Estimation of HIV Incidence in the United States

H. Irene Hall, PhD[1], **Ruiguang Song, PhD**[1], **Philip Rhodes, PhD**[1], **Joseph Prejean, PhD**[1], **Qian An, MS**[2], **Lisa M. Lee, PhD**[1], **John Karon, PhD**[3], **Ron Brookmeyer, PhD**[4], **Edward H. Kaplan, PhD**[5], **Matthew T. McKenna, MD**[1], and **Robert S. Janssen, MD**[1] **for the HIV Incidence Surveillance Group**

[1] Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30333

[2] The Ginn Group Inc., Peachtree City, GA 30269

[3] Emergint Corporation, Louisville, Kentucky, 40243

[4] Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205

[5] Yale School of Management, Department of Epidemiology and Public Health, Yale School of Medicine, and Yale School of Engineering and Applied Science, New Haven, CT 06520

Abstract

Context—HIV incidence in the United States has not been directly measured. New assays that differentiate recent versus long-standing HIV infections allow improved estimation of HIV incidence.

Objective—To estimate HIV incidence in the United States.

Design—Remnant diagnostic serum specimens from patients diagnosed with HIV during 2006 in 22 states were tested with the BED HIV-1 capture enzyme immunoassay to classify infections as recent or long-standing. Information was reported to the Centers for Disease Control and Prevention through June 2007. HIV incidence in the 22 states during 2006 was estimated using a statistical approach with adjustment for testing frequency and extrapolated to the U.S. Results were corroborated with back-calculation of HIV incidence for 1977–2006 based on HIV diagnoses from 40 states and on AIDS incidence from 50 states and the District of Columbia.

Setting—Data from 22 states were extrapolated to the U.S.

Patients—Persons newly diagnosed with HIV (age ≥ 13 years).

Main outcome measure—Estimated HIV incidence.

Results—An estimated 39,400 persons were diagnosed with HIV in 2006 in the 22 states. Of 6,864 diagnostic specimens tested using the BED assay, 2,133 (31%) were classified as recent infections. Based on extrapolations from these data, the estimated number of new infections for the U.S. in 2006 was 56,300 (95% confidence interval [CI] 48,200, 64,500); the estimated incidence rate was 22.8 per 100,000 population (95% CI 19.5, 26.1). Forty-five percent of infections were among blacks and 53% among men who have sex with men. The back-calculation (n=1.230 million HIV/AIDS cases reported by the end of 2006) yielded an estimate of 55,400 (95% CI 52,700, 58,100) new infections per year for 2003–2006, and indicated that HIV incidence increased in the mid-1990s, then slightly declined after 1999 and has been stable thereafter.

Conclusions—The estimates are the first direct estimate of HIV incidence in the United States using laboratory technologies that were previously only implemented in clinic-based settings. New HIV infections in the United States remain concentrated among men who have sex with men and African Americans.

Knowledge about trends and current patterns of HIV infections is essential to plan and evaluate prevention efforts and for resource allocation. In the past, data on AIDS incidence and, more recently, data on HIV diagnoses and prevalence have been used for planning and targeting HIV prevention programs. Timely information on national HIV incidence among key U.S. populations can provide a more accurate picture of the HIV epidemic and likely lead to improved reach and impact of domestic programs. However, the incidence of HIV infection in the U.S. has never been directly measured.[1]

In the early 1990s, back-calculation models using AIDS incidence data and the probability distribution of the incubation period from HIV infection to AIDS[2–5] provided historical trends of HIV incidence, but these models could not provide timely data on current transmission patterns. In addition, with the change in the AIDS case definition in 1993 and the advent of effective treatments that slow disease progression to AIDS, back-calculation models based exclusively on incident AIDS cases are no longer valid because the incubation period from HIV infection to AIDS is difficult to estimate and inconsistently ascertained on a population level. Estimates of the annual number of new infections in the U.S. have also been derived from HIV incidence observed in cohort studies.[6] However, this method was based on small, select populations that did not produce population-based estimates and did not provide trends in incidence over time.

The development of laboratory assays that differentiate recent versus long-standing HIV infections now makes it possible to directly measure HIV incidence.[7–9] Building on the existing infrastructure of the Centers for Disease Control and Prevention's (CDC) national HIV/AIDS case reporting system, we used the new technology to implement population-based HIV incidence surveillance. As a part of the new system, remnant serum specimens from persons who have a new diagnosis with a confirmed positive HIV antibody test are tested with a second antibody assay, the BED HIV-1 Capture Enzyme Immunoassay (BED),

[8] that distinguishes recent (on average, 156 days after seroconversion on standard diagnostic assays) from long-standing infections. The BED assay uses antibodies to detect all HIV subtypes (i.e., HIV-1 subtypes B, E, and D gp41 immunodominant sequences are included on a branched peptide used in the assay). The assay detects levels of anti-HIV IgG relative to total IgG and is based on the observation that the ratio of anti-HIV IgG to total IgG increases with time shortly after HIV infection. If a confirmed HIV-1 positive specimen is reactive on the standard sensitive EIA and has a normalized optical density of <0.8 on the BED assay, the source patient is considered recently infected. The combination of diagnostic testing (confirmed HIV antibody positive) followed by testing for recent infection is known as the serologic testing algorithm for recent HIV seroconversion (STARHS).[9]

Estimation of HIV incidence with extended back-calculation models which incorporate all known infected cases and which attempt to use more information about cases than just their AIDS diagnosis date has been performed in Italy, England, and Australia for about the last ten years.[10–12] In the United States, national AIDS surveillance data were used historically for back-calculation of HIV incidence[2–5] while information for extended back-calculation was not available. Recent advances in HIV case surveillance in addition to AIDS case surveillance in the United States have made the use of this approach feasible at the national level. The purpose of this analysis was to estimate HIV incidence in the United States in 2006. We estimated incidence based on the STARHS method and we corroborated this estimate with an extended back-calculation approach using information on HIV diagnoses and AIDS incidence.

METHODS (see Appendix for more detail)

Since 1982, all 50 U.S. states and the District of Columbia have reported AIDS cases to CDC using a standardized case report form. In 1994, CDC implemented data management for national reporting of HIV integrated with AIDS case reporting, at which time 25 states with confidential, name-based HIV reporting started submitting case reports to CDC. Over time, additional states implemented name-based HIV reporting and started reporting these cases to CDC. In 2004, CDC funded selected areas to implement HIV incidence surveillance.[13]

All data were collected as part of routine HIV/AIDS surveillance as mandated by state or local laws or regulations. The HIV incidence surveillance activity was reviewed according to CDC's Guidelines for Defining Public Health Research and Public Health Non-Research[14] and, based upon CDC's interpretation of Title 45 Part 46 of the Code of Federal Regulations[15], CDC has determined in 2005 and again in 2007 that HIV incidence surveillance is not a research activity, and therefore does not require review by an IRB under that regulation. Demographic information, including race/ethnicity, is collected from medical records as part of routine HIV/AIDS surveillance. Because the rates of HIV/AIDS vary widely by race/ethnicity[16] and this information is used to prioritize populations for HIV prevention and care efforts and resource allocation, we included analyses by race/ethnicity. The data analyses for this paper were generated using SAS software, Version 9.1.3 of the SAS System for Windows, SAS Institute Inc.[17] and APL*PLUS III, Manugistics Inc.[18]

Stratified Extrapolation Approach

Analyses were based on all HIV cases (HIV diagnosed with or without concurrent AIDS diagnosis, age ≥ 13 years) diagnosed in 2006 in 22 states (Alabama, Arizona, Colorado, Connecticut, Florida, Georgia, Illinois, Indiana, Louisiana, Michigan, Mississippi, Missouri, New Jersey, New York, North Carolina, Oklahoma, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, Washington) that had confidential, name-based HIV case reporting and HIV incidence surveillance implemented in 2006. Information on HIV cases

was reported to CDC through June 2007. The areas comprise about 73% of all AIDS cases diagnosed in 2006 in the U.S. Information was obtained on age, sex, race/ethnicity (white, black, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native), transmission category (men who have sex with men [MSM], injection drug users [IDU], MSM and IDU [MSM/IDU], heterosexual contact, other), HIV testing history, STARHS result, and antiretroviral treatment. Infections in persons diagnosed with AIDS concurrently or within 6 months after HIV diagnosis were classified as long-standing infections.

