A REPORT OF THE FINDINGS OF THE NATIONAL HOUSEHOLD SEROPREVALENCE SURVEY FEASIBILITY STUDY

D. G. Horvitz, M. F. Weeks, W. Visscher, R. E. Folsom, Research Triangle Institute
P. L. Hurley, R. A. Wright, J. T. Massey, T. M. Ezzati, National Center for Health Statistics
D. G. Horvitz, P. O. Box 12194, Research Triangle Park, NC 27709

INTRODUCTION

The primary objective of a National Household Seroprevalence Survey (NHSS) would be to determine the prevalence of human immunodeficiency virus (HIV) infection among the noninstitutionalized U.S. civilian population between the ages of 18 and 54. The NHSS would ask a probability sample of persons residing in households to provide a blood sample to be tested for the presence of HIV antibodies and to complete a questionnaire about behaviors which put them at risk for HIV infection.

This paper reports on the findings of a study to assess the technical feasibility of carrying out a NHSS (Research Triangle Institute, 1990), particularly with respect to participation by persons selected into the sample, the quality of reported risk behavior data, and the quality of HIV prevalence estimates derived from the survey data.

CURRENT HIV PREVALENCE ESTIMATION

In the absence of a national population-based survey, data on HIV prevalence have been generated by a surveillance system developed by the Centers for Disease Control (CDC) (Dondero et al., 1988). The system includes a family of HIV seroprevalence surveys covering various segments of the general population. Routine HIV testing of blood donors, military recruits, and Job Corps entrants, as well as special surveys of college students and prisoners, are designed to assess levels of infection in these groups. Other surveys, involving some 39 metropolitan areas and 426 clinics and hospitals (as of September 1989) focus on infection rates in specific subgroups in selected geographic areas, including HIV testing of newborns in participating hospitals and patients in sexually transmitted disease (STD) clinics, drug treatment centers, women's health clinics, tuberculosis clinics, and sentinel hospitals.

The metropolitan areas, clinics, centers, hospitals and individual participants in the CDC family of surveys are not probability-based samples.

National HIV prevalence estimates have been developed using two methods. The first method, "extrapolation," estimates overall HIV prevalence by combining data from selected seroprevalence surveys on the rates of infection in specific HIV risk groups (e.g., homosexual men or intravenous drug users) with the estimated sizes of these groups nationally. The method is deficient in several respects. The HIV infection rates used to produce the estimates have unknown sampling biases, reliable current data for estimating the sizes of the risk groups are not available, and only a relatively small proportion (estimated to be less than 15 percent) of the total U.S. population is covered (HIV prevalence rates for the remainder of the population are not available).

The second method, "back calculation," uses data on cumulative AIDS cases at the end of each time period (month or quarter) and the latency distribution (i.e. the probability of AIDS manifestation in each time period following HIV infection) to estimate current HIV prevalence. This method also has its deficiencies. The main difficulty arises because information pertinent to defining the latency distribution with sufficient accuracy to produce precise national HIV prevalence estimates is lacking. In addition, the level of underreporting of AIDS cases is unknown.

RATIONALE FOR A NATIONAL SURVEY

A survey based on a national probability sample of U.S. households could produce direct and, potentially, more accurate estimates of HIV prevalence in the general population than are available at present (Turner et al.,1987; De Gruttola and Fineberg, 1989; Karon et al., 1988). Further, such a survey could provide reliable data on HIV prevalence for important subgroups of the population defined by demographics, by geography, by risk behavior, and by HIV/AIDS knowledge and attitudes. Comparisons of prevalence estimates across these subgroups would provide new and valuable epidemiologic data bearing upon the spread of HIV infection in the U.S. population--information critical to improving current predictions of the future course of the AIDS epidemic.

To date, information on the sizes of the population groups in the U.S. which engage in specific risk behaviors has been rather meager, to say the least. A NHSS could provide estimates of the number of persons in each HIV risk behavior category nationally, as well as by demographic subgroups, geographic areas, HIV status, and HIV/AIDS knowledge and attitudes. Knowledge of the patterns of risk behavior, their distribution in the U.S. population, and their relationship to HIV infection levels and to HIV/AIDS knowledge and attitudes is essential for better understanding of the epidemic, for modeling the epidemic, and for cost-effective allocation of resources to educational programs and other prevention efforts designed to reduce the numbers of persons engaging in HIV risk behaviors.

Currently, the CDC HIV surveillance system is not based on probability samples of the various hospitals, clinics and treatment centers that serve the population groups of interest. Hence, there is no valid statistical basis for inferring national HIV prevalence estimates from the CDC data. However, a national household-based survey, that included a significant number of the metropolitans areas in which the CDC family of surveys is currently measuring HIV prevalence among mothers of newborn babies and among patients treated at STD clinics, drug treatment centers and general hospitals, could provide data on the extent of sampling biases in the CDC HIV data. Thus, CDC could develop an empirically derived method of adjusting for sampling bias in its future national HIV prevalence estimates for those risk groups covered by the family of surveys data.

Finally, a national survey would provide a set of nationally representative blood samples linked to risk behavior questionnaires that could not only be tested for antibodies to HIV and hepatitis B (a sexually transmitted virus), but also could be stored and available for additional epidemiologic testing. Through densitometric analysis (or other similar analyses) of the HIV-positive blood samples, an estimate of HIV-related disease progression could be derived for the HIV-infected population.

