STATISTICAL ISSUES IN MEASURING THE PREVALENCE OF HIV INFECTION IN A HOUSEHOLD SURVEY

James T. Massey and Trena M. Ezzati, National Center for Health Statistics
Ralph Folsom, Research Triangle Institute
J.T. Massey, 6525 Belcrest Road, Hyattsville, MD 20782

INTRODUCTION

A major concern for a household HIV survey is the ability to produce a scientifically valid estimate of the prevalence of the human immunodeficiency virus (HIV) infection that will be representative of the population being studied. The major sources of error which might lead to an invalid HIV prevalence estimate include coverage bias, nonresponse bias, response bias, and measurement error bias. These sources of error and their potential effect on the results from a household HIV survey were previously discussed by Massey et al., 1989. Because of the concern for potential bias in the estimate produced from a household survey, it became apparent during the early planning phases of the pilot and pretest of the National Household Seroprevalence Survey (NHSS) that a number of procedures would need to be integrated into the survey design to validate the survey results.

In the Pretest of the NHSS in Dallas County, Texas, a number of procedures were incorporated to assess the quality of the results including the response and nonresponse biases. The key question related to response bias was whether respondents could and would accurately report risk behavior. For nonresponse bias, the key issue was whether persons at high risk to HIV infection would participate in the survey at the same rate as persons having less risk to HIV infection. This paper focuses on these two sources of error and describes the procedures in the Dallas Pretest to assess and adjust for these sources of error.

METHODS TO ASSESS DATA QUALITY

There were a number of methods proposed and evaluated for the Dallas Pretest to assess the quality of the Dallas results. These methods and procedures were also being evaluated for their utility in a national survey. Some of these methods are described briefly below.

Geographical stratification by risk

While the primary goals of stratification in the Dallas Pretest were to improve the sampling efficiency and to oversample persons at higher risk of HIV infection in order to better test field and questionnaire procedures, stratification by risk behavior was also used to assess data quality.

The sampling frame was stratified using demographic data from the Census files and summary public health statistics on HIV risk indicators by Census-defined geographic areas. Data from the 1980 Census block-level summary file were used as a starting point for stratification. Using the race/ethnicity, sex, and marital status data for persons 18 to 54 year of age, each segment in the frame was classified into one of four categories:
- high concentration black
- high concentration Hispanic
- high concentration never-married white males
- other (predominantly white).

The next step was to classify the sample segments into one of three HIV risk strata (high, medium, or low), using summary statistics on HIV risk indicators. The summary statistics used were as follows:
- AIDS morbidity rates, by Census tract
- syphilis morbidity rates, by Census block
- hepatitis B morbidity rates, by Census block
- relative number of males reporting sexual behaviors related to HIV risk (including rectal intercourse, sex with males, or a rectal gonorrhea culture), by Census block
- relative number of persons admitting to IV drug use, by Census block
- relative number of persons admitting to cocaine use, by Census block
- syphilis morbidity rates for males, by Census block
- a measure of global risk based on data on male sexual behaviors related to HIV risk and IV drug or cocaine use.

The first three items above were obtained from the Dallas County Health Department's surveillance programs for the disease cited. Other items were based on statistics provided by the Dallas County Health Department's Sexually Transmitted Disease Clinic. The approximate period covered by these statistics was 1988 and the first quarter of 1989.

Using the databases separately, each segment was classified as high, medium, or low, based on its ranking from high to low on either rate or relative number. For each risk indicator segment scores were allocated as follows: 10 for "high", 5 for "medium", and 0 for "low". The three risk strata were balanced on race/ethnicity such that their distributions by percent black, percent Hispanic, and percent white were approximately equal to the overall race/ethnicity percentage for Dallas County. The final high risk stratum contained about 9.7 percent of the age-eligible blacks, Hispanics, and whites in Dallas County (based on 1980 Census block-level counts). The final medium stratum contained about 20 percent of each race/ethnicity group and the

low stratum contained the remainder.

