Journal of Official Statistics, Vol. 19, No. 1, 2003, pp. 31–43

Sequential Weight Adjustments for Location and Cooperation Propensity for the 1995 National Survey of Family Growth

Vincent G. Iannacchione[1]

Distinct patterns of location and cooperation were found among women selected for Cycle 5 of the National Survey of Family Growth (NSFG-5). For example, minority women were harder to locate than were other sample women. Once found, however, these same women were more likely to participate in the survey than other sample women. Two logistic regression models were developed sequentially so that predictors related to the locating process could be distinguished from those related to the cooperation process. For the location propensity model, the design-weighted logistic regression algorithm was modified to compute adjustment factors for the sampling weights that preserved the full-sample weighted means for specified analysis domains among women who were located for the NSFG-5. For the cooperation propensity model, a similar process was used to preserve the location-adjusted weighted means among women who participated in the survey. Receiver Operating Characteristics (ROC) curves were used to assess the overall predictive ability of the combined models. The linkage of the NSFG-5 to the 1993 National Health Interview Survey (NHIS) enabled a large number of candidate predictors obtained from the NHIS interview to be considered for each model. As expected, predictors indicating the presence or absence of NHIS contact data (e.g., a telephone number) were significant factors affecting location propensity. However, the refusal by some NHIS participants to provide certain types of contact data also adversely affected cooperation propensity and was interpreted as a predictor of resistance to participation in the NSFG-5.

Key words: Weight adjustments for nonresponse; linked sample design; logistic regression; ROC curves.

1. Introduction

The National Survey of Family Growth (NSFG) is designed and administered by the National Center for Health Statistics (NCHS), an agency of the U.S. Department of Health and Human Services. Established in 1971, the purpose of the survey is to produce national estimates and an information base on factors affecting pregnancy and birth rates, such as sexual activity, contraceptive use, infertility, sources of family planning services, and the health of women and infants. Interviewing for the first cycle of the NSFG was conducted in 1973; other cycles were conducted in 1976 (Cycle 2), 1982 (Cycle 3), 1988 (Cycle 4), and in 1995 (Cycle 5). Interviewing and data processing for the 1995 NSFG were conducted by the Research Triangle Institute, under a contract with NCHS.

[1] RTI International, 1615 M St., Suite 740, NW, Washington, DC, U.S.A. Email: vince@rti.org
Acknowledgment: The views presented in this article are those of the author alone and do not necessarily reflect the views of the National Center for Health Statistics.

The sample for the 1995 NSFG-5 (Potter et al. 1998) consisted of a subsample of 14,000 women between the ages of 15 and 44 from households that participated in the 1993 National Health Interview Survey (NHIS). The NHIS is a stratified multistage household survey that covers the civilian noninstitutionalized population of the United States. The NHIS is redesigned each decade using data from the most recent census. The 1995 NSFG used the NHIS sample based on the design developed for the period 1985 to 1994 (Massey et al. 1989). The objectives of linking the two surveys are to reduce the cost of the NSFG while maintaining the statistical accuracy of the survey estimates, and to expand analytic opportunities by linking data from the NHIS to the NSFG.

All households in the 1993 NHIS containing Hispanic and non-Hispanic black women were included in the NSFG sample, along with about 55 percent of NHIS households containing white and other (nonblack, non-Hispanic) women. As a result, black and Hispanic women were sampled at a higher rate than other women for the NSFG. In households with more than one eligible woman, one was randomly selected for the NSFG. Sampled women who moved between the 1993 NHIS and 1995 NSFG were traced to their new addresses.

Sampling weights, adjusted for nonresponse and noncoverage reflected the complex design and the unequal sampling rates used to select the NSFG-5 sample. The weight assignment process consisted of four adjustment factors that were applied to the NHIS weights of the 10,847 NSFG participants to reflect the subsampling of the NHIS sample as well as to adjust for nonlocation, noncooperation, and noncoverage of sample women:

$\mathbf{W0_i}$ (the NHIS weight assigned to sample woman i)
$\cdot\,\mathbf{A1_i}$ (the adjustment for differential subsampling rates for the NSFG-5)
$\cdot\,\mathbf{A2_i}$ (the adjustment for inability to locate)
$\cdot\,\mathbf{A3_i}$ (the adjustment for refusal to cooperate)
$\cdot\,\mathbf{A4_i}$ (the post-stratification adjustment)
$= \mathbf{W4_i}$ (the fully adjusted NSFG-5 weight).