NEED FOR A FEASIBILITY STUDY

To this point, the primary focus of the paper has been to demonstrate that a national household seroprevalence survey would generate a range and depth of data essential, first, to fuller understanding of the AIDS epidemic and, second, to more effective planning and allocation of the resources needed to bring it under control. However, a

NHSS poses issues and challenges that are usually not present in health surveys.

In the United States, AIDS has been identified as primarily a disease of homosexuals and intravenous (IV) drug users, segments of the population that are particularly sensitive to the consequences of revealing their lifestyles for fear of subsequent discrimination by employers, landlords, insurance companies, and others. In the absence of hard data on the likelihood of participation in a household-based seroprevalence survey by gay men, IV drug users, and others engaging in behaviors that put them at risk of HIV infection, one must assume that estimates derived from such surveys have a clear potential for significant bias due to differential nonresponse between those at higher risk and those at lower risk of infection in the sample.

The public may also be reluctant to participate in an AIDS-related survey which requires not only a blood sample drawn in their home, but also completion of a risk-behavior questionnaire that emphasizes sexual behavior, a topic generally considered to be a very private matter. Hence, both high- and low-risk respondents might not answer the more sensitive questions honestly, most likely denying some high risk behaviors. Again, in the absence of hard data on the validity of responses to questions about those sexual behaviors that increase a person's risk for HIV infection, one must assume a clear potential for significant response biases in survey estimates based solely on the responses to these sensitive questions.

In the face of these issues, rather than undertake the NHSS immediately, the National Center for Health Statistics (NCHS), an agency of CDC, initiated a feasibility study to determine first hand whether obstacles to participation and valid reporting of risk behavior in a household-based HIV prevalence survey can be overcome.

DESIGN AND CONDUCT OF THE NHSS FEASIBILITY STUDY

Work began on the feasibility study on May 1, 1988. The initial primary objectives of the pilot study were to determine the impact on participation rates when blood samples and risk-behavior data are collected under strict anonymity procedures as compared to confidential procedures. In addition, the impact of a $50 incentive versus no monetary incentive was to be tested. All eligible respondents were to be shown a videotape featuring Surgeon General Koop and covering information about HIV testing, privacy procedures, test result reporting, the importance of honest answers to the risk-behavior questionnaire, and the importance of participation. All participants were also to have the option of obtaining their blood sample results, anonymously, through a counselor.

Washington, DC was selected as the initial site for the Pilot Study. Fieldwork, with an expected sample of 800 eligible persons, was scheduled to begin on July 30, 1988. In the last week of July, however, concerns expressed in the media by the DC Health Commissioner and, subsequently, other community leaders, created adverse publicity and the DC study was cancelled. Allegheny County (Pittsburgh), PA was then selected for a far less ambitious Pilot Study, focussing almost exclusively on gaining community acceptance for the survey and on assessing the feasibility of drawing blood samples in the homes of eligible sample persons.

The Allegheny County Pilot Study

The Pilot Study was conducted in cooperation with the Allegheny County Health Department (ACHD), which played a key role in media and community relations. In order to minimize the potential for repeating the adverse Washington, DC experience, ACHD established a Community Advisory Committee consisting of 21 local civic leaders and representatives of concerned groups. The Committee reviewed and endorsed all aspects of the study design, and suggested ways to encourage community acceptance and participation. Of particular importance, in view of the Washington, DC experience and the sensitivity of the survey, was a decision to use only procedures which would ensure respondents that their name and address could never be linked to their blood test result or to their risk-behavior questionnaire responses; that is, that the survey data set would be an anonymous data set. To add further insurance that these privacy protection procedures were, in fact, being carried out, an independent Privacy Committee was established to review and audit all RTI data handling protocols in the field and in the central office.

A geographically stratified, equal probability sample of 473 occupied residential units (with strata constructed to help ensure a sample reflecting the demographic distribution of households in Allegheny County) was selected for the Pilot Study. Survey teams, consisting of a field interviewer and a phlebotomist, visited each sample household in January, 1989. The interviewer determined, by interviewing a resident, whether any age-eligible (i.e. 18-54) persons were members of the household. If there were age-eligible residents, the interviewer selected one of these persons at random to be asked to participate in the study. The study was presented in detail to the sample person through the use of an explanatory letter, a special videotape presentation, and a consent form.

If the sample person agreed to participate, the phlebotomist collected a small blood sample, using either venipuncture or a fingerstick technique. The sample person was then asked to complete, in private, a self-administered 10-minute questionnaire. The questionnaire requested basic demographic information as well as information about factors related to risk of HIV infection (for example, IV drug use, male-to-male sex, number of sexual partners, sex with prostitutes, and sex with IV drug users). Upon completion, the questionnaire was sealed in a return envelope by the sample person for mailing to RTI. Persons who provided both a blood sample and a completed questionnaire were paid $50. Blood specimens were shipped to a laboratory via air express to be analyzed for antibodies to HIV.

After the fieldwork was completed, a separate follow-up study was conducted with nonrespondents in an attempt to better understand their reasons for not participating.

A total of 450 of the 473 sample households, or 95.1 percent, were screened for eligible persons. Of these 450 screened households, 308 contained an eligible sample person. Of the 308 sample persons, 263 participated in the Pilot Study by providing a blood sample and completing the sample person questionnaire, for a sample person response rate of 85.4 percent. The overall response rate was 81.2 percent, computed as the product of the screening rate (95.1) and the sample person response rate (85.4).