Assuming that the stratification by risk was reasonably effective, indications of biases in the Pretest results could be observed by comparing the Pretest results across risk strata. With effective stratification one would expect significant differences in HIV infection and reporting of risk behavior across strata. The potential for a significant nonresponse bias would be indicated by different response rates across risk strata.

Special nonresponse follow-up survey

A special Quality Assessment Study (QAS) among a sample of initial survey nonrespondents was conducted following the main survey to estimate the magnitude and adjust for the nonresponse bias. The QAS was randomly split into two treatment groups, A and B. Nonrespondents assigned to Group A were asked to provide a blood sample and complete a risk questionnaire in the QAS. Nonrespondents assigned to Group B were only asked to complete a questionnaire. The monetary incentives were increased in the QAS to estimate the maximum impact of a nonrespondent follow-up study. The total QAS sample included 30 final screening refusals and 175 sample person refusals. The split sample design was used to determine whether a greater percentage of the initial nonrespondents would complete the questionnaire without a blood sample in the QAS. The risk behavior data collected for Group B could be used to adjust for the nonresponse bias based upon the assumption that the relationships observed between risk behavior and HIV infection among blood and questionnaire respondents apply as well to questionnaire only respondents.

Correlates of risk

Several HIV risk and correlates of risk questions and an auxiliary blood sample test were included in the study to specifically help assess the quality of the results. In the questionnaire, information was collected about other sexually transmitted diseases, ever tested for HIV infection, and primary reasons for refusal. A serologic test was included to test for antibodies to the hepatitis B core antibody (anti-HBc). Hepatitis B virus is transmitted by behavior similar to that through which HIV is transmitted. All of these additions were made to compare with other known results and to study their correlation with HIV risk behaviors. Estimates of hepatitis B and sexually transmitted diseases such as gonorrhea, syphilis, and genital herpes can be estimated from independent sources. If these diseases were significantly underreported in the Pretest, response or nonresponse biases would most likely be present for HIV infection.

Comparisons to other prevalence estimates

Comparisons of the prevalence of HIV from the Dallas household survey were also made to estimates produced from backcalculation models. These models used the cumulative number of AIDS cases over time to estimate both the number of HIV infected persons giving rise to the AIDS cases and the current number of HIV infected persons. The two major sources of error associated with the backcalculation estimates are uncertainties about the incubation period between infection and onset of AIDS and the level of underreporting of AIDS cases.

RESULTS OF DATA QUALITY ASSESSMENT

Table 1 presents response rates, risk behavior, and infection rates by risk strata. The response rates in the high risk strata are slightly lower than the response rates in the medium and low risk strata. This result holds for the screening rate, the blood sample response rate, and the questionnaire sample response rate. It is not clear, however, whether nonresponse bias is associated with the lower response rates in the high risk strata. Horvitz et al., 1990 reported that the race/ethnicity substrata response rates were generally higher than the blood and questionnaire rate 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. Additional analysis was conducted to examine response rates within the separate risk strata and to compare the response patterns with response patterns in the National Health Interview Survey (NHIS).

Table 2 represents some of the additional analysis within risk strata and shows response rates by age, sex, and marital status within risk strata. Although the sample size within cells is often too small to draw definitive conclusions, some response patterns do emerge. Response rates for 35-54 year olds were generally lower than the response rates for other age groups. There were few differences by marital status and age. The lower response rates for older persons was observed across almost all of the domains. Response rate differences among demographic subdomains did not help explain differences between strata when the Pretest results were compared. In order to compare the Pretest rates to the NHIS response rates, the Dallas sample segments for the NHIS were classified using the Pretest risk strata definitions. The response rate for the NHIS households in the high risk strata had a slightly lower rate than the NHIS response rates in the medium and low risk strata. Thus, it appears that the response patterns observed in the Pretest reflect general response patterns in other health surveys. This result, however, does not preclude the possibility of a nonresponse bias.