A flowchart of the weight assignment process for the NSFG-5 is shown in Figure 1. The remainder of this article focuses on the location and cooperation propensity models that were developed and used in the weight adjustment process.

2. Development of the Location and Cooperation Propensity Models

The linkage between the NHIS and the NSFG provided a large amount of data about NSFG sample members including those who could not be located and those who were located but refused to participate in the NSFG. Between the 1993 NHIS and the 1995 NSFG, many women in the sample moved, and a substantial effort was made to identify their new addresses. Tracing activities were successful in locating 94.6 percent of all sample women. The cooperation rate among sample women who were located and eligible for the survey was 83.2 percent. Only 1.6 percent of the sample women who were located were not contacted. The overall response rate among eligible sample women was 78.7 percent.

Two propensity models were developed for the processes related to nonresponse:

1. The process of locating sample women (especially those who moved during the two years between the NHIS and the NSFG-5);
2. The process of contacting and enlisting cooperation among sample women who were located.

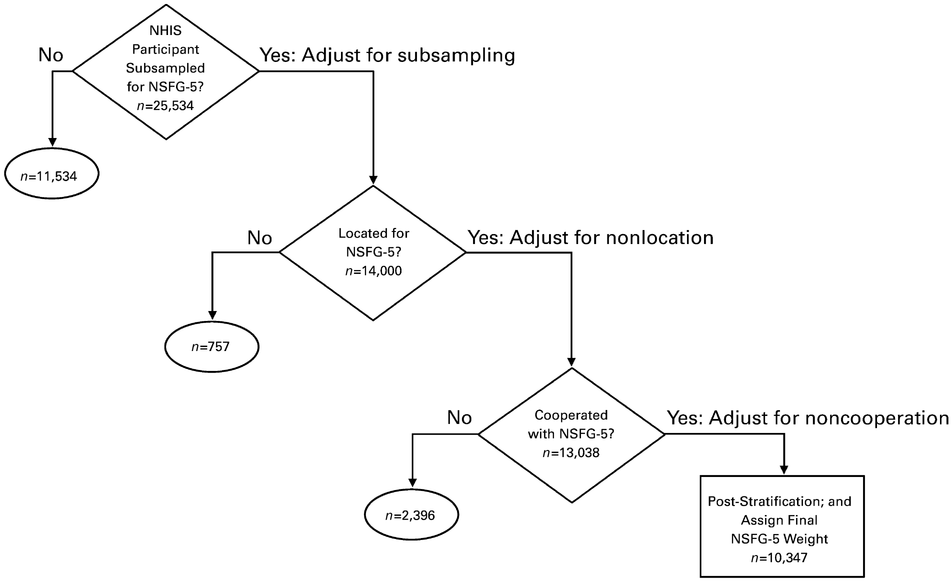

Fig. 1. NSFG-5 weight adjustment process

The contact and cooperation processes were combined because contact was made with all but 210 of the 13,243 sample women who were located. As a result, the propensity to cooperate in the NSFG-5 was largely conditional on contact being made.

When available, NHIS information about the sample woman was used for tracing, as were contact data for the NHIS reference person (typically a spouse) and the NHIS contact person (usually a relative, neighbor, or friend). About 38 percent of all sample women had one or more pieces of tracing information missing: 31 percent had missing Social Security Numbers (SSNs), 25 percent had no contact person listed, and 11 percent either did not have a telephone or refused to give their telephone numbers.

Some of the variables from the NHIS indicated varying degrees of ability to locate, while others indicated varying degrees of resistance or hostility to surveys. For example, failure to obtain a telephone number was related to inability to locate, whereas refusal to give an SSN or the name of a contact person indicated resistance to participation. These distinctive patterns suggested that the location process be treated as a different outcome variable than the cooperation process among those who were located.

Groves and Couper (1998) point out that the location and cooperation processes are different in most household interview surveys. This was certainly the case for the NSFG-5. The location process was largely dependent on the success of the tracing activities which included address and telephone matching, DMV record searches, and field tracing by the U.S. Bureau of the Census, whereas the cooperation process was affected by numerous factors including the sample woman's reaction to the NHIS interview two years prior, and the interaction between the interviewer and the sample woman. As a result, the bias caused by the inability to locate an NSFG-5 sample member may not be the same as the bias caused by the located sample member's refusal to participate.