Participants responded to the individual questionnaire items at a rate of 98 percent or higher for all but three questions. Six percent of sample persons reported at least one of five selected HIV risk behaviors. However, the overall sample was too small to estimate HIV prevalence or the prevalence of risk behaviors in Allegheny County. Further, it was not possible to evaluate whether persons at different HIV risk levels participated at comparable rates or whether sample persons reported their risk factors accurately.

The Pilot Study was considered successful from both an operational and a sample person participation perspective. It contributed significantly to the development of procedures for outreach to the community and established the feasibility

of collecting blood samples to test for HIV infection and HIV risk behavior data in the homes of sample persons.

The Dallas County Pretest

Encouraged by the Pilot Study findings, a pretest was conducted in Dallas County, Texas, in the fall of 1989. It consisted of two components: a "main" survey, conducted between September 30 and November 19, and a Quality Assessment Study (QAS) involving recontacts with a sample of refusals from the main survey during the period November 29 through December 20.

The experience with the Pilot Study reinforced the need for community involvement in planning the Pretest. The Dallas County Health Department (DCHD) participated in every aspect of the planning process. In addition, the Dallas County Commissioner's Court appointed 29 individuals, representing a broad cross-section of community groups and interests, to serve on a Community Advisory Panel (CAP). The Panel reviewed each aspect of the Pretest design, recommended modifications that reflected community values, approved the final set of survey procedures, and assisted in encouraging community acceptance of the survey. Several issues of concern were raised by the CAP that required changes in the proposed Pretest design. The primary concerns of the Panel were in the areas of privacy protection, sample design, questionnaire terminology, and follow-up of persons initially classified as refusing to participate.

The design of the Pretest reflected several changes from the Pilot Study. These included:

- a larger sample
- construction of strata on the basis of expected HIV risk level
- oversampling in strata with higher expected prevalence of at-risk persons
- expansion of the questionnaire to include additional questions about HIV risk behaviors
- testing blood samples for hepatitis B virus infection
- a special follow-up study of a sample of nonrespondents

A geographically stratified probability sample of 2,141 occupied housing units was selected for the Pretest. The sampling frame was stratified by race/ethnicity, sex and marital status (using 1980 Census block-level data) and by expected level of HIV risk (using summary statistics, provided by the Dallas County Health Department, of HIV risk indicators by Census-defined geographic areas). Three risk strata (high, medium, and low) were constructed, with each one reflecting, at the insistence of the CAP, the same race/ethnicity distribution as in Dallas County as a whole. This requirement gave assurance that all race/ethnicity groups within each risk-level stratum would be sampled at exactly the same rate. The stratification was designed to assure representation of both high-risk and low-risk persons in the sample. By oversampling in the high- and medium-risk strata, the expected number of persons in the sample at high risk for HIV was greater than for a proportionately allocated sample of the same size.

The Pretest, like the Pilot Study, was conducted as a completely anonymous survey. The field procedures were much the same as used in the Pilot Study, except that additional precautions were taken in the field to give added assurance of privacy protection and special procedures were implemented in response to CAP objections to any refusal conversion efforts during the main survey.

The CAP objected initially to recontacting any sample person refusal, but subsequently agreed to a compromise in order to accommodate the special follow-up study of nonrespondents, or QAS. The compromise permitted sample person refusals to be classified as either "informed" or "uninformed" depending on whether or not the interviewer had read the Consent Form to the sample person prior to his/her refusal. Recontacts with either type of refusal were not permitted during the main Pretest survey data collection period. Only "uninformed" refusals were eligible for the QAS.

A total of 2,061 of the 2,141 sample households were successfully screened in the main survey for a screening response rate of 96 percent. The screening process identified 1,715 sample persons, of which 1,351 (79 percent) provided both a blood sample and a completed questionnaire. Of the 359 noninterviews, 301 were "uninformed" refusals and 29 were "informed". The remaining 29 noninterviews were eligible persons who were away for the duration of the survey or were incapacitated or had a language barrier.

The QAS was crucial to determining the extent to which lack of participation was related to HIV risk behavior. The objective of the QAS was to collect the survey data from enough main survey nonrespondents to provide a basis for estimating and adjusting for nonresponse bias in the survey estimates.

The Quality Assessment Study was conducted with a sample of 30 screening nonrespondents, who were offered a $25 incentive to provide the screening data, and with a sample of 175 persons who had refused to participate in the main survey. Half of the QAS sample persons (actually 87 refusals plus 2 screening nonrespondent eligibles) were asked to provide both a blood sample and a risk behavior questionnaire and were offered an incentive of $175 to do so; 23 complied for a response rate of 26 percent. Each person in the second half of the QAS sample (actually 88 refusals plus 7 screening nonrespondent eligibles) was asked only to complete the risk behavior questionnaire and was offered a $100 incentive; 52 QAS sample persons or 55 percent complied.

PRETEST FINDINGS

Findings pertinent to assessing the feasibility of a NHSS are reported in this section. Primary emphasis is given to information bearing upon the potential for bias due to nonresponse and bias due to invalid responses to the risk behavior questionnaire.

Participation Rates

The weighted response rates realized in the Pretest are estimates of the level of participation by sample households and sample persons that would have occurred if all households in Dallas County had been included in the survey. Pretest participation was similar to that of the Pilot Study, with an overall proportion of 82 percent willing to provide a blood sample to be tested for HIV antibodies, compared to 81 percent in Allegheny County. The screening response rate was higher in the Pretest, 98 versus 95 percent, but 85 percent of the sample persons in the Pilot Study were willing to provide a blood sample, compared to 84 percent in the Pretest. Of particular interest is the high Pretest questionnaire overall response rate of 88 percent, since it indicates that relevant information can still be obtained from many respondents who refuse to provide a blood sample.