Statistical Analyses—We estimated population-based HIV incidence using a statistical approach that is analogous to that used to estimate a population total from a sample survey. [19] In a sample survey, the weight for a sampled person is the inverse of the sampling probability, and the population total (here all persons infected, including those not diagnosed) is the estimated sum of the weights. All persons infected in 2006 (including those not diagnosed) represented the sampling frame and those identified as recently infected represented the sample selected from the sampling frame. Each sampled case was weighted according to the inverse of the estimated probability that a case of similar demographic and risk characteristics was in the sample. The estimated weight depends on the estimated probability that an infected person was tested within 1 year after infection, the probability that a person diagnosed with HIV had a BED test result, and the probability that the BED result for a person tested within 1 year after infection was "recent". The probability of being tested within 1 year after infection was estimated separately for those whose first HIV test was HIV-positive (first time testers) and those who had a previous negative test (repeat testers). For persons who were previously tested, this probability was estimated assuming that the infection date was uniformly distributed from the date of the last negative HIV test to the date of the first positive test. For persons with no previous test, this probability was estimated from a competing events model, the events being an HIV test or an AIDS diagnosis, assuming the HIV testing hazard (likelihood of having an HIV test) was a constant after infection until AIDS diagnosis.

Because HIV testing history and BED results were not available for most cases diagnosed in 2006 (Table 1), a 20-fold multiple imputation procedure[20] was used (36% of cases [n=12,067] had testing history information and 30% of cases [n=6,864] with HIV [not AIDS within 6 months] had a BED test). First we imputed BED values (recent or long-term infection) for HIV cases without AIDS (no AIDS within 6 months after HIV diagnosis) and missing BED test results; then we imputed previous testing status (previously tested or not tested) for cases with missing information on this variable. The time from the last negative test to the first positive test was also generated for cases with missing information on previous test date but assigned to the previously tested group through imputation. See the Appendix for more details.

Case counts were adjusted for reporting delays.[21] Cases reported without risk factor information were redistributed among transmission categories based on the classification of transmission category, by sex, race, and region, of cases that were diagnosed 3 to 10 years earlier and initially reported without risk factor information but that were later reclassified based on information obtained through follow-up investigations.[22] Incidence data from the 22 states were extrapolated to all 50 states and the District of Columbia. We assumed that the ratio of HIV incidence to AIDS incidence in the 22 states is equal to the ratio in the other areas when cases are stratified by sex, race/ethnicity, age, and transmission category.

Point estimates are the mean values of the estimates from the 20 multiple imputation data sets. Confidence interval (CI) estimates were obtained by normal approximation with standard errors of estimates derived using the delta method and include the variability among the 20 data sets.[20,23] We conducted sensitivity analyses to determine whether data on

people who sought testing because of a specific exposure event would bias incidence estimates.

Crude incidence rates per 100,000 population were calculated by sex, race/ethnicity, and age (population denominators were not available by transmission category). Population denominators for rates were based on official post-census estimates for 2006 from the U.S. Census Bureau[24] and bridged-race estimates for 2006 obtained from CDC's National Center for Health Statistics.[25]

Extended Back-calculation Approach

We used an extended back-calculation model based on the earliest time that a case was known to be infected with HIV[11] and a dichotomous measure of disease severity at diagnosis: whether the case received an AIDS diagnosis in the same year as first diagnosed to be HIV-positive. While the original back-calculation methods that relied exclusively on AIDS data could no longer be used because of the changes in the AIDS case definition and the unpredictable time from HIV diagnosis to AIDS in the era of HAART, the extended back-calculation methods we used included information on HIV diagnoses. The current method is not affected by treatments that delay the time until an AIDS diagnosis. We accounted for the revised case definition by calculating AIDS hazards (the AIDS hazard in a designated year is the probability that an individual is diagnosed with AIDS in that year given that he/she was AIDS-free at the beginning of the year) that took into account the different ways of diagnosing AIDS (e.g., OI vs. CD4). We estimated the national HIV incidence per year for 1977–2006 using information from the national HIV/AIDS Reporting System on cases ≥ 13 years of age diagnosed with HIV prior to the end of 2006 and reported to CDC by the end of June 2007. AIDS cases were reported by all states and the District of Columbia for the entire reporting period. Forty states provided both HIV and AIDS diagnoses while 10 states (California, Delaware, Hawaii, Illinois, Massachusetts, Maryland, Montana, Oregon, Rhode Island, Vermont) and the District of Columbia provided only AIDS diagnoses. We included year of HIV diagnosis, year of AIDS diagnosis, state of residence at diagnosis, sex, race/ethnicity, transmission category, and age at first diagnosis.

Adjustments were made to the surveillance data to obtain the estimated number of HIV diagnoses by year and disease severity (i.e., whether a case had AIDS). Adjustments were made for reporting delay, underreporting of cases, detection and elimination of duplicate reports, and misclassification of the first diagnosis date; these adjustments were based on information from prior studies.[21,26]

Statistical Analyses—We first describe the basic structure and assumptions of the original AIDS-only back calculation models that are similar to our extended back calculation model. We then describe the differences and additional features that distinguish the extended model from both earlier AIDS-only models and from other extended models.

Original back-calculation models used the date of AIDS diagnosis in order to estimate HIV incidence. These models estimated the distribution of the time of infection of the observed AIDS cases using assumptions about the distribution of the incubation period for an AIDS diagnosis following HIV infection and the possible shape of the HIV incidence curve. The assumptions about the incubation period also indicated the proportions of infected individuals by year of infection that would be expected to be AIDS-free at the date specified for the analysis. The two sets of estimates were then combined to provide estimates of HIV incidence by year. However, since the proportion of cases progressing to an AIDS diagnosis in the first few years following infection is very low, most of the analyses using the original back-calculation methods did not attempt to estimate HIV incidence across the entire time period from which AIDS cases were available. (do we need the last sentence?)

Much of the information about the distribution of the incubation period from infection to AIDS diagnosis was obtained from studies of the natural disease history of infected individuals during an era of the epidemic prior to important changes in AIDS case definition and prior to the introduction of treatments that increased the incubation time. These factors have been important modifiers of the distribution of the AIDS incubation period. The original back calculation models were sensitive to the accuracy of the assumptions about the incubation period. Thus, in analyses that include eras of the epidemic after the AIDS case definition was changed (i.e., after 1993) and effective antiretroviral treatment was available, it is crucial that appropriate modifications are made to the incubation distribution. This is not an easy task and the difficulty in doing so was a major reason that estimation of HIV incidence in the United States using these models was discontinued around 1997.

By contrast, in our extended back calculation model we include all persons diagnosed with HIV by the end of 2006. The disease history information of interest is the calendar year in which the individual was first diagnosed with HIV along with an indicator of whether the individual was also diagnosed with AIDS during the same calendar year. Our extended back-calculation model includes more cases than an AIDS-only model using the same cut-off date. For many cases, our extended model uses a date closer to the date of infection, i.e., an HIV diagnosis that occurred in a calendar year prior to the calendar year of the AIDS diagnosis.

