HIV Incidence Estimation Consultation
February 23, 2007
Response to Consultant Comments

Because there was considerable overlap in the comments made by the consultants, the responses below have been organized based on the overall set of questions and issues raised rather than as separate, specific responses to each consultant's comments.

1. Accuracy of Self-reported HIV Test Dates (SSA/SEA)

Several consultants raised concerns about the accuracy of the self-reported time from the last negative HIV test to the first positive HIV test, T. Given that the observed proportion of BED-recent specimens among persons with T < 1 year is significantly lower than the expected proportion, it appears that a telescoping bias exists in recalling the most recent testing date because the estimated BED window distribution is unlikely to be incorrect. On the other hand, the observed proportion of cases with AIDS at HIV diagnosis among those with T > 3 years is significantly lower than the expected proportion. One consultant recommendation was to adjust T so the observed and expected proportions match. We performed an analysis based on adjusting T to match (a) the observed and expected proportions of BED-recent specimens among persons with T < 1 year and (b) the observed and expected proportions of AIDS at HIV diagnosis, but found no substantial changes in the incidence estimates. Therefore, we decided not to adjust the self-reported T.

2. One Combined SSA/SEA Method (SSA/SEA)

Several consultants suggested combining the two BED-based methods since they use the same data to estimate incidence. After careful comparison of the advantages and disadvantages of each method, we have developed a new method, the *stratified* extrapolation approach (i.e., a new SEA, but not the *simplified* extrapolation approach), that takes advantage of the unique strengths of each of the previous methods. We use multiple imputation to handle missing data as we did before with both methods. We use stratification to form testing- and incidence-homogenous groups as was done with the SSA method, estimate the probability of being detected as BED recent as was done with the (previous) SEA method, adjust for reporting delay and redistribute missing risk cases, and extrapolate the results to get incidence estimates for larger population groups.

3. HIV Testing Hazards (Back-calculation)

Consultants raised two distinct issues with regard to HIV testing hazards: whether the hazard function should be dependent on time since infection, and whether the HIV testing probability in the year following infection, p_1, used in the SSA/SEA methods could be incorporated into the back-calculation model.

The current, simple, form of the back-calculation model assumes that the HIV testing rates have the form $h_{ij} = h_j$, which makes the testing hazard appear to be a function only of calendar time.

However, because of the discrete-time aspect of the model, this functional form has some unique implications for recently infected cases compared with older cases.

The back-calculation model is set up as a discrete-time (calendar year) model, which allows for a number of simplifications in the estimation procedures. However, as a result, most or all of the event probabilities over a full calendar year (e.g., HIV testing hazards) are not small. This has the effect that the HIV testing hazard for a particular calendar year has a different implication when considering time intervals defined in terms of time since infection (e.g., within one year of infection) for the diagnosis of cases infected in that same year compared to still undiagnosed cases infected in earlier years. For example, suppose the HIV testing hazard for 1999 is 0.15. The current model structure assumes that 15% of new infections in 1999 will be diagnosed via HIV testing in 1999 (assuming essentially none of these new infections are diagnosed with AIDS in 1999). However, in terms of the time interval "within one year of infection" in the underlying continuous model, some of the cases infected in 1999 but diagnosed via HIV testing in 2000 will have been diagnosed within one year of infection. Thus, the model actually predicts that the 1-year probability of diagnosis via testing for cases infected in 1999 is approximately 0.228 (additional assumptions would be needed to calculate the exact value). We did fit alternative versions of the model that forced the 1-year diagnosis probability for new infections to equal the same value for older infections in the same calendar year. While the estimates of the testing hazard parameters did change somewhat, the total estimated infections as well as incidence estimates for particular time periods showed relatively little change.

The difficulty with having the HIV testing hazards depend *only* on time since infection becomes apparent when one considers that, at least early in the epidemic, the availability of HIV testing depends strongly on calendar time. For example, diagnoses via HIV testing were rare prior to 1985 but became common in 1985, when HIV testing first became widely available. On the other hand, evidence from other studies supports the notion that HIV testing hazards depend on time since infection along with calendar time. For example, the high probability of testing within the year following infection used in the BED analysis supports this idea since a continuation of such a high rate would imply that almost all cases would be found prior to the onset of AIDS symptoms, which does not conform to current experience.

