SAMPLE DESIGN & PROCEDURES

STUDY POPULATION

The study population for the 1996 National Sexual Health Survey (NSHS) is defined to include all English- or Spanish-speaking adults 18 years of age or older contacted and interviewed by telephone from June 28, 1995 to April 30, 1996. Eligible respondents must have resided in housing units, other than on military reservations, in the forty-eight contiguous states. This definition excludes persons living in Alaska or Hawaii. Interviews were conducted in English or Spanish languages by a total of 55 interviewers.

SAMPLE DESIGN

The 1996 NSHS is based on a standard Waksberg-Mitofsky probability sample selected from the Bellcore file of active area code-prefix combinations.

First Stage Sample Selection

A stratified first stage sample was selected, according to standard Waksberg-Mitofsky procedures. Using the April 1995 Bellcore file of active area code-prefix combinations, all 800 and 900 area codes (and others not in the geographic area or known not to be used for residential telephones) were eliminated from the file. To the remaining area code-prefix combinations pairs of digits from 00 to 99 were added to each area code-prefix combination. Thus, each area code-prefix combination produced 100 potential Primary Sampling Units (PSUs). This list was sorted by the four main Census regions (Northeast, South, Midwest, and West). Alaska and Hawaii were excluded from the sample. A target number of needed PSUs was determined for each region. It was estimated that a total of 2148 PSUs would be needed for the main study. The survey population for the survey was persons age 18 years of age and older. Within each region, a simple random sample of potential PSUs was selected. To each selected PSU a random pair of digits was added to produce a telephone number. The percentage of telephone numbers generated in this manner which turn out to be working residential numbers varies by region. These percentages were taken into account to select a sufficient number of potential PSUs to reach the target.

Screening Potential PSUs

Each of the generated telephone numbers was called to determine residential status. During a brief screener interview, it was verified that the correct telephone number was reached, that it was a private household and located in the expected geographic location. Up to 20 callbacks were made to each selected number, spread over different times and days of the week. 2146 PSUs were identified.

Sample Management

Each PSU contained 100 telephone numbers and was arranged in numerical order starting with the identified primary number. Each PSU had a unique identification number. Additionally, all telephone numbers were assigned a sequence number. The Waksberg-Mitofsky design requires that a fixed number of working residential numbers (WRNs) be identified in each PSU. This fixed number is called the cluster size. The cluster size is not the number of interviews in a PSU. A WRN includes all of the following call results: refusals, home recorders (that can be determined to be households), partial interviews, non-English speaking households, households where the interview cannot be conducted, households where interviews cannot be conducted due to a respondent disability, households that do not contain a person 18 or older, and interviews. In other words, every WRN counts toward the cluster size. Ideally, one does not want to exceed the specified cluster size. Hence, telephone numbers were released sequentially and those that were identified as not WRNs included: numbers not in service, numbers temporarily disconnected, business or organization numbers, government agencies, fax and computer lines, car phones.

Sampling began with releasing in each PSU a random set of telephone numbers equal to the cluster size. For this study, we estimated that a cluster size of 6 would be necessary. To release the first sample, we treated the primary number as a random start, released it and the next 5 numbers in sequence after it. Hence, in each of the 2146 PSUs, the first 6 numbers were released. As non-WRNs were identified, they were replaced sequentially.

SAMPLE DESIGN SPECIFICATIONS

The targeted completed interview sample size for the 1996 NSHS survey was n = 8,466 total cases.

Table A. Original Sample Design Specifications 1996 National Sexual Health Survey

Total Sample			26,513
(I) Completed Interviews (81, 82, 85, 87, 88, 89)			8,467
(P) Partial Interviews (70-74, 77-79, 93, 97)			230
(NC) Noncontacts			3,035
	(NC.E) Eligible (50-53, 60-63, 92, 95, 96)	931	
	(NC.U) Eligibility Unknown (40-43, 45-49, 94))	2,104	
(NA) Never Answered (21-24, 28, 29)			547
(R) Refusals (54, 57-59, 67-69)			1,427
(NE) Not Eligible			12,807
	(NE.NOT) Not Eligible : Not WRN	12,349	
	Discontinued (10)	6,694	
	Business / Government (11)	4,241	
	Modem / FAX (12, 26, 27)	997	
	Car / Cellular Telephone (13)	248	
	Not a Housing Unit (14)	169	
	(NE.WRN) Not Eligible : WRN	458	
	Not Regular Housing Unit (31)	69	
	Age Ineligible (33)	18	
	Duplicate Household (34)	2	
	Language Barrier (55, 65)	114	
	Not Interviewable (56, 66)	255	