The relevant 'incubation period' in our extended back calculation model is the time from infection to first HIV diagnosis. The distribution of this period depends on both the rate of progression to AIDS and the rate of diagnosis by HIV testing prior to AIDS among undiagnosed infected individuals. That is, in order to remain undiagnosed from infection to some later time period, an infected individual must avoid diagnoses by either of those reasons in each intervening time period. Since treatments only occur after initial HIV diagnosis, they do not affect the type of incubation period used in the extended model. This is a major strength of the model presented here compared to both the original models and other extended models that make use of dates of both initial HIV diagnosis and a (potentially) later AIDS diagnosis.

The extended model estimates the year of infection conditional on both the calendar year first diagnosed and the stage of disease at diagnosis, i.e., for diagnoses from any particular year, cases that have an AIDS diagnosis at or soon after the initial HIV diagnosis will have a different distribution for the estimated year infected compared with those cases without an AIDS diagnosis at or near the initial diagnosis. The cases with a simultaneous AIDS diagnosis will have an earlier estimated average year of infection compared to those cases without a simultaneous AIDS diagnosis.

The estimation of the year of infection involves three sets of parameters: (1) AIDS hazards by time since infection in untreated infected individuals; (2) HIV testing rate by year in infected individuals prior to AIDS diagnosis; and (3) number of HIV infections by year. The AIDS diagnosis hazards were based on the published literature and assumed to have been correctly specified in our model. The two sets of parameters for HIV testing hazards and the number of HIV infections were estimated by the model subject to some assumptions about relationship of the parameters within each set which is necessary to ensure the stability of the model. Within each set we grouped together calendar years to form time periods in which the parameters within a set were assumed to be constant. For example, for HIV incidence, the thirty years covered by the analysis (1977–2006) was reduced to a smaller number of time intervals, e.g., the model was forced to estimate that the same number of infections occurred in the years 2000, 2001 and 2002. It is important to note that the HIV testing parameters that are estimated here do not represent the rate of HIV testing that occurs

in the general population. Rather, they reflect the rate of removal by HIV testing from the pool of undiagnosed infected individuals that are not close to an AIDS diagnosis. In the simple version of the model where these rates depend only on calendar time but not time since infection, the estimated HIV testing rate for a single calendar year would be calculated as a proportion with the numerator equal to the number of new diagnoses without an AIDS diagnosis in that year divided by a denominator which is equal to the estimated number of undiagnosed cases carried over from the previous calendar year plus new infections occurring in the current calendar year minus the number of new diagnoses that are simultaneous HIV/AIDS cases in the current year.

While fitting models, estimates and goodness of fit statistics were examined to determine whether any adjustments needed to be made to the specified time periods (e.g., whether time periods needed to be broken into smaller time periods). The defining of time periods required a compromise between avoiding too many time periods (and thereby unstable models due to more estimated parameters) and the need for smaller time periods (especially for the early years of the epidemic) to capture the variation likely to be present in the data. The number and lengths of the intervals used to estimate HIV incidence was chosen based both on prior information about the likely shape of the incidence curve at different stages of the epidemic (e.g., steep increases in incidence in the early 1980s , relatively stable incidence from the mid 1990s to the present) and experience gained by evaluating a variety of models with varying numbers of intervals and interval lengths. The HIV testing rates are restricted to be dependent on calendar time, not on time since infection.[27] However, this assumption does not preclude the possibility that within any year there may be groups of infected individuals with different rates of HIV testing (e.g. variation by time since infection). Rather, the assumption merely requires that the average probability of diagnosis via HIV testing is the same across years that were grouped together.

Sensitivity analyses were conducted for the effect of the specified AIDS hazards. We assessed the sensitivity of the model results to the particular values we used by re-fitting the back-calculation model using alternative values for the AIDS hazards that were proportionally larger or smaller than the original values (up to 20% larger or smaller).

RESULTS

Stratified Extrapolation Approach

A total of 33,802 persons aged ≥ 13 years were diagnosed with HIV in 2006 in the 22 states and reported to CDC through June 2007 (adjusted for reporting delays, 39,400). A total of 6,864 HIV cases who were not diagnosed with AIDS within 6 months after HIV diagnosis had BED results (2,133 [31%] were classified as recent infections and 4,731 as long-term). Of 12,067 cases with information on having had a previous test, 7,604 (63%) had a previous negative test. Among the cases that had their specimens BED tested, a slightly higher proportion were black and in younger age groups compared to all cases diagnosed in the 22 states in 2006 (Table 1).

An estimated 56,300 adolescents and adults were newly infected with HIV in 2006 in the U.S. (95% CI 48,200, 64,500) (Table 1), with a rate of 22.8 per 100,000 population (95% CI 19.5, 26.1) (Table 2). Seventy-three percent of the infections occurred among men; 45% among blacks, 35% among whites, and 17% among Hispanics (Table 1). More than half of the infections were attributed to MSM (53%). HIV incidence rates were seven times as high among blacks (83.7; 95% CI 70.9) as among whites (11.5; 95% CI 9.6, 13.4) (Table 2). Rates among Hispanics (29.3; 95% CI 23.8, 35.0) were almost three times as high as rates among whites.

Sensitivity analyses based on data from people who sought testing because of a specific perceived exposure event showed that the incidence estimate would be less than 7% lower than our current estimate, which is within the 95% confidence interval of our estimate.

Back-calculation Approach

Through June 2007, 1.230 million individuals (age ≥ 13 at diagnosis) had been reported to CDC as having been diagnosed with HIV infection (with or without AIDS) by the end of 2006. Accounting for reporting delays, state systems providing only AIDS cases, and under–reporting of HIV cases an estimated 247,000 additional individuals were diagnosed with HIV by the end of 2006 but not yet reported to CDC.

The model estimates indicated that HIV incidence rose sharply after 1977 with a peak in 1984–85 of about 130,000 infections per year (Figure 1). Incidence declined after 1985 and reached a low point in the early-1990s with about 49,000 infections per year. Incidence again peaked in the late 1990s at approximately 58,000 incident infections and decreased to 55,000 per year in the most recent intervals 2000–2006. Incidence among males mirrored the overall trend but among women, incidence rose more slowly until the late 1980s, declined towards the early 1990s and then remained relatively stable.

Throughout most of the epidemic, except in the late 1980s and early 1990s, MSM (not including MSM/IDU) had the largest estimated incidence (Figure 2). The trend in HIV incidence for MSM has been steadily increasing since the early 1990s. For 2003–2006, MSM continued to account for over half of the estimated incidence (Table 1). Blacks, whites and Hispanics, respectively, accounted for about one-half, one-third and one-sixth of current incidence. HIV incidence rose sharply after 1977 among whites with a peak in 1984–85 of more than 72,000 infections per year (Figure 3). Incidence rose more gradually after 1977 among blacks and Hispanics, with peak incidence of about 46,000 infections per year among blacks and about 16,000 infections per year among Hispanics during the late 1980s.

Sensitivity analyses based on re-analyzing the data using different values for the AIDS hazards (±20 %) while retaining the same set of time periods for the testing hazards and the numbers of infections did not change results substantially (data not shown).