The back-calculation model we use can be extended to include a specified dependence of the HIV testing hazard on time since infection. However, because of the change from essentially no testing to high rates of testing during the mid 1980s, it would be reasonable to expect that the effects of calendar time and time since infection show some, if not strong, interaction. In other words, as testing first becomes available, calendar time is the dominant time factor; after testing has been available for, say, 5–10 years, time since infection might become the dominant factor. Another thing to consider is that while such effects can be forced into the back-calculation model, the only validation that these additional effects are accurately modeled is indirect. That is, there is nothing in the surveillance diagnosis data that can directly validate the notion that HIV testing rates are higher for persons recently infected compared to those with older infections. Thus, the validation, or lack of it, is derived from whether the predicted diagnosis values have been "improved" by incorporating the additional effects.

With regard to incorporating the HIV testing parameter used in the SSA/SEA methods into the back-calculation model, one strength of the back-calculation model is that the actual values of the testing rates are estimated from the data, conditional on the particular specified parameterization. Also, the BED model does not provide any direct information on what the testing probabilities should be outside the 1-year period following infection. Additionally, we have investigated the robustness of the HIV incidence estimates from a back-calculation model which only looks at calendar time dependence when in fact the dependence is on both calendar time and time since infection. The general result was that if one were to add a sufficient number of parameters (time periods) for the HIV testing hazards to the model, the amount of bias in the HIV incidence estimates would be minimal. One can evaluate the goodness-of-fit of the model at various stages in order to assess whether more time periods for HIV testing rates should be introduced. Models using data sets with large numbers of cases such as the CDC surveillance data should be able to tolerate a large number of time periods for the testing hazards.

In summary:

- There is some, limited, dependence of HIV testing hazards on time since infection built into the current model.
- It is useful to explore the ability of the current back-calculation model to provide an accurate assessment of a more general type of dependence of the HIV testing rates on time since infection, but at this stage we do not have sufficient experience with the model to feel confident that adding complications to the model will clarify rather than confuse the situation.
- Other modeling strategies such as increasing the number of calendar time periods in the model provide an alternative means to control possible bias in the incidence estimates.
- Forcing one value, p_1, from a very different type of model into the back-calculation model, while seemingly appealing, may overlook large differences in the two models.

4. Additional Modeling

SSA/SEA: Using the (new) SEA method, we estimated 2005 incidence by sex (male, female), race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, A/PI, and AI/AN), age (13–29, 30–39, 40–49, 50+), and transmission category (MSM not IDU, IDU, MSM/IDU, and high-risk heterosexual contact). Because of limited numbers of cases, incidence estimates for the A/PI and AI/AN race groups are not stratified by other variables, and estimates for the MSM/IDU transmission category are not stratified by age.

Back-calculation: The back-calculation model has also been used to provide HIV incidence estimates separately, but not for all possible combinations, by sex, race/ethnicity, age, and transmission categories, using the same breakdown as above except that estimates for an "all other" transmission category were included. The current manuscript contains back-calculation estimates for each group for the time period 2003–2005. Back-calculation estimates for the entire epidemic through 2005 by race/ethnicity and by transmission category are also included in the manuscript.

5. Multiple Imputation (SSA/SEA)

Because the incidence data are not missing completely at random, any method that simply extrapolates results based on the proportion of missing data is potentially biased. Multiple imputation is an appropriate approach for handling missing data. To impute the missing values of the two key variables, previous test status and BED result, we examined the correlation between the missingness of these two variables and other observed variables. Variables that were associated with either the missingness or the distribution of the two key variables were included in the imputation model so that biases due to these correlations are eliminated. The imputed values can still be biased if missingness depends on something that cannot be explained by the observed variables, but we are not able to check this based on the available incidence data. Since the consultation, we have increased the number of iterations to make sure that the generated samples are independent and have also increased the number of imputations to make estimates more stable. Neither increase had a substantial impact on the incidence estimates.

One specific concern regarding multiple imputation is the treating of missing transmission category as an additional categorical level. Unfortunately, the incidence dataset does not contain all the information required to generate an unbiased distribution of transmission category. For example, generating this distribution requires information on the reclassification of cases with no reported risk (NRR) at their initial reports and this information takes many years to collect. The incidence dataset contains only the most recent year of diagnosed cases. In addition, we have demonstrated in a separate imputation study that even if the incidence dataset contained all necessary information about the reclassification of NRR cases, there is no significant difference between treating missing transmission category as a separate group and imputing the transmission category when missing.