Folsom (1991) modified the design-weighted logistic regression algorithm to derive a response propensity adjustment procedure for sampling weights that achieves equalization of respondent and full-sample weighted means for categorical as well as continuous response predictors. The procedure uses logistic regression to model the functional relationship between a set of response predictors and a dichotomous response outcome. If the response predictors are correlated with the substantive variables of interest, and if the response propensities are nonzero over conceptual repetitions of the study, the model-based adjustment factors applied to the sampling weights greatly reduce the potential for nonresponse bias. In addition, response propensity modeling provides a formal statistical setting for evaluating variables believed to be related to response. This was particularly useful for evaluating the large number of potential predictors collected for the NHIS.

Two logistic regression models were developed for NSFG-5 sample women so that predictors related to the locating process could be distinguished from those related to the cooperation process. Both models used the same set of potential predictors available from the NHIS and shown in Table 1. The two models enabled separate weight adjustment factors to reflect the distinct patterns of availability, including change of address, lack of some or all contact information, and resistance to participation. Mobility was an expected artifact of the NSFG-5 sampling design because of the linkage of the 1993 NHIS and the long time period between the two surveys. The lack of contact information was generally considered as indirect resistance to participation but also represented, in some cases, a failure to collect accurate and complete contact information during the NHIS interview.

Segmentation analysis using the CHAID (Chi-squared Automatic Interaction Detection) software (Magidson 1993) was used to enable a parsimonious model-building process that focused on segments that showed distinct location and cooperation propensities, and avoided the dilemma of examining ‘‘all possible interactions’’ that is inherent to regression modeling. The segmentation algorithm successively split the sample into smaller subgroups based on location or cooperation rates. Some of the multi-level categorical variables used in the segmentation (e.g., age category) were collapsed if the algorithm did not detect differences at the 0.05 level of significance. The splitting process continued until either no other predictors could be found or the segment size fell below 100 sample women. For design consistency, a weighted segmentation of the sample using CHAID was performed and then included with main-effect predictors in the logistic regression procedure in SUDAAN (Shah et al. 1997).

The overall response propensity for each sample woman i was subdivided into the following components.

$$L_i = \begin{cases} 1 & \text{if sample woman } i \text{ was located, and} \\ 0 & \text{otherwise} \end{cases}$$

and:

$$R_i = \begin{cases} 1 & \text{if sample woman } i \text{ cooperated, and} \\ 0 & \text{otherwise} \end{cases}$$

Then, the overall probability that sample woman i responded was written as:

$$\begin{aligned} P[R_i = 1] &= P[L_i = 1]P[R_i = 1 \mid L_i = 1] \\ &= \lambda_i \rho_i \end{aligned}$$

Table 1. Candidate predictors of location and cooperation propensity[1]

Candidate predictor	Significant predictor[2]	
	Location model	Cooperation model
NHIS demographic variables		
Family income	X	X
Age category	X	X
Marital status	X	X
Poverty level	X	X
Region of country	X	X
MSA status	X	X
Education	X	X
Health status	X	X
Race	..	X
Ethnicity	..	X
Major activity	..	X
Class of worker	..	X
Employment status	..	X
Number of children in household	..	X
Number of doctor visits in past year	..	X
Time since last doctor visit	..	X
Number of conditions	..	X
Predominantly black area	X	..
Family relationship of sample woman	X	..
Urban/rural	..	..
Living quarters	..	..
Number of families in household	..	..
Relationship to NHIS reference person	..	..
NHIS contact variables		
Name not provided	X	X
Record of calls	X	X
Contact name provided	X	X
NHIS respondent status	..	X
Telephone number refused	X	X
SSN refused	X	X
Number of calls	..	..
Number of additional contacts	..	..
Number of callbacks for SSN	..	..
Number of callbacks for immunizations	..	..
Refused height, weight, or health status	..	..

X: Variable is significant at $\alpha = 0.10$

[1] The list of candidate variables comprised all variables on the 1993 NHIS Public Use Files believed to be potentially related to location and/or cooperation propensity.

[2] Predictors included as main effects and/or segments in the final logistic regression models.

The overall probability of response was estimated with two logistic regression models. The first model for location propensity was applied to the entire sample of 14,000 sample women. The second model for cooperation propensity was applied to the 13,038 sample women who were located and eligible for the study.

3. Location Propensity Model

The following logistic model was developed to estimate the probability that sample woman i was located:

$$\hat{\lambda}_i = P[L_i = 1 \mid X_i, \hat{\beta}]$$
$$= [1 + \exp(-X_i\hat{\beta})]^{-1}$$

where X_i is a vector of NHIS location predictors (main effects and interaction terms).