Exhibits 1 and 2 provide data on participation rates for various subgroups defined by design variables and by demographic variables available from the screening interview. The participation rates are uniformly high for all of these subgroups. The overall rates in the high risk stratum were slightly lower for blood samples, 79 percent versus 82 percent in the other strata, and for the questionnaire, 79 percent versus 87 and 89 in the medium and low risk strata.

The race/ethnicity substrata overall rates were generally higher than the blood and questionnaire averages realized for the county as a whole except for the "high percent never-married white male" substratum in which 77 percent provided a blood sample and 79 percent a questionnaire.

The weighted response rates shown for demographic subgroups in Exhibit 2 are sample person rates. They do not include the two percent screening nonresponse. Again, these rates tend to be uniformly high for all subgroups, with higher than average reluctance to provide a blood sample or to complete the risk-behavior questionnaire among older males, whether married or not.

The response rates achieved in the Pretest indicate that the household population will respond to an HIV seroprevalence survey at rates that equal or exceed those usually achieved in national health surveys when assured of anonymity of the data and when paid an incentive. Taken as a whole, the response rate data also suggest a somewhat greater reluctance to participate by those who belong to subgroups with higher proportions of persons at risk of HIV infection than is found in the general population.

Risk Behavior Data Quality

The risk-behavior questionnaire data were edited by computer to check for and insert codes for inconsistencies. Less than five percent of the questionnaires required edit adjustments for inconsistent data. Just under seven percent of the questionnaires had more than two blank or unusable items. Another 11 percent of the questionnaires had one or two such items and 82 percent had no item nonresponse whatsoever. Sample persons with unusually high rates of item nonresponse were more likely to be from the high-risk stratum, to be older (35-54), and Hispanic.

To facilitate subsequent estimation of HIV status from risk behavior reports, it was necessary to impute questionnaire responses for items left blank by sample persons. At least one missing item was imputed for 118 or 8.2 percent of the 1,446 usable questionnaires. A key risk behavior (IV drug use, male-to-male sex, receptive anal intercourse) was imputed in 15 instances among 71 cases in which there was no response to the item.

Exhibit 3 shows Pretest estimates of the proportion of males in Dallas County who have engaged in male-to-male sex since 1978 and the proportion engaging in receptive anal intercourse in the past 12 months by design variables, by demographic characteristics and by QAS status. Overall, an estimated 7.6 percent of males 18-54 in Dallas County engaged in male-to-male sex since 1978 and 2.0 percent in receptive anal intercourse in the 12 months prior to interview. Both of these high-risk behaviors for HIV infection vary significantly across population subgroups defined by the risk strata and by the race/ethnicity substrata. For example, the male-to-male sex rates since 1978 range from 6.3 percent in the low-risk stratum to 17.7 percent in the high-risk stratum and from 4.5 percent in the "high percent black" substratum to 19.2 percent in the "high percent Hispanic" substratum.

There is also wide variation in both these behaviors across subgroups of the population defined by age, race, marital status, and QAS status. For example, the estimated proportion of males in Dallas County engaging in receptive anal intercourse in the past 12 months ranges from 0.3 of one percent among blacks to 4.7 percent among all never-married males.

The overall coverage of males engaging in high-risk behaviors was dramatically improved by the QAS. For example, the main survey estimated 5.1 percent of the males to have engaged in male-to-male sex since 1978, whereas 16.8 percent of males eligible for the QAS are estimated to have engaged in this behavior. The proportion of males reporting IV drug use since 1978 was significantly higher in the QAS than in the main survey, 12.2 percent compared to 2.8 percent (see Exhibit 4). Although the estimated proportion of males engaging in receptive anal intercourse is greater for the QAS, 2.8 percent versus 1.8 percent for main survey male respondents, this difference is not statistically significant.

Since the estimated proportions of persons reporting IV drug use since 1978 and in the past 12 months did not vary significantly across subgroups defined by design variables, the data are not shown. However, IV drug use did vary significantly across subgroups defined by demographic variables and by QAS status. These data are shown in Exhibit 4. An estimated 3.8 percent of persons 18-54 in Dallas County used IV drugs since 1978 and 1.3 percent in the past 12 months. Significantly more IV drug use, both since 1978 and in the past 12 months, was reported by sample persons in the 25-34 age group and by divorced and separated individuals. More males used IV drugs than females in the past 12 months, 2.2 percent versus 1.2 percent.

The data on HIV status reveals significant relationships with self-reported risk behaviors. Of the 15 sample persons who were HIV positive, 14 reported at least one behavior known to increase the risk of HIV infection. Of the 13 HIV-positive males, ten reported male-to-male sex since 1978, eight had 10 or more male sexual partners since 1978, and ten had engaged in receptive anal intercourse since 1978. Similar relation-ships are observed between reported risk behaviors and the prevalence of hepatitis B infection, a disease which is transmitted by many of the same behaviors that lead to HIV infection.

In the absence of a record check study, the actual level of response bias in the Pretest risk behavior estimates remains unknown. However, the Pretest risk-behavior questionnaire data clearly suggest that concerns over whether respondents in a household-based survey of HIV prevalence would deny sensitive behaviors that increase the risk of HIV infection or, otherwise, would tend to not report them accurately, do not appear to be very well founded. Significant numbers of sample persons reported risk behaviors in the main survey and, even more so, in the QAS. Item nonresponse rates for questions on risk behaviors were generally low. The patterns of reported risk behaviors across survey design variables and demographic variables were consistent with expectations. Further, the data are consistent with the expected relationships between risk behavior and HIV status.