An additional concern regarding multiple imputation is the bias associated with the imputation of binary variables. We had noticed this problem and developed a correction method that we applied to the imputation process. The bias is more complex when the imputed categorical variable is not binary, which provides another reason for not imputing the transmission category variable.

6. AIDS Hazards (Back-calculation)

We have re-fit several data sets using alternative AIDS hazards that have a convex shape in the first 7 years, then are constant starting in year 8, but have a median time to diagnosis which is similar to that of the AIDS hazards currently used in the model. We forced a convex shape on the hazard function by employing a Weibull-type hazard (of the form $a*t^b$, where t = time). The general effect we have observed in our initial investigations has been an increase in the number of estimated infections in the early years of the epidemic. The peak incidence has been moved back in time about 2 years and the overall incidence curve is considerably flatter near the maximum compared to the results for the AIDS hazards we generally used to fit our back-calculation models. In general the number of estimated infections was increased although usually this increase was not large. Estimates in recent periods generally were unaffected by the change in AIDS hazards.

7. Sensitivity of Model to Changes in Diagnosis Data Early in the Epidemic (Back-calculation)

We approached this issue from two slightly different points of view. First, our back-calculation model produced estimates of HIV incidence for the time period 1978–2005. One might ask: How do we know where to start estimating? Would it make a big difference to the overall model results if we included earlier time periods, say 1976–1977, into the model? What if we included several earlier time periods and in fact these periods went back well before the actual start of the epidemic? The answer to these questions is that the back-calculation model will behave in a proper manner if one includes time periods that are too early. This is because the AIDS hazards are regarded as known in the version of the model used here and these hazards only depend on time since infection. Thus, if one includes an early time period but sees no, or very few, cases within a reasonable time following this period then the model will sensibly estimate that no, or very few, infections occurred during the early time period. Of course, the model does not provide automatic protection from starting too late, so in the initial modeling stages one should begin by erring on the side of too early in order to establish a reasonable starting point.

The second approach is to ask a different set of questions: What would happen if the numbers of cases reported as diagnosed in some early years of the epidemic were incorrect and in fact the actual numbers were somewhat more or less than the reported numbers? What if the disease stage at diagnosis were misclassified in an appreciable proportion of these early cases? The answers here depend in part on (a) how early, (b) how much error, and (c) the estimates on which we choose to focus. To a great extent, our interests in the present study are HIV estimates for recent years, after 2000, and estimates of such quantities as the current number of undiagnosed cases. Errors in the diagnosis data in early years of the epidemic will have almost no effect on these estimates of main interest to us, for two reasons. One, the numbers of diagnoses in the early years are substantially smaller than the numbers in recent years. More importantly, the size of the assumed, untreated, AIDS hazards and the estimated HIV testing hazards imply that only a very small number, or proportion, of cases infected early in the epidemic, through, say, 1989, would remain undetected in 2005. Thus, these cases cannot make much of a contribution to the estimated pool of undiagnosed cases in 2005. Additionally, those cases first diagnosed in 2003–2005 also have only a small chance of having been infected during the early part of the epidemic.

Thus, while changes in the number or status of early diagnosed cases may result in changes in parameters associated with the earlier time periods, such changes have little or no effect on estimates of parameters associated with recent time periods.

8. Uncertainty

SSA/SEA: For the SEA estimates, in addition to the uncertainties that had been considered earlier—due to imputation and estimating the probability of being detected as BED-recent—we have now incorporated uncertainty due to extrapolation, reporting delay and risk redistribution adjustments, and the estimation of the mean window period.

Back-calculation: There are four major sources of variability for the HIV incidence estimates obtained from the back-calculation model, two owing to the necessity of adjusting the input diagnosis data, and two due to the model itself.