Differences in risk behavior and HIV infection across strata were in the expected direction. Men reporting having male-to-male sex since 1978 increased from 6.3 percent in the low risk strata to 17.1 percent in the

high risk strata (Table 1). It is clear that stratification was effective in increasing the proportion of persons at risk within geographical areas and that a significant number of males did report risk behavior. The overall reporting of male-to-male sex of 7.7 percent is plausible. The reporting of high risk behaviors for HIV infection also varied by the race/ethnicity substrata (Horvitz et al., 1990).

The prevalence of HIV infection also increased substantially across strata from 0.1 percent in the low risk strata to 2.2 percent in the high risk strata. The overall infection level of 0.4 percent produces an interval estimate of 2200 to 7500 with a point estimate of 4000 HIV infected persons in Dallas County. This estimate is compared to estimates from backcalculation models in a later section.

Selected results from the Quality Assessment Study (nonrespondent follow-up survey) are shown in Table 3. The response rates for the QAS are shown in the paper by Horvitz et al. (1990). The "high percent never-married white male" substratum sample persons participated in the QAS at a rate slightly below the average for all persons in the QAS. Table 3 indicates the levels of reporting of risk behaviors in the main survey and the QAS. In the paper by Horvitz et al., significantly higher levels of reporting of male-to-male sex by marital status were shown (8 percent in the main survey compared to 47 percent in the QAS. Although the number of respondents in the QAS is small, there do appear to be higher levels of risk reported by persons responding in the QAS. This is the clearest indication that a greater proportion of high risk persons refused to participate in the main survey. The critical question still unanswered is whether the nonrespondents in the QAS had even higher levels of risk behavior and HIV infection than the QAS respondents.

The main correlates of HIV risk are shown in Table 1. Ever having gonorrhea was reported by 10.3 percent of respondents while 7.3 percent of the respondents tested positive for hepatitis B. Both of these estimates provide supporting evidence that high risk persons did participate in the Pretest. Both of these estimates are in the expected range based on external information. Table 4 compares the hepatitis B virus prevalence estimates from the Pretest with the prevalence estimates from the National Health and Nutrition Examination Survey II conducted by the National Center for Health Statistics (NCHS) from 1976 to 1980. While time and location differences make this comparison suspect, the slightly higher prevalence estimates in the Pretest indicate that the estimates have a reasonable face validity.

Another analysis that was conducted was an examination of the primary reasons for refusal by persons declining to participate in the main survey and the QAS. It was hoped that the reasons for refusal would provide

some insight into the relationship between nonparticipation and risk behavior or HIV status. Table 5 shows the distribution of major reasons for refusal. The two primary reasons for refusal, fear of giving blood and lack of interest, correspond to reasons given for refusal in other surveys conducted by the NCHS. These two reasons accounted for almost 75 percent of all refusals in both the main survey and QAS. The purpose of the survey was given as a reason for refusal more often in the QAS than in the main survey (17.3 percent versus 8.6 percent). Further analysis of the reasons for refusal have been previously presented (Ezzati et al., 1990). It was shown that a slightly higher percent of refusals gave purpose of the survey as the primary reason for refusal in the high risk strata. It was also shown that the QAS response rates were highest among the initial nonrespondents who had a fear of giving blood and or a lack of interest. This is consistent with many persons only being asked to complete the questionnaire in the QAS. Since information on reasons for refusal are somewhat subjective, it is difficult to draw any definitive conclusions from the analysis of the reasons for refusals.

To compare the backcalculation estimates with the Pretest HIV prevalence estimate, three separate backcalculation models were used to estimate a range for the number of HIV-infected persons in Dallas County in 1989. The first two sets of estimates were provided by the CDC AIDS program and the third by the National Cancer Institute (NCI). The personal contributions, in the preparation of these estimates, of Drs. Meade Morgan and Robert Byers at CDC and Drs. Robert Bigger and Philip Rosenberg at NCI are acknowledged.