DISCUSSION

The national HIV incidence estimates for the United States for 2006 from both methods used are within the range of estimates from back-calculation models in the early to mid 1990s but higher than the CDC estimate from 2001.[6] A back-calculation that accounted for the age-dependent AIDS incubation distributions estimated 55,000 new infections (95% CI 49,500–60,700) for the U.S. each year during 1987–1991.[3] Using an alternative back-calculation method, Rosenberg[4] later reported an average of 40,000 to 80,000 new infections each year from 1987 to 1992. The prior back-calculation estimates were based on national AIDS surveillance data provided by CDC. Another method extrapolating from incidence estimates from studies among convenience samples of MSM to the general U.S. population estimated HIV incidence at approximately 40,000 infections per year.[6] The independence of the methods we used and time frames studied suggest the similar results for 2006 have validity. The discrepancy between our estimate for 2006 based on the stratified extrapolation method and CDC's earlier estimate of 40,000 new infections per year[6] could be due to bias in the current estimate, limitations of the methods used for our previous estimate (e.g., incidence may not have been as low as 40,000), or an increase in HIV incidence.

Our incidence estimate based on the STARHS method could be an overestimate if the proportion of cases classified as recently infected in our sample is higher than that which

would have been observed in the general population of people diagnosed with HIV, or if we underestimated the probability of testing within one year after infection. People who get tested more frequently are more likely to get tested within one year after infection and to be identified as being recently infected. National surveys show differences in testing frequency; for example, a higher proportion of MSM report having had a test within the preceding 12 months[28] than in the general population.[29, 30] However, we attempted to control for a possible bias in our sample by multiple imputation and stratified analyses.

The minor differences between our estimates within some of the subpopulations is likely because of differences between the methods and also because the stratified extrapolation approach provides estimates for 2006 while the extended back-calculation model provides estimates averaged over 4 years (i.e., the confidence intervals reflect model uncertainty but cannot be used to compare the models). The extended back-calculation approach is less suited to identify very recent changes in trends. However, the extended back-calculation model can also provide prevalence estimates that, in context with reported HIV diagnoses and deaths, further corroborate the plausibility of our methods.

Our incidence estimates continue to demonstrate the disproportionate distribution of HIV infection among blacks (incidence rate = 83.7/100,000) and Hispanics (29.3/100,000) compared with whites (11.5/100,000). [16] CDC is working with public health partners and community leaders to address disparities in HIV disease through the Heightened National Response to the HIV/AIDS Crisis among African Americans. [16] Not only will novel, sustained efforts be needed to reduce incidence among African Americans and Hispanics, but increasing the availability of programs will be critical as well.

Overall trends in HIV incidence can mask trends in subpopulations. Based on the back-calculation results, for example, incidence increased nationally in the late 1990s but remained relatively stable among IDUs throughout the mid and late 1990s and then declined. Overall, HIV incidence among IDUs has declined about 80% in the United States. Over that time, IDUs have reduced needle sharing by using sterile syringes available through needle exchange programs or pharmacies and have reduced the number of people with whom they share needles.[31,32] However, the relative contribution of each of these interventions has been difficult to determine.

Currently, we do not have STARHS-based trend data to determine whether the changes in HIV diagnoses in recent years are due to changes in HIV transmission or testing for HIV. [33,34] The results from the extended back-calculation model suggest that HIV incidence among MSM was lowest in the early 1990s and rose thereafter. During this time, annual HIV diagnoses decreased until 1999 and then increased in the 25 low to moderate prevalence states that had HIV reporting.[35] Increases in HIV diagnoses have also been observed in other Western countries.[36] This shows that without incidence data, delays may occur in recognizing a resurgence of HIV infections among certain populations which in turn may delay implementation of needed prevention efforts.

Based on the back-calculation results, incidence trends are also different for the various racial/ethnic groups. The annual HIV incidence among blacks surpassed the incidence among whites in the late 1980s, when incidence among whites declined. Incidence among blacks did not decline substantially until the early 1990s. Incidence among Hispanics, while lower, mirrors the trends among blacks rather than whites. Incidence is low among Asians/Pacific Islanders and American Indians/Alaska Natives, and therefore trends are more difficult to interpret.

Our estimates depend on a number of assumptions that may affect the accuracy of the results. In the stratified extrapolation approach, we assume that information on previous

tests and BED results are missing at random after accounting for all variables known to be associated with missing values in the multiple imputation models. For example, HIV incidence surveillance was implemented in some areas by first enrolling public laboratories to submit specimens for BED testing and then adding additional laboratories; therefore, we controlled for facility type in the imputation models. However, there is the possibility that there are unobserved variables associated with missing previous test or BED results and that associations cannot be explained by the observed variables. We further assumed that testing behavior has not changed significantly over several years, which would affect the probability of testing within 1 year after infection. There is evidence that testing rates have changed little[37] and such changes would have a small effect on our results because a large proportion of persons diagnosed with HIV have been previously tested. A further assumption is that testing and infection are independent; however, there may be a tendency for persons who are recently infected to test in the immediate period following HIV infection. Sensitivity analyses on data from those who sought testing because of a possible exposure event, showed the incidence estimate would be less than 7% lower than our estimate, which is within the 95% confidence interval of our estimate. Bias due to heterogeneity of testing frequency and other possible reasons for early testing such as having a concomitant STD is also minimized by stratifying the population as in our model.

The accuracy of the information on whether cases had a previous negative test is unknown; future studies are needed to validate this information. We extrapolate estimates of HIV incidence from the 22 incidence surveillance states to 50 states and D.C., assuming that the ratio of HIV incidence to AIDS incidence in the 22 states is similar to the ratio in the other areas after adjusting for sex, race/ethnicity, age, and transmission categories. As a proxy, we compared the ratio of HIV to AIDS diagnoses in the 22 states included in our analyses to that ratio in other areas with HIV reporting that were not part of our analyses, and found similar results. The confidence intervals presented reflect random variability and may not reflect model-assumption uncertainty, and should therefore be interpreted with caution. Finally, population denominator data are needed to calculate rates for at-risk populations in the future.

Concerns have been raised about the accuracy of the BED test as incidence appeared to be overestimated when using BED results in Africa and Thailand.[38,39] The primary concern is the misclassification of specimens as "recent" among persons with long-term HIV infection or AIDS, which overestimates the proportion of specimens classified as "recent". To reduce this concern in the U.S., the BED test is not used for people with AIDS. Instead, incidence surveillance systems collect information on disease severity (AIDS) and we classified infections among cases diagnosed with AIDS within 6 months after HIV diagnosis as "long-term". However, we cannot rule out potential misclassification among those who have been infected several years but have not been diagnosed with AIDS. Other factors also differ between the U.S. and some other countries; for example, in the U.S. there are low levels of chronic co-infection (that is, few persons have hypergammaglobulinemia that may yield false recent BED results) and additional information is collected (e.g., last negative test).[40]

Several factors may affect the accuracy of incidence estimates from the extended back-calculation approach, resulting in under- or overestimates of incidence. First, accurate adjustments for reporting delay, underreporting of cases, detection and elimination of duplicate reports, and misclassification of the first diagnosis date need to be made to the surveillance data. Errors in assumptions about contributions from reporting delays and duplicate reports will have much larger effects on estimates of diagnoses in recent years (e.g. 2005, 2006) compared to earlier years. Such errors then would also have a similar pattern of effects on estimates of HIV incidence. The method further depends on accurate specification of the AIDS incubation distribution. Variation in the AIDS diagnosis hazard

appeared to have little effect on results. While fitting models, time periods are combined (i.e., with similar incidence) and an estimate for a particular year may change considerably depending upon the time period in which that year is placed. Finally, the version of the model presented here makes the assumption that the HIV testing hazard is mostly dependent on calendar time and not on time since infection. However, this simplification generally does not distort the HIV incidence estimates as long as the model contains a sufficiently large number of time periods for the HIV testing hazards.

