QUALITY ENHANCEMENT OF THE NATIONAL HOUSEHOLD SEROPREVALENCE SURVEY

James T. Massey[1], Ralph Folsom[2], and Trena Ezzati[1]
[1]National Center for Health Statistics and [2]Research Triangle Institute

INTRODUCTION

The National Household Seroprevalence Survey (NHSS), an area probability sample of 50,000 households in the United States, is being designed in order to obtain a direct estimate of the prevalence of human immunodeficiency virus (HIV) infection for the total United States civilian, noninstitutionalized adult population. A direct assessment of the rate of HIV infection in the general population is needed so that health officials can more accurately predict the nation's health care and financial needs for the acquired immunodeficiency syndrome (AIDS) and HIV epidemic.

The survey will be conducted in two phases with the feasibility phase consisting of tests of field procedures and survey methodologies for determining whether it is possible to conduct a nationwide survey to assess what portion of the population has the HIV virus or risks getting AIDS. The second phase will consist of the national survey based on a probability sample of households.

A major conern for the NHSS is the ability to produce a scientifically valid estimate of the prevalence of the HIV antibody that will be representative of the total U.S. civilian, noninstitutionalized adult population ages 18 to 54 years. Because of the concern for potential bias in the estimates produced from the survey, it became apparent during the early planning phases of the survey that a number of procedures would need to be integrated into the survey design to validate the survey results.

The purpose of this paper is to describe the major potential sources of error for the NHSS, to estimate the possible impact of these sources of error on the survey estimates, to describe the methodological procedures proposed to measure and adjust for the sources of error, to discuss the reactions and objections to the statistical procedures by NHSS advisors and community leaders, and to discuss the current status of the feasibility phase of the survey.

BACKGROUND

The NHSS is the first attempt to determine the prevalence of HIV infection in the adult U.S. civilian population. The survey will be based on an area probability sample of 50,000 households. One adult 18-54 years of age will be randomly selected in each sample household and asked to anonymously provide a sample of blood to be tested for antibody to the AIDS virus and to answer questions about sexual behavior and intravenous drug use in addition to standard demographic questions. Survey data collection teams will consist of an interviewer and a person specially trained to draw blood.

As part of phase one of the NHSS, a Pilot study was conducted in Allegheny County, Pennsylvania in January of this year with 85 percent of the eligible sample providing a blood sample and completing the risk behavior questionnaire. A $50 payment was provided to each respondent for their time and participation. On 19 of 22 items on the questionnaire, survey participants responded at a rate of 98 percent or higher. This included some questions on high-risk behavior. About 6 percent of survey participants reported at least one of the five risk behaviors addressed in the questionnaire. However, this does not represent a valid estimate of risk behavior in Allegheny county since no validation procedures were included to evaluate whether persons of differing HIV risk levels participated at comparable rates or whether respondents reported their risk factors accurately. Further, since the sample size was small, it was not possible to produce an estimate of the prevalence of HIV infection in Allegheny County.

A second feasibility study, a Pretest, of approximately 3,400 households in Dallas County, Texas is scheduled to begin in September of this year. The Dallas County Pretest will be the first attempt to determine the overall prevalence of HIV infection in the general population of the county. This Pretest along with the Pilot study results will help determine whether it is feasible to conduct a national survey of 50,000 households.

POTENTIAL SOURCES OF ERROR IN THE NHSS

The NHSS poses a number of methodologic issues that are in part due to the survey's primary objective of estimating HIV prevalence in the general household population. Due to the sensitivity of the topic of the NHSS and the potential for bias, various sources of measurement error have been identified.

The potential sources of error in the NHSS and their possible magnitude are presented in Table 1. The results are based on a sensitivity analysis conducted to evaluate the major sources of error in the estimates from the NHSS. A brief description and summary of these nonsampling errors are given below.