About 15% of all eligible, observed cases in the CDC surveillance data set had a missing transmission category. Using cases with a known transmission category, we modeled the relationship between reported transmission category and various factors such as sex, race/ethnicity, region of the country, time first diagnosed, disease status, and vital status. We fit separate models by sex since males are classified into five possible transmission categories while females are classified into only three. For purposes of data tabulation and obtaining initial back-calculation estimates, we used the modeled probabilities as weights that were attached to the multiple instances of these cases (weights added to 1 within each case). At a later stage, however, we used the modeled probabilities to create ten multiple imputations of the missing transmission category variable for each of the cases missing this information in order to incorporate variability due to this additional estimation step. (Note that this uncertainty only affects incidence estimates for transmission category sub-groups.)

The input data, initial HIV diagnoses by calendar year and disease status, to the back-calculation model require several types of adjustments because of delays in case reporting, delayed reporting of case information, later ascertainment of duplicates among reported cases, under-reporting of cases from states with AIDS-only reporting systems or recently instituted HIV reporting systems, and misclassification of dates first diagnosed with HIV. We created multiple imputed versions of the diagnosis data by taking into account our assumptions or empirical observations about the variability associated with each of the adjustments mentioned above. These data sets could differ both on the total number of diagnoses as well as their relative distribution by time and disease stage at diagnosis.

Once the adjusted diagnosis data were input to the back-calculation model, we used an EM-type algorithm to obtain estimates of HIV incidence in a specified set of time periods and estimates of rates of HIV testing in other specified time periods. For estimation of HIV incidence, one can specify an equivalent linear Poisson model where the input data are the diagnosis counts categorized by time period and stage of disease at diagnosis, the HIV incidences are the unknown parameters to be estimated, and the predictor variables are represented by functions of both the AIDS and HIV testing hazards. If both sets of these hazards are assumed to be known, the predictor variables in this linear Poisson model are then also known and the variability of the estimated HIV incidences can be obtained as part of the usual output for such models. However, since in this instance the HIV testing hazards are estimated rather than known, this provides an additional source of variability for the HIV incidence estimates.

Additional uncertainty results from such model choices as (a) choices related to the untreated AIDS hazards (e.g., whether values are estimated or specified, the type of structure imposed on either estimated or specified values), (b) choices related to the number and widths of time periods in which the numbers of infections are assumed constant, and (c) choices related to the structure of the HIV testing parameters (e.g., whether these parameters depend on calendar time, time since infection, or both; the number and widths of time periods in which testing hazards are constant, even if only calendar time dependence is assumed). The implications of these choices vary widely for different types of the estimated parameters. For example, the detailed structure of estimated HIV testing rates may be completely changed by the assumption of a more complex model if the additional parameters introduced are in fact relevant—one version, say, may assume calendar time dependence for HIV testing rates while another version introduces additional

parameters which allow the rates to also depend on time since infection. If the estimates of these additional parameters are highly significant then the testing rates in the two models likely will exhibit substantial differences. In another context, simply changing the number of calendar time intervals or the composition of these intervals will affect the estimates of HIV testing rates in particular years. That is, the estimated rate in 1999 may differ if 1999 is included in the interval 1997–1999 as opposed to the interval 1999–2001. Similar effects would also hold for incidence estimates for individual years. For the most part, the effects of the above types of choices on global estimates such as total estimated infections are limited. Changes in the AIDS hazards, however, can induce moderate changes in the estimated total infections, but such changes will also result in changes in the overall goodness of fit of the model (the comparison of expected and observed diagnoses by year and type), so this would imply some constraint on the amount or type of change that would be plausible to consider for the AIDS hazards.

9. Assumption on Constant Incidence (SSA/SEA)

We assume that incidence is constant in the most recent few years in order to obtain an estimate for the single year corresponding to the most recent year of diagnosis in the data. Successive incidence estimates based on BED results, therefore, may not accurately reflect the magnitude of incidence change, but they should reflect the direction of the change. Also, successive estimates based on BED results should be more sensitive than other available methods to recent changes in incidence.

10. Extrapolation (SSA/SEA)

Extrapolation of incidence from the 19 areas to the entire U.S. could introduce bias if the ratio of HIV incidence to the number of AIDS diagnoses is different between the 19 areas and other areas. Since HIV incidence is unknown we cannot check whether this assumption is true. What we can do is to check whether the ratio of HIV diagnoses to AIDS diagnoses is similar between the 19 areas and the 33 states with at least 5 years of HIV reporting. Results show that they are very close, with differences of less than 2%. To minimize bias, extrapolation is implemented within each group stratified by sex, race/ethnicity, age, and transmission category.