NONRESPONSE ADJUSTMENTS FOR A TELEPHONE FOLLOW-UP TO A NATIONAL IN-PERSON SURVEY

by Hüseyin Göksel, David R. Judkins and William D. Mosher
David Judkins, Westat, Inc., 1650 Research Blvd., Rockville, MD 20850

Keywords: Response propensity, Automatic Interaction Detection

Abstract

The National Survey of Family Growth (NSFG), Cycle IV Baseline, was based on in-person interviews in 1988 with 8,450 women selected from several cycles of the National Health Interview Survey (NHIS). Nonresponse adjustments for Cycle IV Baseline were described in a paper for the 1989 ASA conference. In 1990, a telephone reinterview, the CATI Phase, was conducted with about 5,700 of the same women who were interviewed in 1988. Some of the NHIS variables that predicted nonresponse to the Baseline also predicted nonresponse to the CATI Phase, but several Baseline variables were considerably better predictors. Mobility was the most powerful predictor, but race, Hispanic origin, education and other socio-economic variables and several variables specific to the subject matter of NSFG were also closely associated with nonresponse. The paper describes how the determinants of nonresponse were modelled, and how the sample weights were adjusted for non-response.

1. Introduction

Longitudinal components to major demographic surveys are becoming more and more common. One feature that all longitudinal surveys share is higher nonresponse than cross-sectional surveys. This feature is, of course, well known. It is tolerated as the price of either the unique information that can be obtained longitudinally or by the cost savings that can be realized by multiple visits, or by the increased correlations over time that improve the reliability of cross-sectional change estimates. Less commonly recognized is the fact that the data obtained on early visits can be used to reduce the risk of bias due to the higher rates of nonresponse on subsequent visits.

In this paper, we describe the nonresponse adjustment for a telephone reinterview of Cycle IV of the National Survey of Family Growth (NSFG). The reinterview together with a first interview of a supplementary sample of women just recently eligible for the survey (teenagers aged 15 to 17 and some months) was known collectively as the CATI Phase of Cycle IV. (CATI refers to Computer-Assisted Telephone Interviews.)

The NSFG is conducted by the National Center for Health Statistics (NCHS) and designed to provide national estimates of factors associated with fertility, contraception, and reproductive health among women 15-44 years of age in the United States. Specific objectives for the CATI Phase were: (1) to update program-relevant statistics on adoption, teenage sexual activity, contraception, and family planning; (2) to add new data on AIDS-related behavior, sexually transmitted diseases, and other topics; and (3) to create a longitudinal database with information on changes over time for individual women.

The 1988 NSFG sample (the "Baseline") was drawn from households who participated in the National Health Interview Survey (NHIS) between October of 1985 and March of 1987. Women who were in the civilian noninstitutional population and 15-44 years of age on March 15, 1988, were eligible for the 1988 baseline interview (which was conducted in-person). Women who participated in the 1988 baseline were eligible for reinterview if they were under 45 years of age on August 15, 1990 .

More details on the NSFG Cycle IV Baseline may be found in Judkins, Mosher, and Botman (1991). In this paper, we concentrate on the methodology for forming nonresponse adjustment cells and on patterns of nonresponse propensity that may carry over to other follow-back telephone surveys.

2. Methodology

For this survey we formed nonresponse adjustment cells using the same general approach previously used by the Institute for Social Research (ISR 1979, 1986) on the Panel Survey of Income Dynamics (PSID), Kalton, Lepkowski and Lin (1985), Mosher, Judkins and Göksel (1989), and Lepkowski, Kalton, and Kasprzyk (1989). This general approach forms cells based upon modeled nonresponse propensity. For the modeling of nonresponse propensity, these papers have utilized software inspired by the AID (Automatic Interaction Detection) approach due to Morgan and Sonquist (1963). To oversimplify, the method forms adjustment cells in such a manner as to maximize the variation in response rates across cells subject to certain constraints.

The basis for forming nonresponse adjustment cells on the basis of predicted nonresponse propensity as modeled by an AID-type program is still intuitive. David, Little, Samuhel and Triest (1983) appeared to come close to a theoretical basis, but Little (1986) and Little and Rubin (1987) pulled back from this position. The theory assumes that there is some set of variables, x, observed on both respondents and nonrespondents such that respondent/nonrespondent status is conditionally independent of substantive variables given x. In that case, the nonresponse is

said to be *ignorable*. (For a discussion of nonignorable nonresponse, see Fay, 1989.)