Coverage bias

The target population for the NHSS is the civilian, noninstitutionalized population of the United States. Coverage bias refers to the error in the survey estimates due to sampling frame and within household undercoverage of the population. The conclusions reached about coverage bias for the NHSS are summarized below:

- Although the NHSS will not include the homeless population, this population is small and, therefore it is assumed, will have a relatively small effect on the national prevalence estimate. (The infection rate for the homeless population is not known, although a recent study in New York City reported an extremely high rate (62 percent) of HIV infection.)
- The major concern with respect to coverage bias in the national study is the undercoverage of persons within households. Intravenous (IV) drug users, a group at high risk of HIV infection, are most likely to be missed in a household survey. The effect on the total prevalence estimate of coverage bias associated with the underreporting of IV drug users is expected to range between 2 and 10 percent.

This assumes that IV drug users are undercovered by 10 to 50 percent.

Nonreponse bias

Nonresponse bias refers to the error in the survey estimates due to the non-random, nonparticipation in the NHSS by selected sample persons. Since the major purpose of the NHSS is to produce an HIV prevalence rate for the total civilian adult population, there is concern whether the sample responding to the provision of blood will be representative of the entire target population or whether the sample will be biased by nonparticipation of a significant number of individuals in high risk subgroups of the population, thus yielding a serious underestimate of the true prevalence. The major findings related to potential nonresponse bias are given below:

- o Diferrential nonresponse related to the risk of HIV infection is the most serious methodological problem associated with estimating the prevalence of HIV infection for the total population. A high level of nonresponse among high risk persons will have the greatest impact on the nonresponse bias.
- o If persons at high risk or persons who already know their HIV status do not participate in the NHSS, the prevalence of the rate of HIV infection could be underestimated by more than 50 percent.
- o For the NHSS to be feasible, the nonresponse bias must be estimated and largely removed with appropriate statistical adjustments.

Misclassification bias

Misclassification bias refers to the response error in the survey estimates due to the misclassification of a sample person's risk status. The concern is that individuals may not accurately report their risk behavior in the self-administered questionnaire. If participation was 100 percent, misclassification bias would not affect the estimated prevalence for the total population, but would affect the prevalence estimates classified by risk. Using self-reported risk data to predict missing blood test results can introduce bias into the total population prevalence estimate. The findings related to misclassification bias are summarized below:

- o The relative bias in the prevalence estimate for high risk persons due to misclassification is expected to range from zero percent to 35 percent.
- o The relative bias in the prevalence estimated for low risk persons due to misclassification is expected to range from approximately 400 percent to 2800 percent. The 400 percent relative bias corresponds to a misclassification rate for high risk persons of 10 percent. Without a successful misclassification bias correction method, such a large potential error makes it virtually impossible to estimate the prevalence of HIV infection for the low risk population.
- o In the NHSS, the major purpose for collection of individual risk information will be to adjust the total estimate for nonresponse bias. The ability to use risk data alone to reduce the bias due to nonresponse depends on the difference in the high-risk denial rates and the true high risk proportions for respondents and nonrespondents. The risk data itself must be validated in order to adjust for the nonresponse. The larger the misclassification rate the greater the chance for a significant difference in the respondent and nonrespondent misclassification rates.
- o The greatest threat to adjustments based on risk data alone is the liklihood that the true proportion of high risk persons is substantially higher among nonrespondents. In this case, nonrespondents have more high risk deniers among those that self-report low risk. This increase in the proportion of high risk deniers among the low risk reporters will be exacerbated if the denial rate is higher for nonrespondents than for respondents. The appropriate prevalence rate for nonrespondents reporting low risk would therefore be substantially larger than that observed among respondents. Imputing the respondent low-risk prevalence rate to blood sample refusals who report low risk can seriously underestimate the global prevalence when the true high risk population has an elevated blood sample refusal rate and the risk denial rate is substantial.