The Pretest risk behavior estimates are subject to nonresponse bias, as well as response bias. Despite excellent participation rates and adjustments for nonresponse, the level of nonresponse bias in the Pretest risk behavior estimates remains unknown.

HIV Prevalence Estimation

Adjustment for Blood Sample Nonresponse - Prior to computing HIV prevalence estimates for the 18-54 year old population of Dallas County, an empirically based procedure was used to adjust for the 16 percent blood sample nonresponse. First, weight adjustments were carried out to compensate for the two percent screening nonresponse and the ten percent questionnaire nonresponse. These adjustments accounted for observed differences in response rates across (i) the sample design variables, (ii) the demographic variables available on the screening form, (iii) a respondent attitude variable observed and recorded by the interviewer, and (iv) QAS eligibility status. This weight

adjusted data file produced the risk-behavior prevalence estimates reported above.

An imputation adjustment was then performed for sample persons without blood test results. This imputation adjustment took advantage of the QAS which provided risk-behavior data for persons without blood test results. The development and use of a logistic model to predict (impute) probabilities of HIV infection from risk-behavior information was possible because of the strong correlation between self-reports of risk behaviors and HIV infection status.

Separate logistic models were developed for males and females. These models initially included the following independent variables: sample design stratum and substratum; QAS status; age, race, sex and marital status; respondent knowledge and attitudes about the survey; and four risk behaviors including male sexual partners since 1978, IV drug use, receptive anal intercourse in past 12 months, and a positive diagnosis for gonorrhea. The dependent variable was a binary indicator for HIV infection status.

The models were fit using those sample persons who reported at least one of the risk factors also reported by the HIV-positive cases. For males, 86 of the 659 laboratory HIV test results were used to fit the model; for females, the model was based on 243 of 702 laboratory HIV results. The final model for males included a linear trend in the probability of HIV infection with increasing number of male sexual partners since 1978. Also males who reported no receptive anal intercourse since 1978 were less likely to be HIV positive. IV drug use since 1978 was a weaker predictor of HIV infection in males. For females, the only significant predictors of HIV infection were five or more sexual partners since 1978 or a diagnosis of gonorrhea.

A total of 15 of the 42 males who reported their risk behavior, but did not provide a blood sample, were imputed nonzero probabilities of HIV infection ranging from 0.005 to 0.248. Nine of the 42 females in the same category were imputed nonzero probabilities of HIV infection of 0.005. The remaining persons with no blood test results were imputed a zero probability of HIV infection.

HIV Prevalence Estimates - Exhibit 5 displays the adjusted HIV prevalence estimates for Dallas County as a whole and also by sample design strata and substrata and by demographics. The total population adjusted HIV prevalence of 0.42 percent translates into an estimated 4,045 infected persons 18 to 54 years of age in Dallas County. The 95 percent confidence limits for this estimate range from 2,186 to 7,464.

Estimated HIV prevalence decreases across the design strata as expected. The estimate for the high-risk stratum is significantly higher than the estimate for the low-risk stratum. Among the race/ethnicity substrata, HIV prevalence is significantly greater in the "high percent never-married white male" substratum than in either the "high percent Hispanic" or the "remainder" substrata. However, there were no significant differences between the racial groups defined by the questionnaire responses.

Patterns observed across the other demographic variables in Exhibit 5 are consistent with known patterns of HIV infection. HIV prevalence is significantly higher among males than among females. Also HIV prevalence is significantly higher among never-married individuals than among ever-married persons. The 25-34 year old group had the highest HIV prevalence among the three age groups.

HIV prevalence estimates for Dallas County by key risk behaviors are presented in Exhibits 6 and 7. Differences in HIV prevalence between those reporting and those not reporting IV drug use, either since 1978 or in the past 12 months, were not statistically significant. However, a clear increase in HIV prevalence with increasing frequency of receptive anal intercourse is seen for males for both time periods. An increase in HIV prevalence also occurs for males as the number of their male sexual partners increases, whether since 1978 or in the past 12 months. No relationship between HIV prevalence and the number of female partners reported by males is evident in this survey. Likewise, there is no evidence that the number of male partners reported by females affected their HIV prevalence.

The size of each subgroup (prevalence of the risk factor in Dallas County) is also shown for the time period since 1978 and, if applicable, for the past 12 months. The estimated prevalence of HIV infection for males reporting male-to-male sex and no intravenous drug use since 1978 is almost nine percent, much higher than the overall HIV prevalence for males of 0.73 percent. The HIV prevalence estimate is even higher (12 percent) for males with this risk factor in the past 12 months. Almost four percent of males reported at least one male and at least one female sexual partner since 1978. The estimated HIV prevalence for these bisexual men is also close to four percent, about five times greater than for all males 18-54 in Dallas County. Men who reported receptive anal intercourse since 1978 show a considerably elevated HIV prevalence (14 percent) compared to that for men overall.

