

Nonresponse Adjustments for a Telephone Follow-up to a National In-Person Survey

Hüseyin Göksel¹, David R. Judkins¹, and William D. Mosher²

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 an earlier article. 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 socioeconomic variables and several variables specific to the subject matter of NSFG were also closely associated with nonresponse. The article describes how the determinants of nonresponse were modeled, how the sample weights were adjusted for nonresponse, and presents the results of an evaluation.

Key words: Response propensity; Automatic Interaction Detection (AID).

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 article, 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 Computer-Assisted Telephone Interview (CATI) Phase of Cycle IV.

¹ Westat, Inc., 1650 Research Blvd., Rockville, MD 20850, U.S.A.

² National Center for Health Statistics, 6525 Belcrest Rd, Hyattsville, MD 20782, U.S.A.

Acknowledgement: The 1988 and 1990 National Surveys of Family Growth were jointly funded by the National Center for Health Statistics, the National Institute for Child Health and Human Development and the Office of Population Affairs. The opinions expressed are those of the authors, and not necessarily those of the above agencies. The computer assistance of David P. Johnson of NCHS is gratefully acknowledged.

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 data base 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 article, 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 Study of Income Dynamics (PSID), Kalton, Lepkowski, and Lin (1985), Mosh-

er, 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 or by logistic regression) is still intuitive. David, Little, Samuhel, and Triest (1983) came close to claiming a theoretical rationale for adjustment on the basis of predicted nonresponse propensity, but Little (1986) and Little and Rubin (1987) used weaker language to justify the technique. The available 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*. (Whether nonresponse is ignorable for a particular substantive variable y given x is, of course, impossible to establish since y is unobserved for nonrespondents. It must be accepted on faith. For a discussion of nonignorable nonresponse, see Fay (1989).)

Let $r = (r_1, \dots, r_n)$ be the vector of nonresponse indicators for the sample, y be the substantive variable of interest, and $p_i(x_i) = P\{r_i = 1|x_i\}$, where the individuals in the sample are indexed by $i = 1, \dots, n$. Then David, Little, Samuhel, and Triest pointed out, based upon the work of Rosenbaum and Rubin (1983), that $p(x) = (p_1(x_1), \dots, p_n(x_n))$ is the coarsest vector, conditional upon which, r is in-

dependent of x .³ 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 data set 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 $p(x)$. There are three problems with this line of reasoning: (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. (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 a particular 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 large numbers of independent variables and is specifically designed for exploratory analysis. The second problem is avoided in practice by setting constraints on minimum cell sizes and observed response rates. For example, we required a minimum cell size of around 60 and a minimum observed res-

ponse rate of around 32%. (The minimum cell size of 60 was set so that each half-sample used in variance estimation would have an expected 30 observations, the traditional point for claiming an acceptable central limit theorem approximation. The minimum observed response rate was initially set to roughly one-half of the overall response rate on the grounds that we did not want anyone's adjustment to be much more than double the average adjustment. The limit of doubling is based upon institutional traditions.) Furthermore, we chose splits according to the maximum chi-square per degree of freedom. Also, we required fairly strong evidence that a variable was related (within the cell) to nonresponse before we allowed it to be used as a splitting factor. The standard chi-square test had to reject the null hypothesis of independence (between the variable and nonresponse) at the 5% level. If we could not find any such variable for a cell, then splitting of the cell ceased.

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

³ 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 $p(x)$.

one substantive variable and dissimilar with respect to another. It is not possible to say that one variable 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. The range of variables considered is broader than traditional at either the Census Bureau or ISR.

At the Census Bureau, 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 restrict attention to these variables when adjusting for nonresponse to subsequent visits. McArthur and Short (1985) showed that more variables were related to nonresponse than were being used in the nonresponse adjustment by the Census Bureau for the Survey of Income and Program Participation. Recent work at the Census Bureau indicates that future adjustments may make

broader use of the available data – especially mobility indicators (King, Chou, McCormick, and Petroni 1990; Singh and Petroni 1988). We encourage that movement.

Work at ISR has included substantive variables from earlier rounds of data collection, but the first inclusion of survey hostility variables appears to have been at Westat by Mosher, Judkins, and Göksel (1989). Mobility does not appear to have been included in the modeling for nonresponse propensity at ISR.

