidence in the United States, 1978-1999. *J Acquir Immune Defic Syndr* 2002;31:188-201.
6. Kaplan EH, Brookmeyer R. Snapshot estimators of recent HIV incidence rates. *Operations Research* 1999;47:29-37.
7. Brookmeyer R, Quinn TC. Estimation of current human immunodeficiency virus incidence rates for a cross-sectional survey using early diagnostic tests. *Am J Epidemiol* 1995;141:166-72.
8. Nakashima AK, Fleming PL. HIV/AIDS surveillance in the United States, 1981-2001. *J Acquir Immune Defic Syndr* 2003;32 (Suppl 1):S68-85.
9. Centers for Disease Control and Prevention (US). HIPAA privacy rule and public health: guidance from CDC and the U.S. Department of Health and Human Services. *MMWR Morb Mortal Wkly Rep* 2003;52(Suppl):1-17, 19-20.
10. CDC (US). Technical guidance for HIV/AIDS surveillance programs, volume III: security and confidentiality guidelines. Atlanta, GA: Department of Health and Human Services (US), CDC, 2006. Available at: URL: <http://www.cdc.gov/hiv/topics/surveillance/resources/guidelines/guidance/index.htm>
11. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, Green T, et al. Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. *AIDS Res Hum Retroviruses* 2002;18:295-307.
12. Rehle T, Puren A, Zuma K, Pillay V, Dana P, Shisana O. National HIV prevalence and BED HIV incidence estimates: South Africa 2005. In: Abstract from XIII Conference on Retroviruses and Opportunistic Infections (CROI); 2006 Feb 5-9; Denver, CO.
13. Glynn MK, Lee LM, McKenna MT. The status of national HIV case surveillance, United States 2006. *Public Health Rep* 2007;122(Suppl 1):63-71.
14. Mortimer JY, Salathiel JA. "Soundex" codes of surnames provide confidentiality and accuracy in a national HIV database. *Commun Dis Rep CDR Rev* 1995;5:R183-6.
15. Council of State and Territorial Epidemiologists. Reciprocal (inter-state) notification of HIV cases. 01-ID-04 [cited 2006 Sep 13]. Available from: URL: <http://www.CSTE.org/ps/2001/2001-id-04.htm>
16. Song R, Karon JM, White E, Goldbaum G. Estimating the distribution of a renewal process from times at which events from an independent process are detected. *Biometrics* 2006;62:838-46.
17. Chaisson RE, Keruly JC, Moore RD. Race, sex, drug use, and progression of human immunodeficiency virus disease. *N Engl J Med* 1995;333:751-6.
18. Longini IM Jr, Clark WS, Gardner LI, Brundage JF. The dynamics of CD4+ T-lymphocyte decline in HIV-infected individuals: a Markov modeling approach. *J Acquir Immune Defic Syndr* 1991;4:1141-7.