The logistic regression coefficients $\hat{\beta}$ were estimated iteratively to satisfy the following estimation equations:

$$\sum_{i \in S} (W_i \div \hat{\lambda}_i)\mathbf{X}_i^T \hat{\lambda}_i = \sum_{i \in S} (W_i \div \hat{\lambda}_i)\mathbf{X}_i^T L_i$$

where

S = Sample of 14,000 women, and
W_i = Initial NSFG-5 sampling weight for sample woman i.

The estimation equations preserve the initial NSFG-5 weight sums of location predictors in the adjusted weights of sample women who were located.

After the location propensity model was finalized, the location-adjusted weight for sample woman i was computed as

$$W_i^L = \begin{cases} W_i \div \hat{\lambda}_i & \text{if } L_i = 1 \\ 0 & \text{if } L_i = 0 \end{cases}$$

Nonzero location-adjusted weights were assigned to the 13,038 sample women who were located and eligible for the NSFG-5. These interim weights were used to develop the following cooperation propensity model.

4. Cooperation Propensity Model

The probability of cooperation given that sample woman i was located and eligible was estimated as:

$$\hat{\rho}_i = P[R_i = 1 \mid L_i = 1, Z_i\hat{\theta}]$$
$$= [1 + \exp(-Z_i\hat{\theta})]^{-1}$$

where Z_i is a vector of NHIS cooperation predictors (main effects and interaction terms).

Analogous to the location propensity model, the logistic regression coefficients $\hat{\theta}$ were estimated iteratively to satisfy the following estimation equations:

$$\sum_{i \in \xi} (W_i^L \div \hat{\rho}_i)Z_i^T \hat{\rho}_i = \sum_{i \in \xi} (W_i^L \div \hat{\rho}_i)Z_i^T R_i$$

where ξ is the sample of 13,038 located and eligible women.

The estimation equations preserve the location-adjusted weight sums of the cooperation predictors in the adjusted weights of sample women who cooperated.

Then, the cooperation-adjusted weight was computed as:

$$W_i^R = \begin{cases} W_i \div (\hat{\lambda}_i \hat{\rho}_i) & \text{if } R_i = 1 \\ 0 & \text{if } R_i = 0 \end{cases}$$

Nonzero cooperation adjusted weights were assigned to the 10,847 sample women who participated in the NSFG-5. The most influential predictors of $\hat{\lambda}_i$, the estimated location propensity for sample woman i, and $\hat{\rho}_i$, the estimated cooperation propensity for sample woman i, are described in the next sections.

5. Factors Affecting Location Propensity

As expected, predictors indicating the presence or absence of NHIS contact data were significant factors in the final location propensity model. The segmentation of the sample shown in Figure 2 suggests that these predictors interacted with a number of demographic factors, especially family income. For example, among sample women with low or unknown family incomes, the presence or absence of a telephone number resulted in a 10.1 percentage point difference in the location rate. In fact, the lowest segment-level location rate (63.4 percent) occurred among sample women with low or unknown family incomes who either refused or did not have telephone numbers and who completed the NHIS but did not provide their names. In contrast, the lack of segmentation among sample women with (known) annual family incomes of 20,000 USD or more suggests that, among sample women who were willing to provide income (a traditionally sensitive item), there was no difference in the location rate as a function of the provision of contact information.