The estimates from CDC are based on quarterly AIDS incidence from October 1981 through March 1989, adjusted for reporting delay. Two incubation period models were used, a Weibull model distribution and a Markov distribution. Although statistical criteria suggest that the Markov model fits the AIDS incidence data better than the Weibull, experience with a San Francisco City Clinic cohort suggests that the Weibull model provides a better estimate of cumulative HIV incidence. The NCI used a Weibull incubation distribution and backcalculations to estimate HIV prevalence as of January 1, 1985, and then extended this estimate to January 1, 1990, by including information about AIDS mortality and plausible rates of recent HIV seroconversions.

Table 6 compares the CDC and NCI backcalculation estimates of HIV prevalence in Dallas County with the Pretest estimate. As indicated, the CDC and NCI estimates have been inflated to reflect an assumed level of underreporting of AIDS cases. CDC suspects that as many as 25 percent of the Nation's AIDS cases are not reported, but underreporting is thought to be lower among persons engaging in male-to-male sex. Since

over 90 percent of Dallas County's AIDS cases are associated with male-to-male sex, Table 6 presents CDC estimated based on assumed underreporting levels of both 10 percent and 25 percent. The NCI estimates are available only for the 10 percent under-reporting assumption.

As indicated in Table 6, the upper bound of the Pretest HIV prevalence estimate falls within the range of that estimated by the CDC Markov and NCI backcalculation models, but it is somewhat lower than the CDC Weibull model estimated range. Since the validity of the various assumptions upon which the backcalculation estimates are based is unknown, it is not possible to determine whether the Pretest HIV prevalence estimate is low or whether the backcalculation estimates are high. However, the information collected in the Pretest gives additional information to Dallas County regarding risk behaviors and estimates of the number of persons infected with HIV by risk group which may be useful for health planning purposes.

NONRESPONSE ADJUSTMENTS AND DEVELOPMENT OF PREVALENCE ESTIMATES

This section describes and compares two methods of adjusting for nonresponse bias that were done prior to computing HIV prevalence estimates--a minimal adjustment procedure and a more extensive adjustment procedure. The results presented suggest that the more extensive adjustment procedures improve the accuracy of the estimates of HIV and hepatitis B prevalence.

Minimally adjusted prevalence estimates

The minimal adjustment procedure involved the calculation of HIV and hepatitis B prevalence estimates after two weight adjustments. The initial weight adjustment compensated for screening nonresponse across sample design strata and for QAS subsampling. The second weight adjustment compensated for missing blood test results. This second adjustment used data from two sets of sample persons for whom blood test results were available--those whose blood sample were tested for HIV antibodies and those that were tested for anti-HBc antibody. Blood test results were not imputed for persons who did not provide a blood sample or whose samples were of insufficient quantity for analysis. The minimal nonresponse adjustment procedure inflated the sampling weights associated with the available regular survey and QAS blood results by dividing these weights by the associated blood test availability rates.

Fully adjusted prevalence estimates

The more extensive adjustment procedures used two statistical techniques to adjust for nonresponse bias. First, weight adjustments were performed to compensate for screening and sample person questionnaire (SPQ) nonresponse. These weights were based on logistic response probability models that took into account observed differences in response rates across design variables, screener demographics, and the Contact Record attitude variable. The second nonresponse adjustment technique used in the calculation of the fully adjusted prevalence estimate was the imputation of HIV and hepatitis B results for sample persons without blood test results. These included individuals who did not provide a blood sample, as well as sample persons whose blood samples were not sufficient for laboratory analysis. This statistical technique capitalized on the unusual situation afforded by the SPQ-only option in the QAS, where risk behavior data were available for persons without blood test results. The imputation of missing blood results from SPQ risk information was possible because (1) there are few people at risk for HIV and hepatitis B infection, (2) there are few risk factors for these infections, and most important, (3) there is a strong relationship between self-reports of risk behaviors and HIV and hepatitis B infection status.