Since 2002, CDC launched new prevention initiatives that included expanding HIV prevention to those living with HIV and increasing HIV testing[41] and expanding the use of proven behavioral interventions in prevention programs for high risk populations.[42] Condoms are highly effective in preventing the sexual transmission of HIV infection, [43] but are frequently not used.[44] HIV counseling and testing has been found to reduce high-risk behavior among those who find they are infected with HIV by about 68%.[45] Most behavioral interventions reduce risk behavior by 20 to over 40%.[46] Many of these interventions have been implemented in prevention programs across the country, but their reach must be considerably expanded to accelerate progress. An estimated one quarter of people living with HIV do not know it, and over a recent one-year period only about 15% of MSM participated in individual-level and 8% in group level interventions, among the most effective behavioral interventions available.[44] Making a substantial reduction in HIV incidence will require wider implementation of the effective interventions currently available and the development of additional interventions, such as antiretroviral chemoprophylaxis or a vaccine. These new HIV incidence data will help ensure that HIV prevention resources are allocated to the populations with greatest need and in the future will be used to monitor the success of these prevention efforts.

References

1. Institute of Medicine. No time to lose: getting more from HIV prevention. Washington: National Academy of Sciences; 2000.

2. Biggar RJ, Rosenberg PS. HIV infection/AIDS in the United States during the 1990s. Clinical Infectious Diseases 1993;(17 Suppl 1):S219–23. [PubMed: 8399920]

3. Rosenberg P. Backcalculation models of age-specific HIV incidence rates. Statistics in medicine 1994;(13):1975–90. [PubMed: 7846404]

4. Rosenberg PS. Scope of the AIDS epidemic in the United States. Science 1995;270:1372–1375. [PubMed: 7481828]

5. Brookmeyer R. Reconstruction and future trends of the AIDS epidemic in the United States. Science 1991;253:37–42. [PubMed: 2063206]

6. Karon JM, Fleming PL, Steketee R, De Cock KM. HIV in the United States at the turn of the century: an epidemic in transition. Am J Public Health 2001;91:1060–1068. [PubMed: 11441732]

7. 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;141:1660172.

8. Parekh BS, Kennedy MS, Dobbs T, et al. Quantitative detection of increasing HIV type I antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. AIDS Research and Human Retroviruses 2002;18:295–307. [PubMed: 11860677]

9. Janssen RS, Satten GA, Stramer SL, et al. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. JAMA 1998;280:42–48. [PubMed: 9660362]

10. Bellocco R, Marschner IC. Joint analysis of HIV and AIDS surveillance data in back- calculation. Statistics in Medicine 2000;19:297–311. [PubMed: 10649297]

11. Cui J, Becker NG. Estimating HIV incidence using dates of both HIV and AIDS Diagnoses. Statistics in Medicine 2000;19:1165–1177. [PubMed: 10797514]

12. Marschner IC. Using time of first positive HIV test and other auxiliary data in back- projection of AIDS incidence. Statistics in Medicine 1994;13:1959–1974. [PubMed: 7846403]

13. Lee LM, McKenna MT. Monitoring the incidence of HIV infection in the United States. Public Health Reports 2007;122(S1):72–79. [PubMed: 17354530]

14. Centers for Disease Control and Prevention. Guidelines for Defining Public Health Research and Public Health Non-Research (Revised October 4, 1999). [Accessed May 7, 2008]. Available at http://www.cdc.gov/od/science/regs/hrpp/researchdefinition.htm

15. Department of Health and Human Services. Code of Federal Regulations, Title 45 — Public Welfare, Part 46, Protection of Human Subjects. Department of Health and Human Services; Washington, DC: [Accessed May 7, 2008]. Revised June 23, 2005. Effective June 23, 2005. Available at http://www.hhs.gov/ohrp/documents/OHRPRegulations.pdf

16. Centers for Disease Control and Prevention. Racial/ethnic disparities in diagnoses of HIV/AIDS—33 states, 2001—2005. MMWR 2007;56:189–193. [PubMed: 17347642]

17. SAS Institute. SAS software, Version 9.1.3 of the SAS System for Windows. Copyright © 2002–2003. SAS Institute Inc; Cary, NC, USA:

18. Manugistics, Inc.. APL*Plus III. Manugistics Inc; Rockville, MD: 1995.

19. Karon KM, Song R, Kaplan E, Brookmeyer R, Hall HI. Estimating HIV incidence in the United States from HIV/AIDS surveillance data and biomarker HIV test results. Statistics in Medicine. (in press).

20. Rubin, DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons, Inc; 1987.

21. Green TA. Using surveillance data to monitor trends in the AIDS epidemic. Stat Med 1998;17:143–154. [PubMed: 9483725]

22. Song R, Hall HI, Frey R. Uncertainties associated with incidence estimates of HIV/AIDS diagnoses adjusted for reporting delay and risk redistribution. Stat Med 2005;24:453–464. [PubMed: 15532097]

23. Bishop, YMM.; Fienberg, EF.; Holland, PW. Discrete multivariate analysis. Cambridge, Mass: MIT Press; 1975.

24. U.S. Census Bureau. Population estimates: entire data set. Jul 12006 [Accessed November 30, 2007]. Also available at: http://www.census.gov/popest/estimates.php

25. National Center for Health Statistics. Bridged-race Vintage 2006 postcensal population estimates for July 1, 2000—July 1, 2006, by year, county, single-year of age, bridged-race, Hispanic origin, and sex. [Accessed December 31, 2007]. Available at: http://www.cdc.gov/nchs/about/major/dvs/popbridge/datadoc.htem#vintage2006

26. Hall HI, Li J, Campsmith M, Sweeney P, Lee LM. Date of first positive HIV test: reliability of information collected for HIV/AIDS surveillance in the United States. Public Health Reports 2005;120(1):175–181.

27. Aalen OO, Farewell VT, De Angelis D, et al. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. Statistics in Medicine 1997;16:2191–2210. [PubMed: 9330428]

28. Centers for Disease Control and Prevention. HIV prevalence, unrecognized infections, and HIV testing among men who have sex with men – five U.S. cities, June 2004—April 2005. MMWR 2005;54:597–601. [PubMed: 15973239]

29. Anderson JE, Carey JW, Taveras S. HIV testing among the general US population and persons at increased risk: information from national surveys, 1978—1996. AJPH 2000;90:1089–1095.