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 the .01 level. (The significance level is conditioned on the observed sample and is only nominal. No corrections were made for the sample design or for multiple comparisons.)

The first variable is an indicator of mobility: whether the woman had moved since the Cycle IV Baseline or not. This variable has the highest chi-square value. On the cautionary side, it was much easier to determine mover status for responding women than for nonresponding women. Out of the nonrespondents that we classified as movers, about half were traced to a new address but not interviewed for various reasons; another third were known by the post office to have moved but could not be located; interviewers classified the remaining sixth as “not located” on the basis of data which were not available to us. Nonetheless, it seemed reasonable to assume that most of these women whom we could not find two years after an in-person interview must have

moved even if the post office could not verify that fact. Looking at possible errors in the other direction, it is possible that some of the nonrespondents that we classified as nonmovers could have made local moves that did not require phone number changes. The post office would have informed us about many but not all such moves.

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. Hispanics have the lowest response rate followed by blacks. The response rate is markedly lower for women with household income under 150% of the poverty level income. There is also a strong positive correlation between education and response propensity. Kalton, Lepkowski, Montanari, and Maligalig (1990), reporting on a panel survey about health implications of life styles, also found lower reinterview rates for blacks and Hispanics, those with fewer years of education and those with lower family income in the original interview.

The next two variables in Table 1 are marital status and age of the woman (at the reinterview). Currently married women have higher response rates than formerly married and never married women. The response rates among women 20 to 29 years old are lower. This is broadly consistent with the finding by McArthur and Short (1985) that SIPP reinterview response rates were lower for persons between 15 and 24 years old.

Census region and metropolitan status are included as variables indicating geographical location and urbanicity. The response rates are lower in the South and West than Northeast and Midwest. In central cities the rates are also lower than in the suburbs and nonmetropolitan areas. Kalton et al. (1990) also found lower rates in central cities than rural areas in the reinterview

but they did not find significant variation among response rates between the regions.

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 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. Here it is interesting that those whom we would have expected to have the most time available for participation (i.e., those keeping house) had the lowest response rate. However, this effect either disappears or is reversed once we control for other variables as discussed in Section 5.

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 which she has heard of), accuracy of knowledge on AIDS, parity (the number of live births the woman has had),

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

	<i>R</i> (%)	<i>N</i>
Mobility ($\chi^2 = 522$)		
Stationary	79	4,491
Mover	55	3,262
Ethnicity ($\chi^2 = 390$)		
Black	59	2,490
White	78	4,453
Hispanic	51	600
Other	60	210
Education ($\chi^2 = 312$)		
0–8 years	41	374
9–11 years	60	1,555
12 years	69	2,665
13–15 years	74	1,946
16 or more years	82	1,213
Income ¹ ($\chi^2 = 310$)		
0–149	55	2,302
150–299	71	1,943
300–399	76	1,247
400 or more	78	2,261
Marital Status ($\chi^2 = 94$)		
Currently married	74	3,987
Formerly married	63	952
Never married	64	2,814
Age ($\chi^2 = 49$)		
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 ($\chi^2 = 29$)		
North East	72	1,505
South	66	3,006
Midwest	73	1,950
West	67	1,292
Metropolitan Status ($\chi^2 = 59$)		
MSA ² , central city	63	2,277
Other MSA	72	3,817
Non-MSA	72	1,659
Number of tracking attempts ($\chi^2 = 302$)		
None	75	5,526
1	61	648
2–3	53	497
4–5	54	749
6 or more	49	333

¹Income is shown as a percent of the poverty level.²MSA stands for Metropolitan Statistical Area.

Table 1. (Continued)

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

³SSN stands for Social Security Number.⁴STD stands for Sexually Transmitted Disease.⁵"Effective contraception" stands for use of sterilization, the pill, or an IUD.

current contraceptive method used (if any), and use of infertility services. Women with four or more children have a lower response rate. They also have lower incomes: 58% are in the lowest income group, compared to 30% 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 non-response 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. (We examined the marginal relationship of mobility with the substantive variables, rather than conditional relationship given the socio-economic and survey-hostility-marker variables that we were working with, mainly because of ease of implementation. As one of our referees pointed out, the conditional relationship is of greater interest.)