The blood results imputation involved imputing missing SPQ information and then using the SPQ data in logistic models to predict the probabilities of HIV and hepatitis B infection. First, entire SPQ records were imputed for three sample persons who had usable blood samples but missing SPQs. These imputations were done through a sequential hot deck matching procedure that gave each of these sample persons the same SPQ data as another sample person with the same demographic characteristics from the same design stratum. Next, missing SPQ responses to key demographic and risk behavior items were imputed using a combination of logical imputations and the hot deck procedures. Further documentation on the guidelines used for the logical imputations is provided in Appendix I of the feasibility study report (Research Triangle Institute, 1990). Finally, HIV and hepatitis B blood test results were imputed using infection probabilities obtained from the logistic regression models described below.

Four logistic models were developed to impute infection probabilities: (1) an HIV prediction model for males, (2) an HIV prediction model for females, (3) a hepatitis B prediction model for males, and (4) a hepatitis B prediction model for females. The dependent variables in these models were binary indicators for HIV and hepatitis B infection. The HIV imputation models were based on a subset of the laboratory HIV test results for sample persons who reported at least one of the HIV risk factors that was also reported by one of 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 hepatitis B imputation models were based on the entire set of laboratory hepatitis B results (642 for males and 684 for females).

The initial models included the main effects of the design variables and the SPQ

demographic and risk behavior data. Linear and quadratic polynomials were also included for the sexual partner variables. The final imputation models were obtained by sequential elimination of terms. Linear regressions were run to eliminate terms of less importance. Terms with a p-value of less than 0.10 were eliminated from the models except for key demographic and risk-related factors.

HIV blood results were imputed using the infection probabilities obtained from the final models for 42 males and 43 females who were missing HIV test results, but who had completed an SPQ. Of these, 15 males and nine females were imputed nonzero probabilities of HIV infection. The nonzero imputed male probabilities ranged from 0.005 to 0.248. The nonzero imputed infection probabilities for females had the common value of 0.005. Hepatitis B results were imputed for 59 males and 61 females who were missing hepatitis B test results; all of these individuals were imputed nonzero probabilities of hepatitis B infection. Persons with no HIV test results who reported none of the risk behaviors associated with the 15 HIV-positive sample persons were imputed a zero probability of HIV infection.

Table 7 displays the final logistic models for the prediction of HIV infection. For males, a linear trend in the probability of HIV infection was seen with an increasing number of male sexual partners since 1978. In addition, males who reported no receptive and intercourse since 1978 were less likely to be HIV positive. Intravenous 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 male sexual partners since 1978 or a diagnosis of gonorrhea.

As with HIV infection, a strong linear trend in hepatitis B prevalence was seen with increasing male-to-male sex since 1978. Intravenous drug use in the past 12 months was a very strong predictor for hepatitis B infection, in both males and females. For males, the logistic model predicted lower levels of hepatitis B infection for 18-24 year olds and higher levels for blacks. Number of female sexual partners and intravenous drug use since 1978 were not predictive of hepatitis B infection in males. Lower levels of hepatitis B infection were predicted for Hispanic women and for women in the 18-24 age group. In addition, receptive anal intercourse in the past 12 months and a doctor's diagnosis of gonorrhea were predictors of hepatitis B infection in females.

Comparison of minimally and fully adjusted prevalence estimates

Table 8 presents both the minimally and fully adjusted prevalence estimates of HIV and hepatitis B infection. The chi-square p-value given for each comparison indicates the statistical significance of a one-sided test of hypothesis that the fully adjusted estimates are higher than the minimally adjusted estimates.

In general, the fully adjusted rates are higher than the minimally adjusted rates. The HIV prevalence estimate increased from 0.25 percent to 0.42 percent as a result of the full adjustment procedure. The fully adjusted HIV prevalence estimate for males is significantly higher than the minimally adjusted estimate. The fully adjusted estimate of hepatitis B prevalence for males also exhibits a significant increase relative to the minimally adjusted estimate. These results suggest that the more extensive adjustment procedure improved the prevalence estimates by effectively compensating for more of the nonresponse bias than the minimal adjustment procedure.

As noted earlier, the correlation between reported risk behaviors and HIV and hepatitis B status, plus the higher level of risk reporting by QAS questionnaire respondents, permitted the prediction of infection probabilities that were used to develop the fully adjusted estimates.