Let $r=(r_1, ..., r_n)$ be the vector of nonresponse indicators for the sample and let $p_i(x_i)=P\{r_i=1|x_i\}$. David *et alii* pointed out, based upon the work of Rosenbaum and Rubin (1983) that $p(x)= (p_1(x_1), ..., p_n(x_n))$ is the coarsest vector, conditional upon which, r is independent of x.[1] Furthermore, if nonresponse is ignorable, then y and r are conditionally independent given p(x). Thus, if nonresponse is ignorable for y given x, then the partition of the dataset induced by p(x) is a fine enough set of nonresponse adjustment cells to avoid nonresponse bias. This suggests the strategy of choosing a large x matrix (since the incorporation of many attributes into x makes the assumption of ignorability more plausible), estimating p(x) by logistic regression, and stratifying the sample into cells by p(x). Three problems: 1. Current software for logistic regression can handle only a rather small number of independent variables at a time; 2 Stratification on an overspecified model for p(x) can lead to unnecessary variance in the weighted estimator; and 3. Even if p(x) were known, the conditions are unknown under which p(x) induces the coarsest partition capable of rendering nonresponse ignorable for y. To illustrate the last point, suppose that y and r are unconditionally independent but there are intricate relationships between x and r. All that is needed is a single adjustment cell, yet the response propensity approach will lead to an abundance of cells and resulting instability in the weighted estimator of y.

The first problem is the reason for recourse to AID-type software. It can handle larger numbers of independent variables and is specifically designed for exploratory analysis. The second problem is avoided in practise by placing lower limits on cell sizes and observed response rates. For example, we required a minimum cell size of around 60 and a minimum observed response rate of around 32%. Furthermore, we chose splits according to the maximum chi-square per degree of freedom. Also, we did not split a cell any further if none of the potential splitting factors passed a chi-square test for independence from nonresponse.

The third problem is the most difficult. Here, we (and prior advocates of the method) argue that if there is a large number of dependent variables, then their joint relationship with response propensity is likely to be very complex, and thus, the predictive mean approach, the natural alternative, will founder. (With the predictive mean approach, a model b(x) is formed for y instead of r.) The predictive mean approach has long been the tradition at the U. S. Bureau of the Census for demographic surveys. (See for example, U.S. Bureau of the Census, 1963 and Shapiro, 1980.) It is very attractive if there is a single variable of paramount interest as in the monthly Current Population Survey (CPS), but for surveys such as the Survey of Income and Program Participation (SIPP), supplements to the CPS, and the NSFG, there are many important dependent variables. Two cells may be very similar with respect to the expected values of one substantive variable and dissimilar with respect to another. It is not possible to say that one item is more important than all the others.

As in so many other aspects of survey research, the trade-off is between certain variance and possible bias. For NSFG, we favored lowering the risk of bias for a broad range of statistics rather than minimizing the variance. We think that this may be the correct balance for many multipurpose surveys and that the Bureau of the Census should also reconsider its traditions.

3. Potential Predictor Variables

We cast a very wide net for potential prediction variables for nonresponse propensity. We considered basic demographic variables, substantive variables from both the baseline interview and the NHIS interview, mobility since the baseline, and indicators from the baseline and the NHIS of hostility to surveys. This contrasts with tradition both at the Census and at ISR.

At the Census, the tradition is for cells to be formed on the limited basis of such variables as region, PSU, metro/nonmetro status, race of housing unit occupants, variables that can usually be determined without any cooperation on the part of designated sample persons. These types of adjustments are of course, still required for adjustment for nonresponse to the initial interview. However, there is no need to stay with these variables when adjusting for nonresponse to subsequent visits.[2] Work at ISR has included substantive variables from earlier rounds of data collection, but the first inclusion of survey hostility

[1] Any other vector with the same property of rendering nonresponse conditionally independent of x will have more categories and these categories will map onto the categories of p(x).

[2] Recent work at the Census Bureau (King, Chou, McCormick, and Petroni, 1990, as well as Singh and Petroni, 1988) indicates that future work there may move in this direction by making broader use of the many available data, but higher priority needs to be given to making the changeover. McArthur and Short (1985) showed that more variables were related to nonresponse than were being used in the nonresponse adjustment.