Table 2 shows that mobility is indeed related to such characteristics. The most likely movers are at the extremes of the parity distribution (there appears to be a quadratic effect), those poorly informed about sexually transmitted diseases and AIDS, and those who have had intercourse but have never used any birth control method. Women who never had intercourse (the inapplicable line to lifetime birth control usage), those currently using less effective methods of birth control, and those having utilized infertility services are markedly less likely to have moved between the two interview attempts. These are important differences. They confirmed for us the importance of at least considering operational variables (such as refusal to supply a contact person's

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

	R(%)	N
Parity ($\chi^2 = 13$)		
None	44	2,995
1	42	1,444
2	39	1,889
3	40	911
4 or more	44	514
Knowledge on STDs ($\chi^2 = 27$)		
High	42	2,603
Moderate	41	4,225
Low	50	925
Accurate knowledge on AIDS ($\chi^2 = 11$)		
Has	40	2,673
Does not have	43	5,080
Contraceptive use ($\chi^2 = 14$)		
Effective	43	3,511
Less effective	37	1,236
Not using	43	3,006
Ever used any method of birth control ($\chi^2 = 16$)		
Not applicable	37	808
Yes	42	6,664
No	49	281
Ever used infertility services ($\chi^2 = 5.8$)		
Yes	38	860
No	43	6,893

phone number) when modeling nonresponse instead of confining attention to standard domain indicators (such as region and metropolitan status) and substantive variables (such as parity).

5. Description of Final Cells

The cells used in the nonresponse adjustment are shown in Table 3 for the 7,753 women in the reinterview sample. The first split was on mover status. Within each 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% for Hispanic movers to 95% for white nonmovers with prior indicators of receptivity to surveys, high knowledge of STDs (sexually transmitted diseases), and a principal activity of keeping house.

The final cells are not shown for the 609 teenagers eligible for a first interview in 1990. The only variable that turned out to be important for them was mover status. Among those who had moved since the NHIS interview we obtained a response rate of 35%, well below the 67% for the balance of the sample.

6. Effect on Variance

Despite the fact that we avoided forming adjustment cells that would have resulted in adjustment factors larger than 3.13 (note that 3.13 is the reciprocal of 0.32 which was the minimum allowed response rate), the nonresponse adjustment resulted in considerable increases in the variability of the weights. Table 4 shows the relative variance (square of the coefficient of variation) in the weights at each stage of the adjustment.

General theory going back to Kish (1967, eq. 11.7.6) holds that each point added to the relative variance in weights by disproportionate allocation of sample to strata adds a point to the design effect (under the assumption of uniform within-strata unit variances and per-unit costs). Ratio adjustments can have similar effects although the situation is clearly more complicated. We therefore use the relative variance in the weights as a rough approximation (probably pessimistic) of the effect of weighting adjustments on variances for substantive

statistics. 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 possibly as much as one-tenth of the variable budget is being devoted to the reduction of the risk of bias. In other words, the survey precision is no better than would have been achieved with a survey that had 10% smaller sample size but less variation in weights.

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 1985), 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.

7. Effect on Bias

Any evaluation of nonresponse bias for the reinterview is speculative, but we did examine the effect of the adjustment on estimates of 1988 characteristics that can be formed from reinterview respondents. The results are displayed in Table 6. The first column of numbers consists of 1988 estimates from all women who responded in 1988 and were

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

	Cell	Number of women	Response rate (%)
All			
Not moved		7,753	69
Black/Hispanic/Other		4,491	79
Education: 0-8 years		1,773	72
Education: 9-12 years	1	118	51
Number of in-person visits to get Baseline: 1		994	70
Black	2	448	77
Hispanic/Other		346	80
Number of in-person visits to get Baseline: 2-5	3	102	66
MSA central city or non-MSA		457	66
Other MSA	4	306	62
Number of in-person visits to get Baseline: 6 or more	5	151	75
Education: 13 or more years	6	89	55
Had telephone, phone number was given in NHIS		661	78
Did not have telephone or no phone number was given in NHIS	7	600	80
White	8	61	61
Had telephone, phone number was given in NHIS		2,718	84
Number of in-person visits to get Baseline: 1		2,567	85
Knowledge on STDs: high		1,889	87
In labor force or going to school		794	89
Keeping house	9	635	88
Knowledge on STDs: moderate	10	159	95
Refused contact person's phone number in NHIS		1,019	86
Provided contact person's phone number in NHIS	11	110	79
Knowledge on STDs: low	12	909	87
	13	76	78