Accuracy of HIV Test

The laboratory testing for HIV antibodies will first involve an enzyme-linked immunosorbent assay (ELISA). If the test is negative, the laboratory analysis result will be reported as negative (HIV negative). If the test is positive, two additional ELISA tests will be performed. If both are negative, the analysis results will be reported as negative. If one or both of the additional tests are positive, a Western blot assay will be performed and the result of that test will be reported as the final HIV test result. All tests will be performed using FDA-approved materials and procedures. The sensitivity and specificity of the laboratory testing protocol are summarized below:

- o The sensitivity of the laboratory protocol is assumed to be 100 percent.
- o The specificity of the laboratory protocol is assumed to be between 99.90 percent and 99.99 percent. The percent of false positives is expected to be less than 5 out of all positives. With 1 percent total population prevalence, this false positive contribution to bias is less than 0.1 percent. For the low risk population, even this miniscule level of false positives may not be negligible. Some projections of the low risk prevalence rate are as low as 2 per 10,000. This rate is only twice the false positive rate at the 99.99 percent specificity level.
- o The complete anonymity of the survey results will prevent respondents from receiving their test results. This avoids the problem of giving false positive results to respondents. The anonymous testing of blood is one of the key features of the survey designed to achieve a high response rate. For those survey participants who want to know their HIV status, they will be referred to a local agency for a test and appropriate counseling.

STATISTICAL PROCEDURES TO EVALUATE DATA QUALITY

Having identified potential sources of error and their potential impact on the survey estimates, a number of methods to evaluate the quality and adjust the results from the NHSS were examined. These

methods are summarized in Table 2. A few of the procedures are described in greater detail below.

Record Check Study (Direct Assessment and Statistical Verification)

Both forward and reverse record check studies were considered as potential methods to evaluate the quality of the results in the NHSS. In a forward record check study (direct assessment) a sample of persons with known characteristics of interest are surveyed. Self-reported survey responses are then compared with information contained in administrative records for the sample surveyed. Because of the sensitive nature of the NHSS, a forward record check study would be conducted so that the confidentiality of the record data would be preserved. This would be accomplished by anonymously incorporating addresses from the record sample into a larger randomly selected household sample and subsequently identifying the record cases using a demographic matching procedure keyed to date of birth. At no time would individual identifiers for the record cases ever be revealed to the NHSS contractor or the sponsoring agency. Since the addresses in the NHSS are destroyed before the matching would be conducted, neither the contractor, the government, nor field staff would ever know the addresses of the record cases. An encrypted city block variable would be required in addition to an accurately reported date of birth to insure a high probability of an exact match.

In a reverse record check study (statistical verification), the demographic characteristics from the NHSS respondents would be matched against the demographic charcteristics from the record source after the survey was completed by a third party. For the matched cases, the information to be validated, such as risk behavior, would be compared from the two data sources.

Some important conclusions reached related to the use of record check studies in the NHSS are given below:

- o A forward record check study is the best method for determining the feasibility of a national study, since it provides the only method of obtaining direct estimates of blood refusal rates and denial of high risk behavior among high risk persons.
- o A record check study can also be used to evaluate the nonresponse followup study and the item count questionnaire.
- o Record check data in combination with item count data can be used to adjust the survey estimates for potential bias associated with risk denial.
- o Because of the nature of the NHSS, a forward record check study is extremely sensitive due to the perception that it is unethical to select the record check sample from a file of known high risk persons without their informed consent. However, procedures are available for completely protecting the anonymity of all sample persons. Further discussion of this issue is presented in the next section.
- o The matching process will be difficult because of the complete anonymity of the NHSS.
- o A reverse record check study is very appealing because it is perceived to be less threatening to high risk individuals, however it is unlikely to produce enough matched high risk cases in the Pretest to accurately evaluate the feasibility of a national study.

Geographic Stratification

The NHSS study will be a multistage area probability sample. At the first two stages of selection the geographical units will be stratified using public information related to the known risks of HIV infection. Stratification is required in the NHSS to improve the sampling efficiency of the survey. The potential use of stratification to assess the quality of the Pretest results is summarized below:

- o Stratification can be used in the Pretest to provide a greater number of high risk persons. Without a record check study, it is absolutely essential to select a large enough sample of high risk persons in the Pretest to evaluate the survey procedures and to determine if survey participation varies across the risk strata.
- o Stratification can be used to complement the results from the record check data, since it will include a more representative sample of high risk persons.
- o Stratification cannot be used alone to validate the results of the survey. If effective stratification risk measures can be found, then one might be able to conjecture that the response rate appears to be related to risk behavior.