Table 1. Response rates, number of women and the chi-square statistics by selected variables

	R(%)	N
Mobility (c^2=522)		
Stationary	79	4,491
Mover	55	3,262
Ethnicity (c^2=390)		
Black	59	2,490
White	78	4,453
Hispanic	51	600
Other	60	210
Education (c^2=312)		
0-8	41	374
9-11	60	1,555
12	69	2,665
13-15	74	1,946
16 and over	82	1,213
Income (c^2=310)		
0-149	55	2,302
150-299	71	1,943
300-399	76	1,247
400 or more	78	2,261
Marital Status (c^2=94)		
Currently Married	74	3,987
Formerly Married	63	952
Never married	64	2,814
Age (c^2=49)		
17-19	71	596
20-24	63	1,265
25-29	66	1,438
30-34	70	1,630
35-39	73	1,536
40-44	73	1,288
Region (c^2=29)		
North East	72	1,505
South	66	3,006
Midwest	73	1,950
West	67	1,292
Metro Status (c^2=59)		
MSA, central city	63	2,277
Other MSA	72	3,817
Non-MSA	72	1,659
Number of tracking attempts (c^2=302)		
None	75	5,526
1	61	648
2-3	53	497
4-5	54	749
6 or more	49	333

	R(%)	N
Number of in-person visits (c^2=373)		
1	77	4,436
2-3	67	1,325
4-5	58	1,444
6 or more	44	548
Has Phone (c^2=324)		
Yes	73	6,749
No	45	1,004
Refused contact person's phone no (c^2=75)		
Yes	59	1,396
No	71	6,357
Refused SSN (c^2=38)		
Yes	61	927
No	71	5,716
Under 18	67	1,110
Labor Force Status (c^2=54)		
In labor force	71	5,164
Going to School	70	842
Keeping House	62	1,747
Knowledge on STDs (c^2=226)		
High	76	2,603
Moderate	69	4,225
Low	50	925
Accurate Knowledge on AIDS (c^2=81)		
Has	76	2,673
Does not have	66	5,080
Parity (c^2=74)		
None	71	2,995
1	69	1,444
2	72	1,889
3	65	911
4 or more	54	514
Contraceptive use (c^2=20)		
Effective	69	3,511
Less Effective	74	1,236
Not using	67	3,006
Ever used any birth control method (c^2=43)		
Not applicable	72	808
Yes	70	6,664
No	52	281
Ever used infertility services (c^2=18)		
Yes	75	860
No	68	6,893

variables appears to have been at Westat by Mosher, Judkins, and Göksel (1989). The present paper appears to be the first where a mobility variable was also introduced into the modeling.

A large number of variables from the NHIS, the Baseline, and the CATI Phase were used as possible predictors of response rates. Table 1 shows response rates in the CATI Phase for the most significant of these characteristics. Along with the response rates, the number of women and the chi-square statistic for independence between the variable and response propensity are also presented. All the variables are significant at .01 level.

The first variable is an indicator of mobility: whether the woman had moved since the Cycle IV Baseline or not (the nonlocatable women in the reinterview were included as having moved). This variable has the highest chi-square value. Ethnicity, education, and income (as a percent of poverty level) were included as indicators of socio-economic status. All three variables are strong predictors of response. Marital status and age were also important. Census region and metropolitan status are included as variables indicating geographical location and urbanicity.

The next two variables are from the Cycle IV Baseline: the number of tracking attempts and the number of in person visits made. These variables are also found to be strongly correlated with response propensity in the reinterview. The number of tracking attempts may be viewed as a measure of mobility, a measure of attachment to society for movers (those with stronger attachments are easier to track), and as a possible indicator of hostility toward surveys. Three NHIS variables may be considered as measures of cooperation and/or availability: whether the woman had a telephone and provided the phone number, or not; whether she provided a contact person's phone number, or not; and whether she gave her social security number, or not. Such "hostility" variables from an earlier survey were also considered by Mosher, et al (1989) and Kalton, et al (1990). They also found these variables to be extremely important determinants of nonresponse. The woman's labor force status as of the Cycle IV Baseline -- in the labor force, going to school, or keeping house -- is also included as an indicator of availability for interview.

The final set of variables presented in Table 1 are from the Baseline which are more directly related to the variables to be employed in substantive analysis. These are: knowledge of sexually transmitted diseases (the number of STDs that she has heard of), accuracy of knowledge on AIDS, parity (number of live births), current contraceptive method used, and use of infertility services. Women with 4 or more children have a lower response rate. They also had lower incomes: 58 percent were in the lowest income group, compared to 30 percent for the entire sample.

4. Importance of Mobility

Even though we believed that forming cells on the basis of nonresponse propensity would give us good protection against nonresponse bias, the mobility indicator was such a strong predictor of nonresponse that we were leery of accepting the increase in variance associated with allowing it into the model without some direct evidence that mobility was related to several of the most important substantive characteristics.