30. Centers for Disease Control and Prevention. HIV Testing – United States, 2001. MMWR 2003;52:540–545. [PubMed: 12803192]

31. Santibanez SS, Garfein RS, Swartzendruber A, Purcell DW, Paxton LA, Greenberg AE. Update and overview of practical epidemiologic aspects of HIV/AIDS among Injection Drug Users in the United States. J of Urban Health 2006;83:86–100. [PubMed: 16736357]

32. Des Jarlais DC, Perlies T, Arasteh K, et al. Reductions in hepatitis C virus and HIV infections among injection drug users in New York City, 1990–2001. AIDS 2005;19(suppl 3):S20–S25. [PubMed: 16251819]

33. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2005. Vol. 17. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2007. p. 10Rev edAlso available at: http://www.cdc.gov/hiv/topics/surveillance/resources/reports/

34. Hall HI, Byers RH, Ling Q, Espinoza L. Racial/ethnic disparities in HIV prevalence and disease progression among men who have sex with men. Am J Public Health 2007;97:1060–1066. [PubMed: 17463370]

35. Centers for Disease Control and Prevention. Diagnosis and reporting of HIV and AIDS in states with HIV/AIDS surveillance --- United States, 1994--2000. MMWR 2002;51:595–598. [PubMed: 12139203]

36. UNAIDS. AIDS epidemic update: special report on HIV/AIDS. Dec2007 [Accessed January 23, 2008]. Available at http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf

37. Centers for Disease Control and Prevention. HIV counseling and testing at CDC-supported sites— United States, 1999–2004. 2006 [Accessed April 30, 2007]. p. 4-7.Available at: http://www.cdc.gov/hiv/topics/testing/reports.htm

38. Karita E, Price M, Hunter E, et al. Investigating the utility of the HIV-1 BED capture enzyme immunoassay using cross-sectional and longitudinal seroconverter specimens from Africa. AIDS 2007;21:403–408. [PubMed: 17301558]

39. UNAIDS. Statement on the use of the BED-assay for the estimation of HIV-1 incidence for surveillance or epidemic monitoring. Report of a meeting of the UNAIDS Reference Group for Estimates, Modeling and Projections. Athens, Greece, December 13–15th 2005. Geneva: [UNAIDS] 2005. Statement following a meeting of the UNAIDS Reference Group for Estimates, Modelling and Projections held in Athens, Greece, December 13th 2005. Uploaded December 2005.

40. Centers for Disease Control and Prevention. Using the BED HIV-1 Capture EIA Assay to Estimate Incidence Using STARHS in the Context of Surveillance in the United States. Also available at: http://www.cdc.gov/hiv/topics/surveillance/resources/factsheets/BED.htm

41. Centers for Disease Control and Prevention. Advancing HIV prevention: new strategies for a changing epidemic—United States, 2003. MMWR 2003;52:329–32. [PubMed: 12733863]

42. Centers for Disease Control and Prevention, HIV/AIDS Prevention Research Synthesis Project. Compendium of HIV prevention interventions with evidence of effectiveness. Atlanta, GA: Centers for Disease Control and Prevention; Nov. 1999 Revised. Also available at: http://www.cdc.gov/hiv/resources/reports/hiv_compendium/index.htm

43. Warner L, Stone KM, Macaluso M, Buehler JW, Austin HD. Condom use and risk of gonorrhea and Chlamydia: a systematic review of design and measurement factors assessed in epidemiologic studies. Sexually Transmitted Diseases 2006;33:36–51. [PubMed: 16385221]

44. Centers for Disease Control and Prevention. Human immunodeficiency virus (HIV) risk, prevention, and testing behaviors – United States, National HIV Behavioral Surveillance System: men who have sex with men, November 2003—April 2005. Surveillance Summaries, July 7, 2006. MMWR 2006;55(SS-6)

45. Marks G, Crepaz N, Senterfitt JW, Janssen RS. Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States. JAIDS 2005;39:446–453. [PubMed: 16010168]

46. Crepaz N, Lyles CM, Wolitski RJ, et al. Do prevention interventions reduce HIV risk behaviours among people living with HIV? A meta-analytic review of controlled studies. AIDS 2006;20:143–157. [PubMed: 16511407]

47. Longini IM, Clark WS, Gardner LI, et al. The dynamics of CD4+ T-lymphocyte decline in HIV-infected individuals: A Markov modeling approach. JAIDS 1991;4:1141–1147.

Appendix: Methods

Stratified Extrapolation Approach

Since males have more transmission categories than females, multiple imputation was carried out separately for males and females. Variables that were associated (chi-square test,

p<0.05) with having a BED test or previous testing status or that were associated with missing values in these variables were included in the imputation models. The variables included in imputing BED values were race/ethnicity, age at diagnosis, transmission category, facility type where HIV was diagnosed, and having ever tested HIV-negative. The variables included in imputing previous testing status were race/ethnicity, age at diagnosis, transmission category, facility type where HIV was diagnosed, AIDS within 6 months after diagnosis, whether the BED result is imputed, and BED result. After imputation, all cases have a BED result and information on previous test.

To control for heterogeneity in testing frequency, newly diagnosed cases were stratified by sex, race/ethnicity, age group, and transmission category. Age groups were formed based on the age at HIV infection. Since the exact age at HIV infection was unknown, it was estimated based on the HIV status at diagnosis. For individuals with a previous negative test, the age at HIV infection was assumed to be the age at the mid point of the interval from the last negative test to the first positive test. For individuals with no test prior to HIV diagnosis, we assigned the age at HIV infection: as either 8 years younger than the age at HIV diagnosis if AIDS was diagnosed at the time of HIV diagnosis; [47] as 4 years younger than the age at HIV diagnosis if AIDS was not diagnosed and the BED result indicated long term infection; or the same as the age at diagnosis if the BED result indicated recent HIV infection. Due to the small numbers of cases, cases in the race groups Asian/Pacific Islander and American Indian/Alaska Native were not stratified by other variables and cases in the transmission category MSM/IDU were not stratified by age.

Within each stratum, cases were further divided into two subgroups based on previous testing status: repeat testers and first-time testers. Within each subgroup, incidence was estimated by the number of BED-recent specimens divided by the probability of being classified as BED-recent. Because all persons without AIDS within 6 months after their HIV diagnosis have a BED result in the imputed data, the probability of these persons being classified as BED-recent is the product $p_1 * p_{1w}$, where p_1 is the probability of being tested within 1 year after infection, and p_{1w} is the probability of having a BED test result indicating recent infection if the test is within 1 year after infection. The latter probability is approximately equal to the mean window period for the BED testing algorithm (156/365 years; CDC unpublished data). The window period is the time from seroconversion to the time when the individual's serum, if tested using the BED test, would reach an optical density level predetermined to distinguish recent from long-standing infections. For repeat testers, p_1 is estimated based on the time from the last negative to the first positive test for each individual (reported or imputed) in the group. For first-time testers, p_1 is determined by the testing hazard, which is based on the proportion of cases with AIDS diagnosed at the time of HIV diagnosis in this group. These estimates are approximately 0.60 (range 0.41 – 0.71) and 0.24 (range 0.13 – 0.51), respectively.

The standard errors for the incidence estimates (derived using the delta method[20]) incorporate uncertainties associated with imputation, the observed number of BED test results indicating recent infection, estimates of p_1 and the mean BED window period, adjustments for reporting delay, risk redistribution, extrapolation to the nation, and the covariance among groups for which estimates were made resulting from the inclusion of p_w in each estimate.

Crude incidence rates per 100,000 population were calculated by sex, race/ethnicity, and age (population denominators were not available by transmission category). Population denominators for rates were based on official post-census estimates for 2006 from the U.S. Census Bureau[24] and bridged-race estimates for 2006 obtained from CDC's National Center for Health Statistics.[25] The bridged estimates were based on the Census 2000 counts and

produced under a collaborative agreement with the U.S. Census Bureau. These estimates result from regrouping the 31 race categories used in the Census 2000 (1997 standard of the Office of Management and Budget) for the classification of data on race and ethnicity to the 4 race categories of the 1977 standard and, therefore, to correspond to the HIV data.