DISCUSSION

A number of design features were incorporated into the Dallas Pretest to evaluate the quality of the results. The results of our evaluation are not conclusive. There are a number of both positive and negative indicators of data quality. The positive indicators presented in this paper and the paper by Horvitz et al. include:
- High interview and blood sample response rates across demographic domains and risk strata
- Low item nonresponse
- Reporting of one or more risk behaviors by all but one of the HIV positives
- High levels of reporting for major risk behaviors
- Higher levels of risk reporting in high risk strata
- High levels of reporting for major correlates of risk and high association between risk behaviors and correlates of risk. For example, 73 percent of hepatitis B positives reported 1 or more risk behaviors
- Ability of nonrespondent follow-up study (Quality Assessment Study) to reduce the nonresponse bias
- Effectiveness of risk strata to reduce sampling variability and nonresponse bias
- Reasons for refusal similar to reasons given for general health surveys
- Response distribution by strata similar to the National Health Interview Survey distribution by strata
- Sample person response rate higher for single males than for married males. On the other hand, the lowest exam response rate across the design strata was in the high risk, never married white male substrata
- High infection rates (over 25 percent)

observed for several of the highest risk groups of the population. The subgroups with the highest infection estimates were too small, however, to produce separate statistics from the survey.

The negative indicators of data quality include:

* Slightly lower response rates in high risk strata and in particular in the high risk never-married white male substratum.
* Lowest response rates among older males
* The HIV point estimate is lower than any of the backcalculation estimates based on reported AIDS cases
* Higher levels of risk reporting in the QAS. It might be assumed that there is an even larger percentage of high risk persons among the nonrespondents
* Very few persons reported being HIV positive in the self-reported questionnaire (0.1 percent)

It is clear from a review of the results that the household Pretest in Dallas was quite successful in obtaining cooperation, drawing blood samples, and collecting valuable information on risk behaviors. A national survey is operationally feasible. It is not as clear whether the HIV prevalence estimate for the Pretest is statistically valid. Although the sample design appears to be reasonably efficient, the small sample sizes, low prevalences being estimated, and lack of direct validation make it difficult to draw statistical conclusions. A record check study was a direct validation method considered but dropped because of the sensitivity of such a study and the absence of informed consent of clients from potential record sources. In a national survey, a record check study or some other method of direct validation for a subset of the sample would be extremely valuable in enhancing the credibility of the results.

Following the Dallas Pretest, a group of experts on survey methodology and public health statistics were asked to review the results of the Pretest and assess their quality. While there was not a complete consensus, the general conclusions reached by the majority of advisors are presented below.

* the Dallas household survey was successful
* the feasibility study provided sufficient evidence to support a national study
* the potential for nonresponse bias remains a primary concern and should be researched further
* the value of a national survey will increase over time as treatments affect the HIV incubation distribution
* information from probability samples are necessary for a scientifically defensible information system
* the benefits of a national survey outweigh the technical problems
* more emphasis should be given to local area estimates

REFERENCES

Ezzati, T., Visscher, W., Massey, J., and McQuillan, G. "Primary Reasons for Refusal in the Dallas County Household HIV Survey." Paper presented at the Sixth International Conference on AIDS, San Francisco, CA., 1990.

Horvitz, D, Weeks, M., Visscher, W., Hurley, P., and Wright, R. "A Report of the Findings of the National Household Seroprevalence Survey Feasibility Study." Invited paper presented at the 1990 Annual Meeting of the American Statistical Association.

Hull, H.F., Bettinger, Gallagher, M.M., Keller, N.M., Wilson, J., Mertz, G.J. "Comparison of HIV-Antibody Prevalence in Patients Consenting to and Declining HIV-Antibody Testing at an STD Clinic." JAMA 1988; 260:935-938.

Massey, J.T., Ezzati, T.M., and Folsom, R. "Quality Enhancement of the National Household Seroprevalence Survey." Proceedings of the 1989 Section on Survey Research Methods of the American Statistical Association.