Nonresponse Followup Study

Another procedure to enhance and assess the quality of the NHSS results is a special followup study of a sample of nonrespondents. After data collection had been completed for the regular survey period, a sample of nonrespondents could be selected and special procedures used in an attempt to gain participation. The sample could be divided into two random subsamples with one-half being offered perhaps an enhanced incentive to provide a blood sample and complete the questionnaire. The other half-sample of nonrespondents would be offered a different incentive to complete the questionnaire only. The HIV status of this subsample could be estimated using a prediction equation developed from the other half-sample. This information would then provide the basis for estimating and adjusting for nonresponse bias. The success of a post-survey nonrespondent followup study would depend on obtaining a relatively high response rate for each subsample.

Testing for Hepatitis B

Although reliable data on the prevalence of HIV antibody and HIV risk behaviors are not available for the general population, an additional quality assessment method would be to test the blood samples for antibody to hepatitis B virus, in addition to HIV antibody. As a surrogate measure for certain high risk behaviors, hepatitis B results could be compared with national data on hepatitis B prevalence from the second National Health and Nutrition Examination Survey. This comparison of the two prevalence estimates could provide an indication of potential nonresponse bias in the NHSS estimates. Further, since the risk factors for HIV and hepatitis B infection are similar, the percentage of hepatitis B seropositive individuals who report no high risk behavior could be viewed as an index of HIV risk denial. This procedure is problematic since no gold standards are available for hepatitis and its exact correlation with HIV risk behavior is unknown.

Item Count

Since there is concern that some individuals will

deny certain high risk behaviors in the standard direct questions in the sample person questionnaire, a relatively new technique called "item count" has been proposed to gauge the level of underreporting or denial in the responses to the direct questions. This method, which is an indirect questioning technique, will allow survey respondents to provide statistical information collectively about sensitive behaviors without having to answer a direct question about their own personal behavior. Respondents provide only a count of items from a list that are true for them, but provide no indication of which individual items are true for them. The purpose of such an indirect questioning method is to enhance respondent's perception of the privacy of their answers, thereby minimizing the liklihood of denial of potentially stigmatizing behavior and its biasing effects. Item count questions consist of a list of behaviors the respondent may or may not have done. Respondents answer the questions by indicating the number of behaviors they have done, but not which particular ones. A split sample design is used whereby half of the respondents receive questions which include the target behavior in the list, while the other half is asked questions about the same list of non-target behaviors without the target behaviors.

REACTIONS TO STATISTICAL PROCEDURES BY NHSS ADVISORS AND COMMUNITY LEADERS

The design of the NHSS poses a significant challenge for survey methodologists not only because of the special methodologic issues but also because of the complex social attitudes and political issues surrounding AIDS and HIV infection. The task of developing a household-based HIV seroprevalence survey that is both feasible and capable of producing valid estimates of HIV infection has dictated the need for cooperation and involvement of the county health department and community advisory panels in the Pilot and Pretest study sites. Equally important has been the need to obtain expert advice from technical advisory groups on the survey procedures and methods to assess data quality.

Discussions with county health officials, community leaders, and technical experts have resulted in diverse reactions to the various proposed survey procedures. Some of the discussions have resulted in changes in survey procedures which were good. For example, in the Allegheny county Pilot study respondents were provided with an option of having their blood drawn by venipuncture or fingerstick. For those who had the fingerstick procedure, it often resulted in insufficient blood to perform the HIV antibody test and therefore the loss of valuable data. In the upcoming Pretest it was recommended that venipuncture only be offered and that the fingerstick procedure be provided only as an alternative and therby minimize the loss of data. Other suggestions regarding stratification and clustering did not significantly alter the survey procedures. Some of the modifications have somewhat altered the effectiveness of the survey. For example, the need for total anonymity means that no prospective survey can be conducted to measure incidence over time. Of a more serious nature, some modifications will severely impair the ability to validate the survey results. The most controversial reactions have centered around record check studies. As a result, direct assessment (forward record check study) will not be done in the pretest. Other modifications, such as limiting the number of recontacts to only one in the special nonrespondent followup study may adversely affect the overall response rate. A summary of the major reactions and objections is presented in table 3.