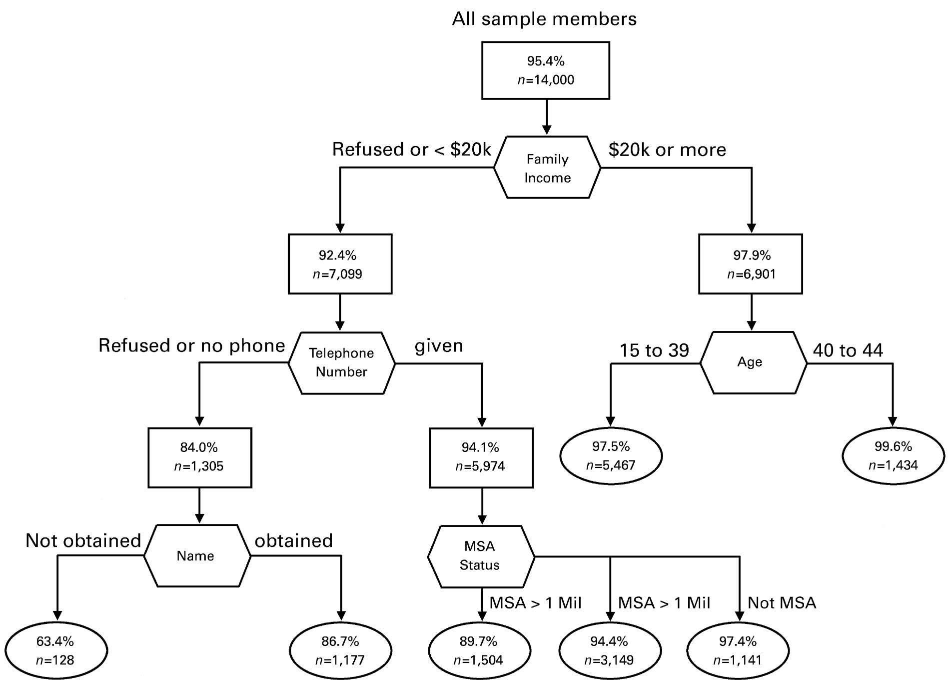

Fig. 2. Segmentation of NSFG-5 sample members by weighted location rate

In addition to the interactive effects identified by the segmentation analysis, several demographic variables were significant ''main effect'' factors in the location propensity model. For example, after adjusting for other covariates in the model, unmarried women were significantly harder to locate than their married counterparts. Similarly, women with no college education had lower location propensities than those with some college education. Region of the country was also a significant factor, with women in the Midwest located more easily than those in other regions of the U.S.

6. Factors Affecting Cooperation Propensity

As in the location propensity model, several of the predictors related to the presence or absence of contact data were also significant in the cooperation propensity model. For example, more than 2,000 sample women refused to provide an SSN during the NHIS but were subsequently located and found eligible for the NSFG-5. Once located, however, these sample women were significantly less likely to participate than those who provided an SSN. Similar patterns can be seen for those who refused to provide a telephone number or the name of a contact person in the segmentation modeling of located eligibles shown in Figure 3. Unlike the location model, the absence of contact data in the cooperation model was thought to be an indication of hostility to the interview process.

Although not significant in the location propensity model, both race and Hispanic background were important predictors of cooperation. Among sample women who provided an SSN to the NHIS, Asians and Pacific Islanders were less likely to cooperate than other racial groups. Among those who refused to provide an SSN, the cooperation rate among Hispanic women was 15.3 percentage points higher than among non-Hispanic women. In

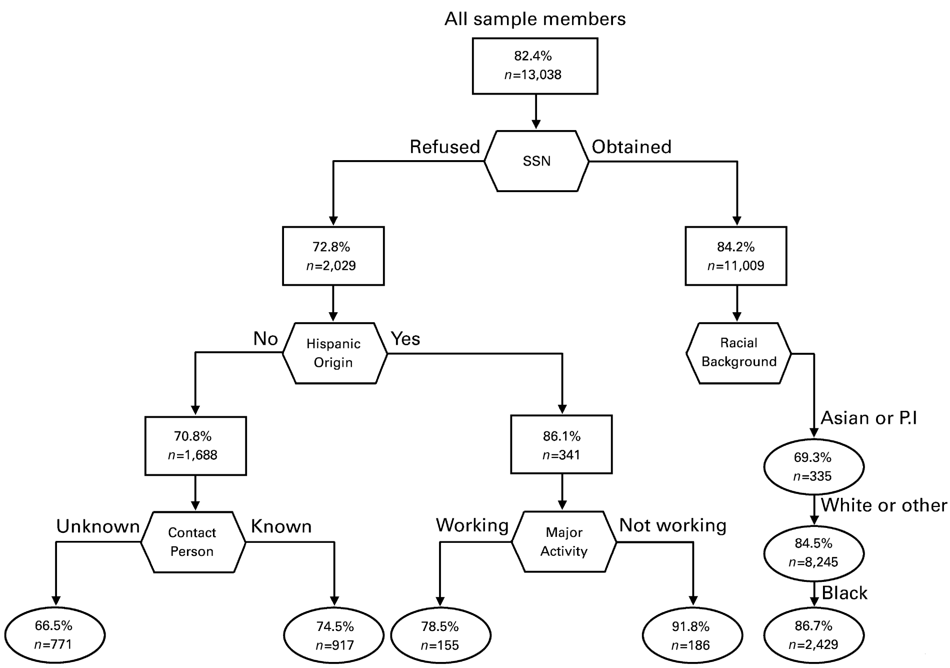

Fig. 3. Segmentation of located and eligible NSFG sample members by weighted cooperation rate

fact, the 91.8 percent participation rate among nonworking Hispanic women who refused to provide (or did not have) SSNs to the NHIS is a notable exception to the general trend of being able to predict hostility to the NSFG-5 by hostility to the NHIS.