Research Triangle Institute. "National Household Seroprevalence Survey Feasibility Study Final Report." April, 1990.

Table 1. Response Rates, Risk Behavior, and Infection Rates by Risk Strata
for Dallas County Household HIV Survey

	Total	High	Strata Medium	Low
Screening response	97.7	94.5	97.1	98.3
Blood sample response	82.0	78.6	82.1	82.4
Questionnaire response	87.6	78.7	86.7	88.9
Male-to-male sex since 1978	7.7(1.0)	17.1(4.4)	8.1(2.1)	6.3(2.6)
Receptive anal intercourse in past 12 months (among males)	2.0(0.6)	8.3(2.9)	3.8(1.5)	0.7(0.5)
Global risk	9.0(1.3)	15.7(2.7)	10.3(1.7)	7.9(1.7)
Gonorrhea*	10.3	16.5	9.8	9.7
Gonorrhea/syphilis/ genital herpes*	13.7	19.8	14.7	12.7
Ever tested for HIV*	14.9	25.9	22.7	11.3
Hepatitis B	7.3(1.2)	11.3(1.7)	8.5(1.1)	6.4(1.6)
HIV Infection	0.4(0.1)	2.2(0.9)	0.8(0.4)	0.1(0.1)

*Chi-square P=value <0.05. Standard errors are shown in parentheses.

Table 2. Regular Survey Weighted Sample Person Response Rates According to Marital Status, Sex, and Age by Risk Strata for Dallas County Household HIV Survey

Demographic Characteristic	Total Sample Size	Total Response Rate	High Sample Size	High Response Rate	Medium Sample Size	Medium Response Rate	Low Sample Size	Low Response Rate
Total	1715	80	489	78	680	80	546	81
Married Males								
18-24	18	88	5	100	4	100	9	85
25-34	138	83	31	73	59	85	48	84
35-54	171	73	35	75	38	75	98	73
Married Females								
18-24	42	84	9	68	21	95	12	80
25-34	149	79	28	68	54	63	67	83
35-54	172	78	25	83	43	72	104	79
Not Married Males								
18-24	121	87	39	83	47	88	35	87
25-34	239	79	102	73	105	82	32	79
35-54	154	72	54	76	69	78	31	68
Not Married Females								
18-24	125	90	37	88	64	84	24	94
25-34	216	85	74	75	97	81	45	89
35-54	170	81	50	90	79	77	41	81

Table 3. Comparison of Reporting of Selected Risk Behavior Between Main Survey and Quality Assessment Study for Dallas County Household HIV Survey

Risk Behavior	Main Survey Percent (S.E.)	Main Survey Percent of Population	Quality Assessment Study Percent (S.E.)	Quality Assessment Study Percent of Population
IV Drug Use Last 12 Months				
Total	0.80 (0.26)	81	3.42 (2.82)	19
18-24	0.69 (0.42)	17	3.42 (3.48)	2
25-34	1.59 (0.61)	31	7.97 (7.57)	7
35-54	0.11 (0.07)	33	0.00 (0.00)	10
Male-to-male Sex Since 1978				
Total	5.09 (0.88)	78	16.80 (7.93)	22
18-24	1.85 (0.84)	17	0.00 (0.00)	2
25-34	8.58 (2.10)	29	21.78 (14.85)	8
35-54	3.72 (1.20)	32	15.73 (9.48)	12
Receptive Anal Intercourse in Last 12 months				
Total	1.79 (0.55)	78	2.75 (1.38)	22
18-24	0.54 (0.33)	17	0.00 (0.00)	2
25-34	3.74 (1.45)	29	3.35 (2.14)	8
35-54	0.73 (0.34)	32	2.71 (2.04)	12

Table 4. Comparison of Hepatitis B Virus Prevalence Estimates in Dallas County Household HIV Survey and NHANES II By Demographic Characteristics