Extended Back-calculation Approach

A Kx2 table (K=number of years) of the estimated number of new diagnoses by calendar year and disease severity at diagnosis (whether AIDS was diagnosed within the same calendar year as HIV) served as the input data for the back-calculation model. A discrete-time probability model (calendar year) for the observed diagnosis data was based on three sets of parameters whose properties are described below. Because surveillance data were incomplete for a variety of reasons, a number of adjustments were necessary. For underreporting of HIV cases, we estimated the number of diagnosed but unreported HIV/ not AIDS cases for areas with either AIDS-only surveillance or with incomplete combined HIV/AIDS surveillance. We defined strata based on sex, race, transmission risk group, and year of diagnosis. Within these strata, for the 30 states with mature HIV/AIDS surveillance systems we computed the proportion of diagnosed cases that were still HIV/not AIDS as of the end of 2006 and adjusted the number of HIV/not AIDS diagnoses in the other states/ areas to match these proportions. Adjustments were also made for reporting delay, detection and elimination of duplicate reports, and misclassification of the first diagnosis date; these adjustments were based on information from prior studies.[21,26]

The model parameters for the extended back-calculation approach include 1) the number of infections per year; 2) the AIDS diagnosis (discrete) hazards; and 3) the HIV testing (discrete) hazards. The number of infections per year was estimated subject to constraints with categorical structure. Time periods were defined such that the number of infections was forced to be the same for each year within a time period. The AIDS diagnosis (discrete) hazards were completely specified, not estimated. These values only depend on time since infection, not on calendar time. The AIDS diagnosis hazard values used here are similar to those described by Aalen[23] from a Markov model which included staged declines of CD4 counts, progression to AIDS by occurrence of opportunistic infections (OIs), and/or diagnosis by HIV testing. The hazards used in Aalen's model were modified to account for the U.S. AIDS case definition which is based on either the occurrence of OIs or immunologic criteria related to CD4 cell counts. One prominent feature of this set of hazards is the flattening of the curve at times distant from infection. The HIV testing (discrete) hazards were estimated subject to categorical constraints. The testing hazards were assumed to depend only on calendar time and not on time since infection. A 'categorical' structure was imposed, i.e., time periods were defined such that years within the same period were forced to have the same testing hazard. Note, in this instance (calendar time dependence) to ensure identifiable and stable estimates, the time periods defined for the HIV testing parameters cannot be identical or too similar to the time periods specified for the number of infections.

The discrete hazards represent conditional probabilities for the two types of diagnosis (disease severity). Due to the discrete time framework, we specified that within the same time period, an AIDS diagnosis took precedence over diagnosis by HIV testing. Thus, within each time period the undiagnosed individual was at risk first to receive an AIDS diagnosis and only if no AIDS diagnosis occurred was the individual then at risk for being diagnosed by HIV testing.

The expected values of the observed data in any year (i.e. the two types of diagnoses by time) can be written as a linear function of the incidence in years prior to and including the

current year with weights that are a function of the AIDS diagnosis and HIV testing hazard values in the same set of years. We assumed that the diagnosis counts have Poisson distributions with expectations that are linear, not log-linear, as described above.

We used an expectation-maximization (EM) algorithm to estimate the unknown parameters in the back-calculation model. After specifying some initial starting values for the unknown parameters, the algorithm alternates between an expectation step, which calculates an "expanded" version of the observed data set that is both consistent with the specified model structure and current "working" parameter values, and a maximization step, which re-estimates the parameter values using the observed and the expanded data. In this case, the expanded data set consists of the number of diagnoses by time of infection, type of diagnosis and time of detection.

Variance estimates for estimated HIV incidence or testing hazard values took into account 1) the variability in the (estimated) diagnosis data that served as input to the back-calculation model and 2) the variability arising from the back-calculation model (including the effects of estimating other parameters). Operationally, the overall variability was estimated by a multiple imputation approach which incorporated multiple estimates of relevant values (e.g. estimated diagnoses by time and disease severity at HIV diagnosis).
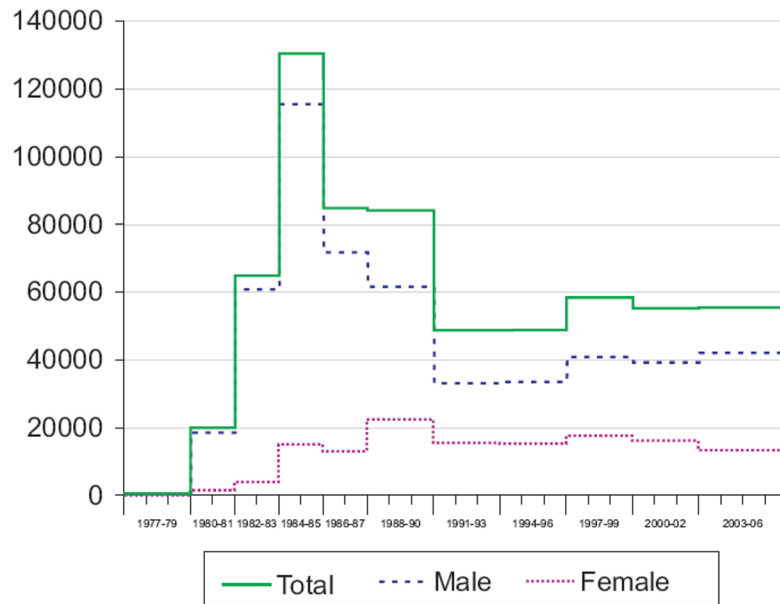
Acknowledgments

Figure 1.
Estimated number of new HIV infections, extended back-calculation model, 50 U.S. states and the District of Columbia, 1977–2006
Extended back-calculation approach.
Tick marks denote beginning and ending of a year.
The model specified time periods within which the number of HIV infections were assumed to be approximately constant.
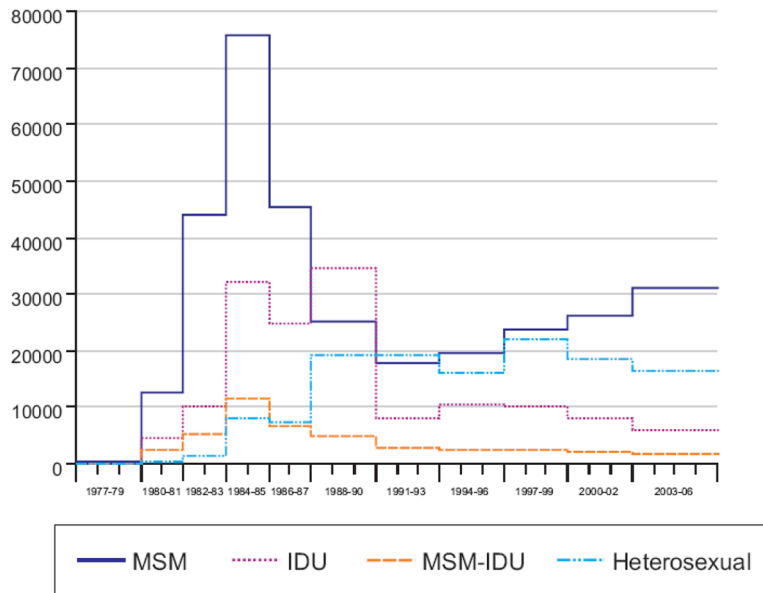
Figure 2.
Estimated number of new HIV infections by transmission category, extended back-calculation model, 50 U.S. states and the District of Columbia, 1977–2006
Extended back-calculation approach.
Tick marks denote beginning and ending of a year.
The model specified time periods within which the number of HIV infections were assumed to be approximately constant.