Response Rates: Calculations and Formulas

$$\text{Response Rate } = \frac{I}{TotalSample - NE} = \frac{8{,}467}{26{,}513 - 12{,}807} = \frac{8{,}467}{13{,}706} = \mathbf{61.8\%}$$

S p1 = total eligibles / total WRN

$$= \frac{I + P + NC.E + R}{I + P + NC + R + NE.WRN} = \frac{8{,}467 + 230 + 931 + 1{,}427}{8{,}467 + 230 + 3{,}035 + 1{,}427 + 458} = \frac{11{,}055}{13{,}617} = \mathbf{0.812}$$

p2 = total eligibles/Total Sample = $\frac{11{,}055}{26{,}513}$ = **0.417**

$$\text{Adjusted Response Rate (cooperation rate) } = \frac{I}{I + P + NC.E + (p1)NC.U + (p2)NA + R}$$

$$= \frac{8{,}467}{8{,}467 + 230 + 931 + (0.812)2{,}104 + (0.417)547 + 1{,}427} = \frac{8{,}467}{12{,}991.547} = \mathbf{65.2\ \%}$$

Acceptance Rate = I / total eligibles = $\frac{8{,}467}{11{,}055}$ = **76.6 %**

SPECIAL NOTE: Survey Methods Group (SMG) was unable to retrieve the data for one completed interview (disposition code 85). Therefore, the provided data files will contain only 8,466 cases.

WEIGHTED ANALYSIS OF 1996 NSHS DATA

The Mitofsky-Waksberg probability sample design for the 1996 NSHS does not result in an equal probability sample of U.S. households. Any constructed weight(s) must be used to compensate for unequal probabilities of (1) selection; (2) multiple telephones in the household; (3) household nonresponse; (4) noncoverage of households without telephones. In addition, if the investigator wishes to make population inferences, the weight(s) must include a post-stratification ratio adjustment (Groves, et al, 1988). This type of adjustment is typically made in telephone surveys to bring survey estimates of subdomain totals in agreement with independent population figures for the subdomains. For the 1996 NSHS, the subdomains consisted of gender, age, education, and race/ethnicity. These subdomains were post-stratified to the 1990 U.S. Census and the 1995 Current Population Survey figures.

In analysis, the weight provided in the data set should be used to compensate for (1) through (4), above, in addition to post-stratification for making population inferences. The use of the design/post-stratification weight in analysis is strongly encouraged.

CONSTRUCTION OF ANALYSIS WEIGHTS

In general, there is one final weight that incorporates both sampling and post-stratification weights. The sampling (design) weights were constructed using three design weights. The post-stratification weights adjusted the 1996 NSHS distributions of four demographic characteristics to those corresponding to the 1990 Census and 1995 CPS data.

Description of weights construction

The first design weight adjusts for non-response within clusters. Clusters with only one completed interview were weighted up to two and those with greater than five interviews were weighted down to five. For clusters where two to five interviews were completed no weighting was done. Therefore, the weight is 5/6 when the completed cluster size is equal to six and 5/7 when the size is equal to seven. When the cluster size is equal to one, the weight will be two

The second design weight corrects for the greater probability of selecting persons in households with multiple telephone numbers. This weight is the inverse of the number of telephone numbers in the households with four or more telephone lines where the weight was kept at ¼. The weight for the person in household without information of telephone lines was one.

The third design weight adjust for the smaller chance of selecting persons in multi-adult households, since only one interview was completed in each selected household. This weight is equal to the number of adults in the household, except in households with ten or more adults where the weight was kept at ten. The weight for persons living in households without information on the number of adults was assigned the same weight as in two-adult households.

In addition to the three design weights, there are also four post-stratification weights. These weights are for gender, age, education, and race. We weighted these four demographics characteristics to adjust their distributions to those adult population (persons 18 years of age or older). The gender, age and education weights were based on 1995 CPS data[1]. Due to the different context of racial/ethnical options used in this study, the race distribution for this survey was not directly comparable to CPS data. Therefore, race distribution was weighted on estimates that were base on both information from 1990 Census and 1995 CPS data.