Characteristic	N	%*	(S.E.)	N	%*	(S.E.)
<u>Total</u>	1446	7.3	(1.2)	6122	5.3	(0.4)
<u>Age</u>						
18-24	276	1.5	(0.5)	1728	3.3	(0.5)
25-34	626	7.4	(1.3)	1900	5.3	(0.7)
35-54	544	9.8	(2.4)	2494	6.9	(0.6)
<u>Sex</u>						
Male	701	7.7	(1.1)	2867	6.6	(0.6)
Female	745	6.9	(1.7)	3255	4.0	(0.3)
<u>Marital Status</u>						
<u>Male</u>						
Never Married	297	6.4	(1.5)	847	6.6	(0.9)
Divorced/Separated	124	16.8	(5.2)	168	12.4	(2.3)
Married/Widowed	208	6.3	(1.4)	1838	6.0	(0.7)
<u>Female</u>						
Never Married	240	5.9	(1.5)	731	3.0	(0.7)
Divorced/Separated	185	5.1	(1.5)	388	5.4	(1.3)
Married/Widowed	320	7.8	(2.7)	2130	4.1	(0.4)

*Adjusted with sample person nonresponse weights

Table 5. Primary Reasons for Refusal by Survey Component for Dallas County Household HIV Survey

Reason for Refusal	Regular Survey (%)	QAS (%)
Total (n)	395	81
Purpose of survey	8.6	17.3
Fear of needles/giving blood/ religious beliefs/pregnant	18.2	0.0
Concerns about anonymity/ confidentiality	3.3	1.2
Don't have AIDS/survey topic not a concern	5.1	2.5
Never do surveys/not interested/ too busy	49.4	75.3
Spouse/relative/friend says "no"	6.8	1.2
Other	8.6	3.7

Table 6. Comparison of Dallas County Pretest HIV Prevalence Estimates with
Backcalculation Model Estimates*

Backcalculation Method	Assumed Percent AIDS Under-reporting	Estimated Range of No. of HIV Infected Persons
Dallas Pretest	--	2,200 - 7,500
CDC Markov Model	10	5,600 - 8,100
CDC Markov Model	25	7,000 - 9,600
CDC Weibull Model	10	9,700 - 13,600
CDC Weibull Model	25	12,100 - 16,000
NCI Model	10	7,100 - 13,800

*The project staff gratefully acknowledge the work of Drs. Meade Morgan and
Robert Byers of the Centers for Disease Control (CDC) and Drs. Philip Rosenberg
and Robert Biggar at the National Cancer Institute (NCI) in preparing these
estimates.

Table 7. Logistic Models for Predicting HIV Infection for Dallas
County Household HIV Survey

Gender	Independent Effects**	Beta Value	Beta SE	Students t-value	Significance Probability
MALES	Intercept	-2.0411	0.4769	4.280	0.0000
	SM78LN	0.9328	0.3822	2.441	0.0167
	AS78NO	-2.3103	0.7898	-2.925	0.0044
	IVDU78	1.6869	1.1159	1.512	0.1343
FEMALES	Intercept	-3.63759	1.0794	3.370	0.0008
	SM785P	-1.67562	1.0993	-1.524	0.1274

**SM78LN   Linear polynomial effect for five categories of male partners since 1978
  AS78NO   Indicator for no receptive anal intercourse since 1978
  IVDU78   Indicator for any intravenous drug use since 1978
  SM785P   Indicator for 5 or more male partners since 1978

Table 8. Comparison of Fully and Minimally Adjusted HIV and Hepatitis B
Prevalence Estimates for Dallas County Household HIV Survey

Prevalence	Minimally Adjusted Estimate Percent	SE	Fully Adjusted Estimate Percent	SE	Chi-Square P-Value
HIV Infection					
Total	0.25	0.09	0.42	0.13	0.013
Males	0.39	0.14	0.73	0.23	0.018
Females	0.12	0.11	0.13	0.11	0.160
Hepatitis B Infection					
Total	6.85	2.06	7.28	1.18	0.338
Males	5.57	1.20	7.66	1.13	0.007
Females	7.98	3.23	6.90	1.67	0.742