Underestimates of HIV prevalence due to nonresponse bias will occur when persons at higher risk of HIV infection participate in seroprevalence surveys at a lower rate than do persons at lower risk of HIV infection. While the extent to which this differential participation occurs is unknown, it is clear that the statistical procedures used in the Pretest were effective in reducing the nonresponse bias in the HIV prevalence estimates. These procedures capitalized on the sample design, the risk-behavior data, the QAS and the strong relationship between risk behavior and HIV status among blood and questionnaire respondents. The QAS was particularly important. It increased the total number of respondents and it provided risk-behavior data of sufficient accuracy for effective adjustment for nonresponse bias in the HIV prevalence estimates. The resulting estimates, presented in Exhibits 5-7, follow expected patterns across the risk strata and other subgroups of the population defined by demographic variables and risk behaviors, further increasing the level of confidence in their accuracy.

FEASIBILITY STUDY CONCLUSIONS

Findings from the NHSS Feasibility Study support the general conclusion that informative and useful data about HIV infection, risk behaviors and their relationships can be obtained in a national household seroprevalence survey. Specific feasibility study conclusions are summarized in this section.

Outreach

- Public sensitivities and concerns about AIDS-related household surveys must be addressed by involving a panel of advisors in the design and conduct of a national survey. This policy advisory panel must include representatives of the constituencies, organizations, and groups concerned about the HIV epidemic, and technical experts as well.
- Public health officials in the areas selected for a national survey must be fully advised about the survey well in advance of its execution.
- Given the high level of public interest in AIDS, it is important that local and national media be fully informed about a national survey.

Privacy and Anonymity

- It is not feasible to conduct a NHSS without using procedures that fully protect the privacy of all participants and that assure that the data collected are anonymous.

Data Collection Procedures

- The use of an interviewer/phlebotomist team to collect the blood sample and questionnaire data is both feasible and effective in gaining the confidence of sample persons.
- It is feasible to collect blood samples by venipuncture in a household setting and to package and ship them to a lab for analysis with minimal loss of sample integrity.
- A $50 incentive payment to sample persons who provide a blood sample and a risk-behavior questionnaire is effective in encouraging and gaining their participation.
- It is not necessary to advise sample persons of the result of their blood test for HIV infection.

Survey Participation

- A high proportion of the population of interest is willing to provide a blood sample in their home for HIV testing, when assured of anonymity of the data and when paid an incentive.
- An even higher proportion of persons in the population of interest is willing to complete a risk-behavior questionnaire, when assured of anonymity of the data and when paid an incentive.
- Significant numbers of persons who initially refuse to participate will, if recontacted and offered an increased incentive, provide both a blood sample and/or a risk questionnaire.

Risk-Behavior Data

- Significant numbers of people who engage in behaviors that put them at risk for HIV infection will report those behaviors in a self-administered anonymous questionnaire.
- Data on risk behaviors reported by sample persons can be used to estimate the sizes of risk groups.

Study Design

- The statistical efficiency of a NHSS can be increased very significantly through geographic stratification at the first and second stages, based on public health statistics and Census data related to risk of HIV infection, and by oversampling never-married males.
- Given the strong correlation between HIV status and specific risk behaviors, a double sampling, or two-phase, design can achieve estimates of HIV prevalence at a specified level of precision at lower cost than a single-phase design. In the double sampling design, a portion of the total sample would be asked to provide both a blood sample and the risk-behavior questionnaire and the remaining portion would be asked to complete only the questionnaire.
- Geographic stratification can also be used effectively to reduce (through nonresponse weight adjustments) the bias which arises because of lower likelihood of participation by sample persons at higher risk for HIV.
- Risk-behavior data collected in a follow-up study of survey nonrespondents can be used to effect significant reductions in the bias in the estimates of HIV prevalence that would normally result from nonresponse.

These conclusions, while comprehensive and supportive of the feasibility of a NHSS, may not reflect completely the feelings and attitudes of those asked to participiate in the Pilot Study and the Pretest. It is our opinion that many of the respondents were motivated to contribute in some way toward solving the AIDS problem. Others, motivated, perhaps, by opposition to the Pretest by the Dallas Gay Alliance, including public demonstrations, were very negative toward the survey. It is of interest to note, however, that six of the seven gay men on the Dallas Community Advisory Panel supported the Pretest and that significant numbers of gay men participated. On the other hand, the QAS data clearly suggest a greater reluctance to participate by persons at higher risk of HIV infection.

While the conclusions reflect the successful application of several techniques for reducing nonresponse bias in the estimates of HIV prevalence, the magnitude of any residual bias, due to differential rates of response by persons at higher and lower risk of HIV infection, remains unknown. Similarly, the level of any residual response bias, due to denial of risk behaviors, remains unknown.

A forward record check study, which can be used to directly assess the levels of response and nonresponse biases, was contemplated for the Dallas Pretest. However, it was not implemented since it required the release of address information from clinic records of persons at risk of HIV infection without their prior consent. This was not considered an acceptable procedure. Direct assessment, in a sample of PSU's, remains a candidate for measuring the quality of HIV relevant data in a NHSS, provided records are available for individuals whose behavior has put them at risk of HIV infection and who have previously signified their willingness to be subjects for future research studies.

Despite the obstacles to including a direct assessment of the levels of response and nonresponse bias in HIV prevalence estimates generated by NHSS data, we remain convinced that a NHSS would reduce the uncertainty associated with current national HIV prevalence estimates, as well as provide useful and valuable data essential to a fuller understanding of the AIDS epidemic and its future course.

References

De Gruttola, V. and Fineberg, H. V., Estimating prevalence of HIV infection: Considerations in the design and analysis of a national seroprevalence survey. Journal of Acquired Immune Deficiency Syndrome, 1989; 2:472-480.