The significant noninteractive (i.e., main effect) factors in the cooperation propensity model included age, number of children, and number of health conditions. Sample women between 15 and 24 were more likely to cooperate than older women while those with one or no children were less likely to cooperate than those with two or more children. Also, sample women with no reported health conditions were less likely to cooperate than those who reported one or more health conditions.

7. Evaluating the Overall Response Propensity

Generalized Wald statistics, adjusted for design effects, were used to test the goodness-of-fit for the location and cooperation propensity models. Although each model was found to be significant, the predicted overall response propensity ($\hat{\lambda}_i \hat{\rho}_i$) was not amenable to a conventional regression analysis because of the lack of independence between the models. Therefore, a Receiver Operating Characteristics (ROC) curve was used to assess the overall predictive ability of the combined models.

As shown in Figure 4, the area under the ROC curve developed for the overall predicted response propensity was 0.65, which corresponds to a highly significant Wilcoxon test statistic (Hanley and McNeil 1982). The curve indicates that in two of every three randomly chosen pairs of sample women, one responding and the other not responding, the predicted overall response propensity of the respondent will be larger than that of the nonrespondent. This level of discrimination implies that the NHIS variables used in the two models are informative but not definitive predictors of a sample woman's overall response propensity.

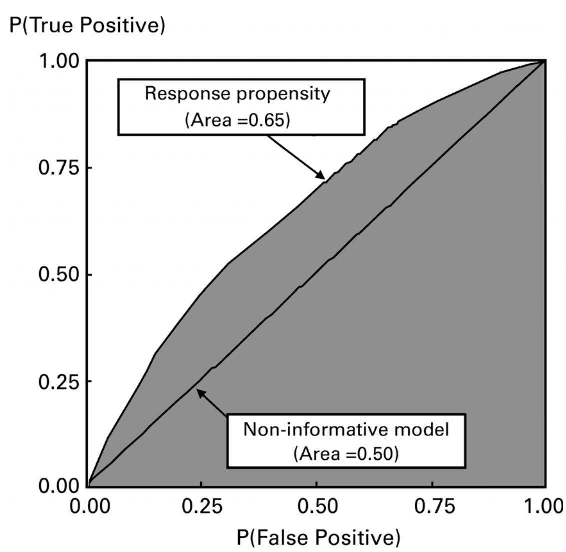

Fig. 4. ROC Curve for Overall Response Propensity

The moderate predictive ability of the combined propensity models suggests that factors other than the contact and demographic information obtained from the NHIS are important predictors of response propensity. For example, some sample women may have avoided being located because of immigration or credit problems; or, sample women who were located may have refused to cooperate because of the sensitive nature of the NSFG-5 instrument. Nevertheless, the propensity models not only offer insight into the factors that affect the location and cooperation process but also provide weight adjustment factors to compensate for their potential biasing effects.

A noticeable distinction between weight adjustments based on response propensity and those based on weighting classes is that the adjustment factors for response propensity are applied to individual sample members rather than a class or group of sample members. As a result, response propensity adjustments can help to reduce nonresponse bias by following the actual response pattern more closely than weighting class adjustments. However, these gains in accuracy may be offset if the variation among the adjusted weights causes an excessive increase in the variances of sample estimates (Potter 1990).

A summary of the adjustments made is to the NSFG-5 weights for NHIS subsampling, location propensity, and cooperation propensity is shown in Table 2. The unequal weighting effects are shown by race/ethnicity along with the mean and maximum adjustment factors. Note that the mean adjustment factor for subsampling the ''Other'' race stratum (2.16) is larger than for the Hispanic or the non-Hispanic Black stratum. This reflects the lower sampling rates that were applied to women in the ''Other'' race stratum. As can be seen, the subsequent adjustments for location and cooperation propensity produced little additional effect on unequal weighting.