Table 2 shows that mobility is indeed related to such characteristics. The most likely movers are those with no or many children, those poorly informed about sexually transmitted diseases, and those who have had intercourse but have never used any birth control method. Women who never had intercourse and those currently using less effective methods of birth control are particularly unlikely to have moved between the two interview attempts. These are important differences. They confirmed for us the importance of allowing mobility and operational variables (such as refusal to supply a contact person's phone number) into nonresponse models instead of confining attention to standard domain indicators (such as region and metropolitan status) and substantive variables (such as parity).

Table 2. Mover rates, number of women and the chi-square statistics by selected variables

	R(%)	N
Parity (c^2=13)		
None	44	2,995
1	42	1,444
2	39	1,889
3	40	911
4 or more	44	514
Knowledge on STDs (c^2=27)		
High	42	2,603
Moderate	41	4,225
Low	50	925
Contraceptive Use (c^2=14)		
Effective	43	3,511
Less Effective	37	1,236
Not Using	43	3,006
Ever Used Any Method of Birth Control (c^2=16)		
Not applicable	37	808
Yes	42	6,664
No	49	281

5. Description of Final Cells

The final cells used in the nonresponse adjustment are shown in Table 3 for the women in the reinterview sample. The first split was on mover status. Within each initial split, race and ethnic origin were quite important. Education appeared to be more important among movers and minority nonmovers than among white nonmovers. The next splits mostly involved the hostility variables. Finally, some substantive and geographic variables entered the model. The response rates ranged from 32 percent for Hispanic movers to 95 percent for white nonmovers with prior indicators of receptivity to surveys, high knowledge of STDs, and a principle activity of keeping house.

Table 3. Nonresponse adjustment cells for reinterviewed women: NSFG Cycle IV, CATI Phase

	Cell	No. of women	Resp. rate (%)
ALL		7,753	69
Not Moved		4,491	79
Black/Hispanic/Other		1,773	72
• Education: 0-8 years	1	118	51
• Education: 9-12 years		994	70
- 1 in-person visit*		448	77
Black	2	346	80
Hispanic/other	3	102	66
- 2-5 in-person visits*		457	66
MSA central city or non-MSA	4	306	62
Other MSA	5	151	75
- 6 or more in-person visits*	6	89	55
• Education: 13 or more		661	78
- Had telephone, phone number was given in NHIS	7	600	80
- Did not have telephone or no phone number was given in NHIS	8	61	61
White		2,718	84
• Had telephone, phone number was given in NHIS		2,567	85
- 1 in-person visit*		1,889	87
Knowledge on STDs: high		794	89
In labor force or school	9	635	88
Keeping house	10	159	95
Knowledge on STDs: moderate		1,019	86
Refused contact person's phone number in NHIS	11	110	79
Provided contact person's phone number in NHIS	12	909	87
Knowledge on STDs: low	13	76	78
- 2 or more in-person visits*		678	80
Currently married**	14	424	83
Formerly or never married		254	75
Under 18 in NHIS	15	88	88
18 or over in NHIS	16	166	69
• Did not have telephone or no phone number was given in NHIS	17	151	68
Moved		3,262	55
Black/other	18	1,228	42
White		1,735	68
• Education: 0-11		375	51
- 0 tracking attempts*	19	196	61
- 1 or more tracking attempts*	20	179	40
• Education: 12 years		573	64
- 0-1 tracking attempts*		416	71
- 0-2 in-person visits*	21	353	74
- 3 or more in-person visits*	22	63	54
- 2 or more tracking attempts*	23	157	55
• Education: 13-15 years		450	73
- Refused contact person's phone number in NHIS	24	61	61
- Provided contact person's phone number in NHIS		389	75
Knowledge on STDs: high		226	80
Ever married**	25	141	85
Never Married	26	85	72
Knowledge on STDs: mod./low	27	163	69
• Education: 16 or more years		337	83
- 0-1 in-person visits*	28	229	88
- 2 or more in-person visits*	29	108	71
Hispanic	30	299	32

* No, of in-person or tracking visits to get baseline interview
**Includes currently informally married

6. Impact on Variance

Despite the fact that we avoided forming adjustment cells that would have resulted in adjustment factors larger than 3.13, the nonresponse adjustment resulted in considerable increases in the variability of the weights. Table 4 below shows the relative variance in the weights at each stage of the adjustment.