Most of these issues reflect the unique aspect of this survey, namely, estimating HIV prevalence in a household-based population; the sensitivity of the survey topic; the low prevalence of the attribute of interest (HIV infection); the disproportionate prevalence among subgroups of the population; concerns about privacy and anonymity; and the potential for nonresponse and response bias.

DISCUSSION

During the feasibility phase of this survey, it has been apparent that there are serious difficulties to be overcome in the conduct of the necessary studies to reliably assess the viability of a household survey to measure the prevalence of HIV infection. Although the Pretest plans include some quality assessment methods, it is not clear whether it will be possible to produce a prevalence estimate for Dallas county. Additional research needs to be done to determine how to best assess the degree of nonresponse bias and adjust the estimates for this bias.

Alternative procedures may need to be developed in order to determine the feasibility of a national survey. One alternative would be a comparison of NHSS results with CDC's family of survey results, for example within select PSU's. Research into the feasiblity of doing this will need to be done first. Due to ethical reasons, the incorporation of a forward record check study into any future feasiblity studies or in selected PSU's in a national study will require that the medical record source obtain informed consent from its patients before the addresses of known high risk persons could be included in any future NHSS studies.

Even though at the end of this year two feasiblity studies will have been conducted, it may still be necessary to field another feasibility study which would include all of the validation methods including direct assessment. Clearly, a great deal of effort remains to be done in order to establish the proper climate among community leaders and scientists that will permit the necessary methodologicial research to be done. The most efficient and accurate validation techniques need to be fully understood and accepted so that a valid estimate of HIV prevalence can be produced and subsequently used by health planners in developing effective treatment and prevention strategies and planning for the health care needs for those affected by HIV infection.

REFERENCES

Horwitz, D.G., Weeks, M.F., Hurley, P.F., and Wright, R.A. "The National Seroprevalence Survey," Proceedings of the Social Statistics Section of the American Statistical Association, pp 1-5, 1988.

Wright, Robert A. "Results of the NHSS Feasibility Study," to be published in the Proceedings of the Social Statistics Section of the American Statistical Association, 1989.

TABLE 1. POTENTIAL SOURCES OF ERROR IN THE NHSS

Type of Error	Cause of Error	Potential Magnitude
Sampling	Selection of sample from population.	±100,000 to 150,000 (confidence interval)
Coverage bias	Non-random coverage and enumeration of population associated with risk to the HIV infection.	2 to 10 percent
Nonresponse bias	Non-random non-participation by sample persons associated with risk to the HIV infection.	-65 to 25 percent
Misclassification bias[1]	Error in survey estimates due to the misclassification of sample persons risk status.	-35 to 35 percent for $\hat{P}_T$ and $\hat{P}_H$ 400 to 2800 percent for $\hat{P}_L$
Accuracy of HIV test results	False positive and false negative results.	0 to 5 percent

[1] $\hat{P}_T$, $\hat{P}_H$, and $\hat{P}_L$ represent the estimated prevalence of HIV for the total, high risk, and low risk population.