For user convenience, the final data has two weights. First, the three design weights were

[1] "Population Profile of the United States 1995", Current Population Reports, P23-189, Bureau of the Census, Economics and Statistics Administration, U.S. Department of Commerce

combined into one variable, DesgnWT1. Second, all the post-stratification weights were combined with design weights and saved as WT1, the overall weight. Both were saved in SPSS system file. If overall weight exceeded five, the weight was set to five. The weighted sample size was adjusted so that it equals the unweighted sample size. For user convenience, selected unweighted and weighted (using the combined weight) demographic variables are shown in Table 1, below.

Table 1. Demographic Characteristics of the National Sexual Health Survey.

	Unweighted		Weighted§	
	%	**(N)**	**%**	**(N)**
Gender				
Male	44.2	(3746)	47.0	(3980)
Female	55.8	(4720)	53.0	(4486)
Ethnicity				
White	77.3	(6528)	74.3	(6273)
African American	9.2	(774)	11.5	(971)
Hispanic	8.6	(725)	8.6	(727)
Other	5.0	(421)	5.6	(474)
Age				
18 to 29	22.5	(1898)	22.8	(1919)
30 to 39	25.4	(2139)	22.1	(1859)
40 to 49	21.6	(1818)	19.1	(1611)
50 to 59	12.4	(1042)	12.5	(1052)
60 and over	18.2	(1537)	23.6	(1988)
Education (years)				
< 12	13.1	(1107)	19.4	(1641)
12	30.3	(2567)	35.4	(2997)
> 12	56.5	(4780)	45.1	(3815)
Income ($1000)				
< 10	12.7	(1007)	14.4	(1131)
1 to 20	16.2	(1287)	17.5	(1377)
21 to 40	31.3	(2484)	31.1	(2438)
41 to 60	20.7	(1646)	20.0	(1568)
60 over	19.1	(1516)	17.0	(1336)
Marital status				
Married	54.0	(4566)	59.1	(5062)
Separated	2.3	(197)	1.8	(151)
Divorced	13.1	(1109)	9.3	(786)
Widowed	7.4	(628)	7.6	(641)
Never married	23.1	(1950)	21.5	(1815)

§ Post-stratified to the 1995 Current Population Survey using gender, age, education, and race of respondent.

Table 2. Demographic Comparisons for Cities of 250,000 or more: Ages 18+ (%)

	NSHS Weighted	1994 CPS§ Weighted	1995 CPS§ Weighted
Gender			
Male	47.7	46.7	48.3
Female	52.3	53.3	51.7
Age			
18-29	25.7	26.5	21.1
30-39	25.0	23.7	23.5
40-49	19.1	17.7	19.6
50+	30.3	32.2	35.8
Education			
<HS	20.2	23.1 (23.1)	22.2
HS	29.7	30.4 (49.7)	29.5
>HS	50.1	46.5 (27.1)	48.3
Marital Status			
Married	49.7	47.9	51.5
S/D/W	20.2	21.5	23.7
Never Married	30.1	30.6	24.8
Race/Ethnicity			
White	53.9	55.7	57.5
Black	23.4	22.5	23.3
Hispanic	15.6	16.5	15.2
Other	7.1	5.3	4.1

§ Current Population Survey

PROCEDURES FOR SAMPLING ERROR ESTIMATION

The 1996 NSHS is based on a stratified multi-stage probability sample of United States telephone households. This can be described as a complex sampling design, a loosely-used term meant to denote the fact that the sample incorporates special design features such as stratification, clustering and differential selection probabilities (i.e., weighting) that analysts must consider in computing sampling errors for sample estimates of descriptive statistics and model parameters. This section of the 1996 NSHS sample design description focuses on sampling error estimation and construction of confidence intervals for survey estimates of descriptive statistics such as means, proportions, ratios, and coefficients for linear and logistic regression models.

Many standard procedures within popular analysis software systems such SAS, SPSS, Minitab assume simple random sampling (SRS) or equivalently independence of observations in computing standard errors for sample estimates. In general, the SRS assumption results in underestimation of variances of survey estimates of descriptive statistics and model parameters. Confidence intervals based on computed variances that assume independence of observations will be biased (generally too narrow) and design-based inferences will be affected accordingly.