Figure 3.
Estimated number of new HIV infections, by race/ethnicity (a. all race/ethnicity categories and b. Asian/Pacific Islander and American Indian/Alaska Native), extended back-calculation model, 50 U.S. states and the District of Columbia , 1977–2006
Extended back-calculation approach.
Tick marks denote beginning and ending of a year.
The model specified time periods within which the number of HIV infections were assumed to be approximately constant.
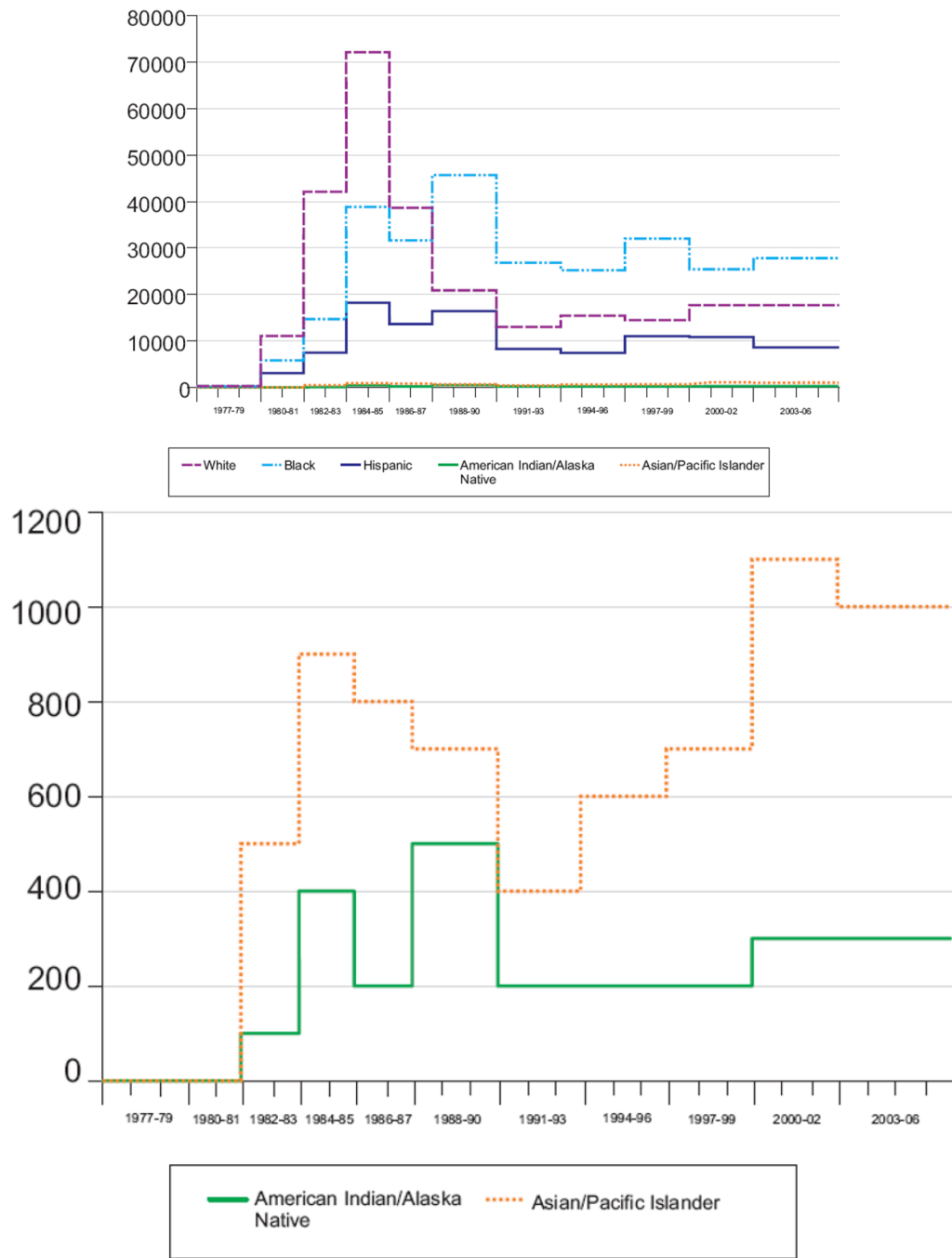
Table 1

Estimated HIV incidence, 50 U.S. states and the District of Columbia

	Stratified Extrapolation Approach										Extended Back-calculation Approach		
	22 States[a]						50 States + DC				50 States + DC		
	BED Tested[b]		2006 Diagnoses[c]		2006 Incidence		2006 Incidence				Incidence per year, 2003–2006		
	No.	%	No.	%	No.	%	No.	%	95% CI[d]		No.	%	95% CI[d]
Total	6864		39400		40800		56300		48200	64500	55400		50000	60800
Sex													
Male	4892	71	28900	73	29300	72	41400	73	35100	47700	42000	76	37400	46600
Female	1972	29	10600	27	11500	28	15000	27	12600	17300	13400	24	11000	15800
Race/ethnicity[e]													
White	1707	25	11400	29	13100	33	19600	35	16400	22800	17700	32	14700	20700
Black	3825	56	20000	51	19600	49	24900	45	21100	28700	27800	50	24200	31400
Hispanic	1190	17	7000	18	6800	17	9700	17	7900	11600	8600	16	6200	11000
Asian/Pacific Islander	78	1	440	1	590	1	1200	2	490	1900	1000	2	200	1800
American Indian/ Alaska Native	21	0	130	0	180	0	290	1	60	500	300	<1	50	700
Age, years													
13–29	2790	41	13100	33	14100	35	19200	34	16300	22200	21200	38	17000	25400
30–39	1892	28	12100	31	12500	31	17400	31	14600	20200	16800	30	13600	20000
40–49	1539	22	9800	25	9900	24	13900	25	11700	16100	12300	22	9100	15500
50–99	643	9	4400	11	4300	11	5800	10	4600	7100	5100	9	2900	7300
Transmission category[f]													
MSM	3582	52	18400	48	20100	51	28700	53	24300	33100	31200	56	25400	37000
IDU	749	11	5600	15	4900	12	6600	12	5300	7900	5900	11	3500	8300
MSM/IDU	182	3	1200	3	1400	3	2100	4	1500	2700	1600	3	400	2800
Heterosexual	2328	34	13100	34	13100	33	16800	31	14200	19400	16400	30	12600	20200

[a] Alabama, Arizona, Colorado, Connecticut, Florida, Georgia, Illinois, Indiana, Louisiana, Michigan, Mississippi, Missouri, New Jersey, New York, North Carolina, Oklahoma, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, Washington

[b] The numbers for "BED tested" do not count cases who were diagnosed with AIDS at or within 6 months after HIV diagnosis and they were risk redistributed but not adjusted for reporting delay.

[c] The numbers for 2006 diagnoses were adjusted for reporting delay and risk redistribution.

[d]CI, confidence interval. The confidence intervals reflect model uncertainty but cannot be used to compare the models.

[e]Race and transmission category subgroup numbers may not add up to the overall total because cases with unknown race or unknown transmission categories are excluded. However, percents are adjusted for the exclusion so they add up to 100%.

[f]MSM, men who have sex with men; IDU, injection drug users.

Table 2

Estimated rates of new HIV infections, 50 U.S. states and the District of Columbia, 2006[a]

	Rate[b]	95% CI	
Total	22.8	19.5	26.1
Sex			
Male	34.3	29.1	39.5
Female	11.9	10.0	13.7
Race/ethnicity			
White	11.5	9.6	13.4
Black	83.7	70.9	96.5
Hispanic	29.3	23.8	35.0
Asian/Pacific Islander	10.3	4.2	16.3
American Indian/ Alaska Native	14.6	3.0	25.2
Age, years			
13–29	26.8	22.8	31.0
30–39	42.6	35.7	49.4
40–49	30.7	25.8	35.6
50–99	6.5	5.1	7.9

[a]Stratified extraploation approach. See table 1 for numerator information.

[b]Per 100,000 population, post-census estimates from the U.S. Bureau of the Census.