Table 2. NSFG-5 weight adjustment summary

Type of weight adjustment	Number of sample women	Mean adjustment factor	Maximum adjustment factor	Unequal weighting effect[1]
Adjustments for NHIS subsampling				
Hispanic	2,097	1.28	7.00	1.30
Black, non-Hispanic	3,205	1.27	6.00	1.41
Other	8,698	2.16	14.24	1.06
Overall	14,000	1.83	14.24	1.20
Adjustments for location propensity				
Hispanic	1,926	1.08	2.54	1.31
Black, non-Hispanic	2,939	1.08	4.02	1.42
Other	8,378	1.04	2.90	1.07
Overall	13,243	1.06	4.02	1.20
Adjustments for cooperation propensity				
Hispanic	1,612	1.15	1.89	1.32
Black, non-Hispanic	2,465	1.19	2.66	1.42
Other	6,770	1.23	3.04	1.09
Overall	10,847	1.20	3.04	1.23

[1] The unequal weighting effect measures the increase in sampling variance attributable to unequal weighting compared to an equally weighted sample.

8. The Influence of NHIS Contact Data on NSFG-5 Location and Cooperation Propensities

The sequential model building process used to adjust the sampling weights for the NSFG-5 provides an opportunity to assess the effects of the presence or absence of contact data obtained from the NHIS on NSFG-5 location and cooperation propensity. The assessment of the contact data may be useful for planning subsequent cycles of the NSFG or other surveys with linked sample designs. For location propensity, this analysis provides insight into the success of the tracing activities when various pieces of NHIS contact data are missing. For cooperation propensity, the analysis provides insight into the predictive relationship between a sample woman's refusal to provide contact data for the NHIS and her eventual reluctance to participate in the NSFG-5.

Predictive margins (Korn and Graubard 1997) were computed to estimate the overall effect of three sources of NHIS contact data (SSN, telephone number, and name of contact person) on NSFG-5 location and cooperation propensities. Each of these variables was a significant predictor in both models and each interacted with a number of demographic characteristics.

The predictive margins shown in Table 3 may be viewed as the expected location and response rates if everyone in the sample provided (or did not provide) a piece of contact information. For example, the observed location rate for sample members who provided the name of a contact person was 6.7 percentage points higher than for those who refused. However, if everyone in the sample had provided the name of a contact person, the location rate would only increase by 2.1 percentage points compared to a sample where no one provided the contact name. A similar result can be seen for refusing to provide an SSN. In contrast, if a telephone number was not obtained or there was no telephone number, the observed location rate was 10.8 percentage points lower than when the number was obtained. Even after adjusting for other location predictors, however, the 4.1 percentage

Table 3. Influence of NHIS contact data on NSFG-5 location and cooperation propensity

NHIS contact data	Location propensity[1]		Cooperation propensity[2]	
	Observed rate	Predictive margin	Observed rate	Predictive margin
	% ± 95% CI	% ± 95% CI	% ± 95% CI	% ± 95% CI
Name of contact person:				
Obtained for NHIS	96.5 ± 0.5	95.8 ± 0.6	84.2 ± 0.9	83.0 ± 1.3
Refused	89.8 ± 1.4	93.8 ± 1.0	74.0 ± 2.4	78.8 ± 2.4
Social security number:				
Given for NHIS	96.3 ± 0.5	95.8 ± 0.6	84.1 ± 0.9	83.3 ± 1.2
Refused	89.8 ± 1.5	93.7 ± 1.5	72.7 ± 2.4	76.3 ± 2.9
Telephone number:				
Given for NHIS	96.4 ± 0.4	96.0 ± 0.6	82.7 ± 0.9	82.5 ± 1.2
Refused or no phone	85.6 ± 2.2	91.9 ± 1.4	79.2 ± 3.0	81.6 ± 3.2

[1] Among 14,000 NSFG-5 sample members. The observed location rates and predictive margins were computed using the NSFG-5 sampling weights.
[2] Among 13,038 NSFG-5 sample members who were located and eligible for the survey. The observed cooperation rates and predictive margins were computed using the location-adjusted weights.

point difference in predictive margins implies that the presence of a telephone number was an important factor affecting location propensity.

As shown in the segmentation analysis, sample members who refused to provide their SSN during the NHIS were significantly less likely to cooperate in the NSFG-5 than those who did provide their SSN. Even after adjusting for other predictors in the cooperation model, the absence of an SSN was found to adversely affect cooperation rates by 7.0 percentage points. The observed cooperation rate among sample members who provided the name of a contact person during the NHIS was 10.2 percentage points more than among those who refused. However, this difference was reduced to only 4.2 percentage points when other cooperation predictors were taken into account. In contrast, after adjusting for other predictors in the cooperation model, the predictive margins for the presence or absence of a telephone number indicate virtually no effect on cooperation propensity.