TABLE 2. STATISTICAL PROCEDURES TO ENHANCE AND ASSESS THE QUALITY OF NHSS RESULTS

Procedure	Purpose	Strenths and Weaknesses
Geographic stratification by risk level	To improve sampling efficiency. To estimate response rate differences between risk defined geographic strata. To increase number of high risk persons in study.	Will improve efficiency and increase number of high risk persons. Response rate differences between strata may not be correlated with risk status due to non-risk related confounders. Only partially accepted by advisors. Stratification may be relatively ineffective.
Nonresponse follow-up	To collect blood or risk information about nonrespondents. To assess presence of nonresponse bias.	Can be used to test alternative strategies for higher response. Non-interviewed persons will participate. Relative scarcity of the HIV high risk population makes it impossible to completely assess nonresponse bias.
Consistency of HIV status and risk behavior	To assess validity of risk information.	Good measure of response validity for respondents who are HIV positive. Cannot measure response validity for respondents who are HIV negative.
Indirect risk behavior questions a) Item count questions b) Global risk questions	To assess denial of high risk behavior.	Questions are less threatening to respondents. Should provide better total estimate of high risk persons. Individual risk status is unknown. Methods have not been validated.
Direct assessment (anonymous forward record check study)	To estimate response and coverage of persons with known HIV or risk status. To evaluate validity of reporting of risk behavior by high risk persons.	Best method to evaluate validity of results and feasiblity of a national study. Will significantly increase number of high risk persons in pretest. Quality of data from record sources is unknown. Results may not be generalized to all high risk persons. Not accepted by advisors.
Statistical verification (reverse record check study)	To evaluate the validity of reporting of risk behavior by high risk persons. To estimate response rate of selected sample persons with known risk behavior.	No pre-matching or seeding of cases required. Accepted by advisors. Number of matches in pretest is too small to be useful. Can't assess quality of record source data or screening response rate.
Testing for hepatitis B	To estimate likelihood of nonresponse bias using surrogate measure for risk behavior.	Risk factors for HIV and hepatitis B are similar. Correlation between HIV and hepatitis B is not known precisely.
Comparisons to other data sources a) NHIS response rates across strata b) Family of surveys (CDC)	To estimate face validity of results and likelihood of a nonresponse bias.	Should detect major failure of survey. Validity of other data is unknown. Differences in definitions make comparisons difficult.
Interviewer observations and reasons for refusals	To adjust for nonresponse bias using logistic regression.	Unlikely to be accurate predictor of HIV status.

TABLE 3. REACTIONS AND OBJECTIONS TO STATISTICAL PROCEDURES BY NHSS ADVISORS AND COMMUNITY LEADERS

Procedure	Reaction/Objection	Solution
Geographic stratification	Targets neighborhoods and subgroups of the population (red lining).	Exclusive use of published census and public health data.
Oversampling of subgroups	Discriminates against groups suspected of having HIV infection. Perception of non-random sample.	Equal probability of selection in Pilot Study. Proportioned representation of race and ethnic subgroup's within strata in Pretest.
Cluster sampling	Perception of non-random probability of selection within segments.	Noncompact segments.
Screening information from neighbors	Creates suspicion of possible AIDS household among neighbors.	No neighbor information to be collected.
Forward record check study	No informed consent was obtained for participation in NHSS. Perception that record source has list of high risk persons. Deception about how sample was selected. Unethical not to reveal all methods and procedures. Mistrust of blinding procedures.	Dropped from consideration.
Data collection under NCHS confidentiality law	Mistrust of government by high risk persons will reduce response rate. Not 100 percent fool proof--any possibility of a link to person I.D. is too high a risk. Requires counselling for HIV positives.	Dropped from consideration
Differential monetary incentive	Discriminatory. Must reveal all procedures to public. Possible revelation not worth risk. Unethical to pay some respondents more than others.	Major reason for dropping D.C. from pilot. Equal incentive for all respondents in main study. Separation of follow-up study from main study. Differential incentive must require differential level of effort by respondent.

TABLE 3. (continued)

Field and office procedures	All procedures must be carefully monitored for any possibility of a violation of anonymity.	Privacy Officer and Privacy Committee established. Written privacy procedures developed for every aspect of survey.
Recontact of refusals	Harassment. Violates perception of anonymity.	Special definitions for refusals. No recontact of fully informed refusals. Only one recontact allowed in Pretest.
Storing of blood for future tests	Unethical without informed consent.	Blood collected in Pilot Study destroyed. Informed consent in pretest.
Risk behavior questions	Questions too specific and sensitive. Questions not specific enough. Street terms too graphic. Clinical terms not understood.	Thorough review of all questions for Pilot and Pretest. Minimal use of street terms that might offend a significant number of respondents.