Dondero, T. J., Pappaioanou M., and Curren, J. W., Monitoring the levels and trends of HIV infection: the Public Health Service's HIV surveillance program. Public Health Reports, 1988; 103:213-220.

Karon, J. M., Dondero, T. J., and Curran, J. W., The projected incidence of AIDS and estimated prevalence of HIV infection in the United States. Journal of Acquired Immune Deficiency Syndrome, 1988; 1:542-550.

Pappaioanou M., Dondero, D. J., Petersen, L. R., Onorato, I. M., Sanchez, C. D., and Curran, J. W., The family of HIV seroprevalence surveys: objectives, methods, and uses of sentinel surveillance for HIV in the United States. Public Health Reports, 1990; 105:113-118.

Research Triangle Institute, National Household Seroprevalence Survey Feasibility Study Final Report. Research Triangle Park, NC, 1990.

Turner, C. F., Fay, R. E., and Widdus, R., Monitoring the spread of HIV infection., National Research Council, National Academy of Sciences, Washington, DC, 1987.

Turner, C. F., Miller, H. G., and Moses, L. E., editors, AIDS:Sexual Behavior and Intravenous Drug Use. National Academy Press, Washington, DC, 1989.

Exhibit 1. Overall Response Rates by Design Variables
Dallas County Pretest

Design Variables	Blood Sample Percent	Questionnaire Percent
Total	82.04	87.61
Risk Strata		
High Risk	78.55	78.74
Medium Risk	82.10	86.69
Low Risk	82.44	88.94
Race/Ethnicity Substrata		
High % Hispanic	82.18	93.30
High % Black	89.14	97.33
High % Never-Married White Male	77.44	79.44
Remainder	82.27	87.79
High Risk Stratum		
High % Never-Married White Male	72.98	74.31

Exhibit 2. Estimated Response Rates* for Sample Persons Completing the Questionnaire and Giving a Blood Sample
Dallas County Pretest

Subpopulations	Questionnaire Number	Questionnaire Percent	Blood Sample Number	Blood Sample Percent
Total	1498	90	1493	84
Marital Status				
Married	587	90	596	80
Not Married	909	90	896	90
Age				
18 - 24	278	96	280	87
25 - 34	656	92	644	85
35 - 54	564	85	569	81
Sex				
Male	733	87	729	82
Female	764	92	763	86
Marital Status by Sex and Age				
Married Males	278	87	287	79
18 - 24	17	96	18	76
25 - 34	124	97	124	81
35 - 54	137	81	145	78
Married Females	308	93	308	82
18 - 24	39	100	40	73
25 - 34	125	91	123	84
35 - 54	144	93	145	82
Single Males	453	88	441	87
18 - 24	112	96	109	90
25 - 34	211	89	207	87
35 - 54	130	76	125	81
Single Females	456	91	455	92
18 - 24	110	95	113	96
25 - 34	195	91	190	91
35 - 54	151	86	152	90

* The response rates in this table do not include the 2 percent screening nonresponse.

Exhibit 3. Weighted Rates of Male-to-Male Sex Since 1978 and Receptive Anal Intercourse in the Past 12 Months by Characteristic
Dallas County Male Population 18 to 54

Characteristic	Since 1978 Male-to-Male Sex Percent	SE	Past 12 Months Receptive Anal Intercourse Percent	SE	Percent of Population
Total	7.65	1.91	2.00	0.55	100.00
Risk Strata					
High Risk	17.74	4.37	8.28	2.85	8.27
Medium Risk	8.14	2.06	3.80	1.46	21.50
Low Risk	6.32	2.60	0.71	0.51	70.22
Race/Ethnicity Substrata					
High % Hispanic	19.17	10.59	1.25	1.14	8.80
High % Black	4.47	2.28	0.00	0.00	6.88
High % Never-Married White Male	13.90	3.32	6.57	2.12	14.72
Remainder	5.19	1.94	1.33	0.65	69.60
Risk Strata by Race/Ethnicity					
High Risk					
High % Never-Married White Male	24.42	6.53	12.98	4.40	5.13
Medium Risk					
High % Never-Married White Male	8.27	2.76	3.15	1.67	9.59
Remainder	7.54	3.88	5.19	2.95	8.20
Age					
18 - 24	1.70	0.74	0.50	0.30	19.06
25 - 34	11.44	3.71	3.65	1.27	36.59
35 - 54	7.08	2.91	1.28	0.60	44.35
Race*					
Black	3.10	1.45	0.28	0.27	13.53
Hispanic	13.52	6.32	1.65	1.57	18.48
Other	6.96	2.07	2.44	0.69	67.99
Marital Status					
Married/Widowed	4.93	2.25	0.40	0.37	57.48
Divorced/Separated	4.01	1.82	2.86	1.59	12.57
Never Married	14.40	4.27	4.71	1.41	29.95
QAS Status by Marital Status					
Main Survey	5.09	0.88	1.79	0.55	78.10
Married/Widowed	3.62	1.19	0.53	0.48	43.78
Divorced/Separated	2.86	1.50	1.26	0.94	9.04
Never Married	8.42	2.03	4.17	1.49	25.27
QAS	16.80	7.93	2.75	1.38	21.90
Married/Widowed	9.12	8.27	0.00	0.00	13.69
Divorced/Separated	6.96	5.95	6.96	5.95	3.53
Never Married	46.71	18.08	7.61	5.12	4.68

* "Black" includes sample persons who reported their race as black or African American; "Hispanic" includes those who, regardless of race, indicated that they were of Spanish or Hispanic origin; and "other" includes those who reported race as white, Alaskan Native or Native American, Asian or Pacific Islander, or some other race.