9. Summary and Discussion

The linkage between the NHIS and the NSFG-5 provided a large amount of contact and demographic information about NSFG-5 sample members, including those who could not be located and those who were located but refused to participate in the NSFG-5. The information obtained from the NHIS was used to develop sequential location and cooperation propensity models for nonrespondents to the NSFG-5. A design-weighted logistic regression algorithm was used to preserve the full-sample weighted means of significant predictors first among located sample women and then among cooperating sample women. The predicted values obtained from the models were evaluated with an ROC curve and found to be informative predictors of a sample woman's overall response propensity. To adjust the NSFG-5 sampling weights for nonresponse, the predicted values for each sample woman were inverted and then used as adjustment factors. The adjusted weights did not produce noticeable increases in the variances of sample estimates.

The development of separate models for location and cooperation propensity revealed distinct factors for each process. As expected, location propensity was affected by the presence or absence of contact data, primarily a telephone number. Unexpectedly, the cooperation propensity model provided insight into the predictive relationship between a sample woman's refusal to provide contact data for the NHIS and her eventual reluctance to participate in the NSFG-5.

The findings obtained from sequential modeling of the response process can provide useful information to planners of other linked surveys. For example, the model parameters used to develop the NSFG-5 models could be applied to the sample members selected for the next version of the survey or to a similar linked design. Then, sample members with low predicted location propensities could be identified and subjected to intensive tracing activities early in the data collection process. Similarly, sample members with a low response propensity could be assigned to specially trained field interviewers to help reduce the rate of refusals.

The sequential modeling of response propensity also can be applied to other sample designs where the data are collected serially. For example, a variation of the sequential weight adjustment process described in this article was used earlier for the 1988 Army Family Research Program (Iannacchione et al. 1991) which surveyed soldiers and their spouses about the role of family factors in retention and readiness among Army personnel. For this survey, each sample soldier completed a questionnaire and then was asked to provide his/her spouse's mailing address for use in mail survey of spouses. The requirement that a spouse address be obtained directly from the soldier motivated the development of two models: the first model predicted the probability of obtaining a spouse address from the soldier, and the second model predicted the probability of a spouse response conditional on obtaining a mailing address. Other applications of a sequential response propensity modeling weight adjustment process include longitudinal surveys.

10. References

Folsom, R.E. (1991). Exponential and Logistic Weight Adjustments for Sampling and Non-response Error Reduction. Proceedings of the American Statistical Association, Social Statistics Section, 197–201.

Groves, R.M. and Couper, M.P. (1998). Nonresponse in Household Interview Surveys. John Wiley & Sons, New York, NY.

Hanley, J.A. and McNeil, B.J. (1982). The Meaning and Use of the Area Under a Receiver-operating Characteristic (ROC) Curve. Diagnostic Radiology, 143, 29–36.

Iannacchione, V.G., Milne, J.G., and Folsom, R.E. (1991). Response Probability Weight Adjustments Using Logistic Regression. Proceedings of the American Statistical Association, Survey Research Methods Section, 637–642.

Korn, E.L. and Graubard, B.I. (1997). Predictive Margins with Survey Data. Proceedings of the American Statistical Association, Survey Research Methods Section, 651–656.

Magidson, J. (1993). SPSS for Windows: CHAID, Release 6.0. Statistical Innovations, Inc., Belmont, MA.

Massey, J.T., Moore, T.F., Parsons, V.L., and Tadros, W. (1989). Design and Estimation for the National Health Interview Survey, 1985–94. U.S. National Center for Health Statistics. Vital Health Statistics 2(101).

Potter, F.J., Iannacchione, V.G., Mosher, W.D., Mason, R.E., and Kavee, J.D. (1998). Sample Design, Sampling Weights, Imputation, and Variance Estimation in the 1995 National Survey of Family Growth. U.S. National Center for Health Statistics. Vital Health Statistics 2(124).

Potter, F.J. (1990). A Study of Procedures to Identify and Trim Extreme Sampling Weights. Proceedings of the American Statistical Association, Section on Survey Research Methods, 225–230.

Shah, B.V., Barnwell, B.G., and Bieler, G.S. (1997). SUDAAN User's Manual, Release 7.5. Research Triangle Institute, Research Triangle Park, NC.

Received March 1999
Revised September 2002