Sampling Error Computation Methods and Programs

Over the past 50 years, advances in survey sampling theory have guided the development of a number of methods for correctly estimating variances from complex sample data sets. A number of sampling error programs which implement these complex sample variance estimation methods are available to NSHS data analysts. The two most common approaches to the estimation of sampling error for complex sample data are through the use of a Taylor Series Linearization of the estimator (and corresponding approximation to its variance) or through the use of resampling variance estimation procedures such as Balanced Repeated Replication (BRR) or Jackknife Repeated Replication (JRR). Newer Bootstrap methods for variance estimation can also be included among the resampling approaches. See Rao and Wu (1988).

1. Linearization Approach

If data are collected using a complex sample design with unequal size clusters, most statistics of interest will not be simple linear functions of the observed data. The objective of the linearization approach is to apply Taylor's method to derive an approximate form of the estimator that is linear in statistics for which variances and covariances can be directly estimated. Kish, 1965; Woodruff, 1971). Linearized variance approximations are derived for estimators of ratio means (Kish and Hess, 1959); finite population regression coefficients and correlation coefficients (Kish and Frankel, 1974); and many other non-linear statistics. Software packages such as

SUDAAN, PC CARP, SAS Proc Genmod, and STATA SVY commands (see below) use the Taylor Series linearization method to estimate standard errors for the coefficients of logistic regression models. In these programs, an iteratively reweighted least squares algorithm is be used to compute maximum likelihood estimates of model parameters. At each step of the model fitting algorithm, a Taylor Series linearization approach is used to compute the variance/covariance matrix for the current iteration's parameter estimates (Binder, 1983).

Available sampling error computation software that utilizes the Taylor Series linearization method includes, but is not limited to: SUDAAN, SUPERCARP, CLUSTERS, OSIRIS PSALMS, OSIRIS PSRATIO, OSIRIS PSTABLES, SAS Proc Genmod, STATA SVY commands. At least SUDAAN, PC CARP, and STATA include procedures for estimation of sampling error both for descriptive statistics such as means, proportion, totals and for parameters of commonly used multivariate models (least squares regression, logistic regression).

2. Resampling Approaches

In the mid-1940's, P.C. Mahalanobis (1946) outlined a simple replicated procedure for selecting probability samples that permits simple, unbiased estimation of variances. The practical difficulty with the simple replicated approach to design and variance estimation is that many replicates are needed to achieve stability of the variance estimator. Unfortunately, a design with many independent replicates must utilize a coarser stratification than alternative designs--to achieve stable variance estimates, sample precision must be sacrificed. Balanced Repeated Replication (BRR), Jackknife Repeated Replication (JRR) and the Bootstrap are alternative replication techniques that may be used for estimating sampling errors for statistics based on complex sample data.

The BRR method is applicable to stratified designs in which two half-sample units (i.e., PSUs) are selected from each design stratum. The conventional "two PSU-per-stratum" design in the best theoretical example of such a design although in practice, collapsing of strata (Kalton, 1977) and random combination of units within strata are employed to restructure a sample design for BRR variance estimation. The half-sample codes prepared for the 1996 NSHS data set require the collapsing of non self-representing strata and the randomized combination of selection units within self-representing (SR) strata. When full balancing of the half-sample assignments is employed (Wolter, 1985), BRR is the most computationally efficient of the replicated variance estimation techniques. The number of general purpose BRR sampling error estimation programs in the public domain is limited. Research organizations such as Westat, Inc. and the National Center for Health Statistics have developed general-purpose programs for BRR estimation of standard errors. Another option is to use SAS or SPSS Macro facilities to implement the relatively simple BRR algorithm. The necessary computation formulas and Hadamard matrices to define the half-sample replicates are available in Wolter (1985).

With improvements in computational flexibility and speed, jackknife (JRR) and bootstrap methods for sampling error estimation and inference have become more common (Rao and Wu, 1988). Few general-purpose programs for jackknife estimation of variances are available to analysts. OSIRIS REPERR has a JRR module for estimation of standard errors for regression and correlation statistics. Other stand-alone programs may also be available in the general survey research community, such as WesVarPC from Westat, Inc. Like BRR, the algorithm for JRR is relatively easy to program using SAS, SPSS or S-Plus macro facilities.