Exhibit 4. Weighted Intravenous Drug Use Rates Since 1978 and in Past 12 Months by Demographics and QAS Status
Dallas County Population 18 to 54

Characteristic	Since 1978 IV Drug Use Percent	SE	Past 12 Months IV Drug Use Percent	SE	Percent of Population
Total	3.80	0.82	1.30	0.59	100.00
Age					
18 - 24	4.76	1.70	1.04	0.57	18.99
25 - 34	6.51	2.00	2.77	1.55	38.44
35 - 54	0.93	0.37	0.09	0.05	42.57
Sex					
Male	4.90	1.47	2.18	0.45	49.28
Female	2.74	0.72	1.16	0.20	50.72
Race*					
Black	2.87	0.74	1.38	0.51	16.49
Hispanic	2.28	1.27	0.29	0.29	18.24
Other	4.46	1.15	1.56	0.89	65.27
Marital Status					
Married/Widowed	1.90	0.76	0.17	0.09	58.75
Divorced/Separated	8.67	4.14	5.45	3.96	14.70
Never Married	5.32	1.31	1.51	0.60	26.55
QAS Status by Marital Status					
Main Survey	3.06	0.57	0.80	0.26	81.00
Male	2.83	0.82	1.09	0.47	38.49
Divorced/Separated	5.80	1.87	1.91	1.16	12.18
Never Married	4.68	1.27	1.43	0.61	22.20
QAS	6.96	3.51	3.43	2.82	19.00
Male	12.25	5.35	6.04	4.72	10.79
Divorced/Separated	22.51	18.87	22.51	18.87	2.53
Never Married	8.59	5.78	1.91	1.94	4.35

* "Black" includes sample persons who reported their race as black or African American; "Hispanic" includes those who, regardless of race, indicated that they were of Spanish or Hispanic origin; and "other" includes those who reported race as white, Alaskan Native or Native American, Asian or Pacific Islander, or some other race.

Exhibit 5. HIV Prevalence Estimates
by Design Variables and Demographics
Dallas County Population 18 to 54

Design Variables	HIV Infection Percent	SE	Sample Size
Total	0.42	0.13	1446
Risk Strata			
High Risk	2.15	0.87	401
Medium Risk	0.76	0.37	578
Low Risk	0.13	0.11	467
Race/Ethnicity Substrata			
High % Hispanic	0.25	0.26	118
High % Black	1.06	0.84	122
High % Never-Married White Male	1.53	0.53	538
Remainder	0.16	0.09	668
Marital Status			
Married/Widowed	0.01	0.01	600
Divorced/Separated	0.43	0.21	309
Never Married	1.35	0.45	537
Sex			
Male	0.73	0.23	701
Female	0.13	0.11	745
Age			
18-24	0.39	0.31	276
25-34	0.73	0.25	626
35-54	0.16	0.08	544
Race*			
Black	0.49	0.35	384
Hispanic	0.43	0.40	207
Other	0.40	0.13	855

* "Black" includes sample persons who reported their race as black or African American; Hispanic" includes those who, regardless of race, indicated that they were of Spanish or Hispanic origin; and "other" includes those who reported race as white, Alaskan Native or Native American, Asian or Pacific Islander, or some other race.

Exhibit 6. HIV Prevalence Estimates
by Key Risk Factors
Dallas County Population 18 to 54

Risk Factor	HIV Infection Percent	SE
Total	0.42	0.13
IV Drug Use, No Male-to-Male Sex		
Since 1978		
Drug Use Reported	0.99	0.56
No Drug Use Reported	0.40	0.13
In Past 12 Months		
Drug Use Reported	1.34	0.83
No Drug Use Reported	0.41	0.13
No IV Drug Use Male-to-Male Sex		
Since 1978	8.68	2.39
In Past 12 Months	11.77	2.54
Bisexual Male		
Since 1978	3.86	2.06
In Past 12 Months	0.00	0.00
Male-to-Male Sex and Receptive Anal Intercourse Frequency		
Since 1978	14.02	3.07
No male partner	0.10	0.05
Male partner(s) but no receptive anal intercourse	0.21	0.12
Male partner(s) and receptive anal intercourse sometimes	10.49	2.57
Male partner(s) and receptive anal intercourse usually or always	25.69	6.48
In Past 12 Months	17.04	5.65
No male partner	0.22	0.10
Male partner(s) but no receptive anal intercourse	7.41	2.63
Male partner(s) and receptive anal intercourse sometimes	12.92	6.04
Male partner(s) and receptive anal intercourse usually or always	26.58	8.31

Exhibit 7. HIV Prevalence Estimates by Number of Sexual Partners
Dallas County Population 18 to 54

Number of Sexual Partners	HIV Infection Percent	SE
Total	0.42	0.13
Since 1978		
Males		
No male partner	0.10	0.05
1-4 male partners	1.39	1.01
5-9 male partners	11.05	1.03
10-99 male partners	21.66	6.22
100+ male partners	41.56	30.14

Number of Sexual Partners	HIV Infection Percent	SE
Total	0.42	0.13
In Past 12 Months		
Males		
No male partner	0.22	0.10
1 male partner	4.27	4.07
2-4 male partners	12.38	1.76
5+ male partners	26.72	9.82