Table 4. Relative variance in weights by adjustment stage

Stage	Blacks	Others	All races
Baseweight	.49	.12	.28
Nonresponse Adjusted	.61	.26	.40
Trimmed	.55	.22	.35
Post-stratified (Final)	.53	.24	.39

The general theory holds that each point added to the relative variance in weights adds a point to the design effect. A trimming stage was introduced to counter the few situations where a respondent with a large baseweight happened to be in a cell with a low response rate. Even so, nonresponse adjustment appears to have added roughly 10 points to the relative variance in weights for nonblacks and for women of all races. These are nontrivial increases since they indicate that one tenth of the variable budget is being devoted to the reduction of the risk of bias.

However, when we calculated actual variances using balanced repeated replications and generalized a large number of items (these terms are used in the classical sense standardized in Wolter, 1984), the design effects for the CATI phase of Cycle IV were actually smaller than those for the baseline. These design effects are shown in Table 5. Some of this is due to a fall in average cluster sizes. Some is due to an additional stage of post-stratification to Hispanic controls. Some of it may be due to an absence of the procedure used in the baseline to subsample initial nonrespondents for intensive conversion attempts (conversion to respondents). Nonetheless, it is clear that the very aggressive nonresponse adjustment for the CATI Phase did not have major adverse effects on variances.

Table 5. Design effects by race and phase

	Cycle IV Baseline	Cycle IV CATI
All races	1.57	1.28
Blacks	1.90	1.46

References

David, M.H., Little, R.J.A., Samuhel, M.E., and Triest, R.K. (1983). Imputation methods based on the propensity to respond. *Proceedings of the Section on Business and Economics, American Statistical Association,* 168-173.

Fay, R.E. (1989). Estimating nonignorable nonresponse in longitudinal surveys through causal modeling. In *Panel surveys* (D. Kasprzyk , G. Duncan, G. Kalton, and M. P. Singh, eds). New York: John Wiley.

Institute for Social Research (1979). *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1978 Interviewing Year, Wave XI, A Supplement.* Ann Arbor: University of Michigan.

Institute for Social Research (1986). *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1984 Interviewing Year, Wave XVII, A Supplement.* Ann Arbor: University of Michigan.

Judkins, D., Mosher, W., and Botman, S. (1991). National Survey of Family Growth: Sample design, estimation, and Inference. *Vital and Health Statistics, Series 2,* No 109, National Center for Health Statistics.

Kalton, G., Lepkowski, J., and Lin, T.-K. (1985). Compensating for Wave Nonresponse in the 1979 ISDP Research Panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 372-377.

Kalton, G., Lepkowski, J., Montanari, G.E., Maligalig, D. (1990). Characteristics of second wave nonrespondents in a panel survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 462-467.

Kasprzyk, D., Duncan, G., Kalton, G., and Singh, M.P. (1989). *Panel Surveys.* New York: John Wiley.

King, K., Chou, S., McCormick, M., and Petroni, R. (1990). Investigations of the SIPP's cross-sectional noninterview adjustment method and variables. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 576-581.

Lepkowski, J. (1989). The treatment of wave nonresponse in panel surveys. In *Panel surveys* (D. Kasprzyk , G. Duncan, G. Kalton, and M. P. Singh, eds). New York: John Wiley.

Lepkowski, J., Kalton, G., and Kasprzyk, D. (1989). Weighting adjustments for partial nonresponse in the 1984 SIPP Panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 296-301.

Little, R.J.A. (1986). Survey nonresponse adjustments. *International Statistical Review* **54**, 139-157.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley.

McArthur, E. and Short, K. (1985). Characteristics of sample attrition in the Survey of Income and Program Participation. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 366-371.

Morgan, J.N. and Sonquist, J.A. (1963). Problems in the analysis of survey data and a proposal. *Journal of the American Statistical Association,* **58**, 415-435.

Mosher, W.D., Judkins, D., and Göksel, H. (1989). Response rates and nonresponse adjustment in a national survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 273-278.

Rosenbaum, P.R. and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.

Shapiro, G.M. (1980). A General Approach to Noninterview Adjustment. U.S. Bureau of the Census internal memorandum, dated March 11, 1980, within Statistical Methods Division.

Singh, R. and Petroni, R. (1988). Nonresponse adjustment methods for demographic surveys at the U.S. Bureau of the Census. SIPP Working Paper Series, No. 8823. Washington D.C.: U.S. Department of Commerce, Bureau of the Census.

Sonquist. J. A., Baker, E.L. and Morgan, J.N. (1973). *Searching for Structure.* Ann Arbor, MI: Institute for Social Research.

U. S. Bureau of the Census (1963). *The Current Population Survey - - A report on Methodology,* Technical Paper No. 7. Washington, D. C.: U. S. Government Printing Office.

Wolter, K.M., (1984). *Introduction to Variance Estimation.* New York: Springer-Verlag.