BRR and JRR are variance estimation techniques, each designed to minimize the number of "resamplings" needed to compute the variance estimate. In theory, the bootstrap is not simply a tool for variance estimation but an approach to actual inference for statistics. In practice, the bootstrap is implemented by resampling (with replacement) from the observed sample units. To ensure that the full complexity of the design is reflected, the selection of each bootstrap reflects the full complexity of the stratification, clustering and weighting that is present in the original sample design. A large number of bootstrap samples are selected and the statistic of interest is computed for each. The empirical distribution of the estimate that results from the large set of bootstrap samples can then be used to a variance estimate and a support interval for inference about the population statistic of interest.

In most practical survey analysis problems, the JRR and Bootstrap methods should yield similar results. Most survey analysts should choose JRR due to its computational efficiency. NSHS data analysts interested in the bootstrap technique are referred to LePage and Billard (1992) for additional reading and a bibliography for the general literature on this topic.

One aspect of BRR, JRR and bootstrap variance estimation that is often pushed aside in practice is the treatment of analysis weights. In theory, when a resampling occurs (i.e., a BRR half sample is formed), the analysis weights should be recomputed based only on the selection probabilities, nonresponse characteristics and post-stratification outcomes for the units included in the resample. This is the correct way of performing resampling variance estimation; however, in practice acceptable estimates can be obtained through use of the weights as they are provided on the public use data set.

Sampling Error Computation Models

Regardless of whether linearization or a resampling approach is used, estimation of variances for complex sample survey estimates requires the specification of a sampling error computation model. NSHS data analysts who are interested in performing sampling error computations should be aware that the estimation programs identified in the preceding section assume a specific sampling error computation model and will require special sample clustering variables. Individual records in the analysis data set must be assigned sample clustering variables which identify to the programs the complex structure of the sample (stratification, clustering) and are compatible with the computation algorithms of the various programs. To facilitate the computation of sampling error for statistics based on 1996 NSHS data, design-specific sample clustering variables are included in all public-use versions of the data set. Although minor recoding may be required to conform to the input requirements of the individual programs, the sample clustering variables that are provided should enable analysts to conduct either Taylor Series or Replicated estimation of sampling errors for survey statistics.

References

- Binder, D. A. (1983), "On the variances of asymptotically normal estimators from complex surveys," International Statistical Review, Vol. 51, pp. 279-292.
- Groves, R. M., Biemer, P. P., Lyberg, L. L., Massey, J. T., Nicholls II, W. L., Waksberg, J. (1988), Telephone Survey Methodology. New York: John Wiley & sons, Inc.
- Kalton, G. (1977), "Practical methods for estimating survey sampling errors," Bulletin of the International Statistical Institute, Vol. 47, 3, pp. 495-514.
- Kish, L. (1965), Survey Sampling. New York: John Wiley & Sons, Inc.
- Kish, L., & Frankel, M.R. (1974), "Inference from complex samples," Journal of the Royal Statistical Society, B, Vol. 36, pp. 1-37.
- Kish, L., & Hess, I. (1959), "On variances of ratios and their differences in multi-stage samples," Journal of the American Statistical Association, 54, pp. 416-446.
- LePage, R., & Billard, L. (1992), Exploring the Limits of Bootstrap. New York: John Wiley Sons, Inc.
- Mahalanobis, P.C. (1946), "Recent experiments in statistical sampling at the Indian Statistical Institute," Journal of the Royal Statistical Society, Vol. 109, pp. 325-378.
- Rao, J.N.K & Wu, C.F.J. (1988.), "Resampling inference with complex sample data," Journal of the American Statistical Association, 83, pp. 231-239.
- The National Election Studies, Center for Political Studies, University of Michigan. Electronic resources from the NES World Wide Web site (http://www.umich.edu/~nes/resources/data/nes94/sample-design.html). Ann Arbor, MI: University of Michigan, Center for Political Studies [producer and distributor], 1995-1998.
- Wolter, K.M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.
- Woodruff, R. S. (1971), "A simple method for approximating the variance of a complicated estimate," Journal of the American Statistical Association, Vol. 66, pp. 411-414.