Table 3. (Continued)

	Cell	Number of women	Response of rate (%)
Number of in-person visits to get Baseline: 2 or more			
Currently married*	14	678	80
Formerly married or never married		424	83
Under 18 years in NHIS	15	254	75
18 years or over in NHIS	16	88	88
Did not have telephone or no phone number was given in NHIS	17	166	69
Moved		151	68
Black/other		3,262	55
White	18	1,228	42
Education: 0–11 years		1,735	68
Number of tracking attempts for Baseline: 0	19	375	51
Number of tracking attempts for Baseline: 1 or more	20	196	61
Education: 12 years		179	40
Number of tracking attempts for Baseline: 0–1		573	64
Number of in-person visits to get Baseline: 0–2	21	416	71
Number of in-person visits to get Baseline: 3 or more	22	353	74
Number of tracking attempts for Baseline: 2 or more	23	63	54
Education: 13–15 years		157	55
Refused contact person's phone number in NHIS	24	450	73
Provided contact person's phone number in NHIS		61	61
Knowledge on STDs: high		389	75
Ever married*	25	226	80
Never married	26	141	85
Knowledge on STDs: moderate/low	27	85	72
Education: 16 or more years		163	69
Number of in-person visits to get Baseline: 0–1	28	337	83
Number of in-person visits to get Baseline: 2 or more	29	229	88
Hispanic	30	108	71
		299	32

*Includes currently informally married.

Table 4. Relative variance in weights by adjustment stage

Stage	Blacks	Nonblacks	All races
Baseweight	0.49	0.12	0.28
Nonresponse adjusted	0.61	0.26	0.40
Trimmed	0.55	0.22	0.35
Post-stratified (final)	0.53	0.24	0.39

then between the ages of 20 and 39. These are the best estimates available for 1988. The next two columns consist of estimates based only on women who responded in both 1988 and 1990. In the first of these (column 2), 1990 respondents were weighted by inverse probability of selection and adjustment for nonresponse to the NHIS and to the 1988 survey. There was no adjustment for nonresponse on the reinterview. Nor was there any post-stratification. The second column of 1988 estimates based upon 1990 respondents (column 3) was formed by using the final reinterview weights. These weights included the nonresponse adjustment described in Section 5 and a post-stratification to special estimates from the Current Population Survey.

Those estimates that were improved by the adjustments are marked with asterisks. The change in the estimate due to the adjustment is expressed as a percentage of the full 1988 estimate. A negative change indicates that the error in the estimate was reduced by the nonresponse adjustment. In the last column, the cv (coefficient of variation, i.e., standard error of the sample estimate for the statistic expressed as a percentage of the statistic) for each statistic is given using

the design effect and sample size of the reinterview.

The displayed statistics were selected on the basis of their importance to the sponsor and perceived sensitivity. The statistics were selected prior to examining differences due to adjustment. Note that none of these substantive variables were included in the model.

We have not done formal tests on the differences because it is not clear how to account for the strong correlation between the estimates. Nonetheless, it is obvious that the adjustments were extremely successful in reducing the bias in estimates of totals. Of course, a less complex approach would probably have captured most of the same improvement. The real test is for estimated proportions – here the results are less clear. Most of the differences between the 1988 estimated proportions and both of the 1990 estimates were small; only a few of the differences dominate the mean squared error. The adjustments helped estimated proportions more often than they hurt for this specific set of statistics and the average improvement was larger than the average worsening. For this set of statistics we thus conclude that the adjustments had a modestly beneficial effect.

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

8. Recommendations

We think that the type of adjustment applied here to NSFSG is a good method for creating multipurpose nonresponse-adjusted weights for demographic surveys with

Table 6. Evaluation of adjustment for selected statistics

	1988	1988 based on 1990 respondents w/o adjustment	1988 based on 1990 respondents and adjusted weights	Change in error as percent of 1988 estimate	CV for 1990 sample
Fecundity Status					
Contraceptively sterile	23.6%	25.2%	23.8% *	− 5.9%	3.2%
Sterile for other reasons	3.8%	4.0%	3.9% *	− 2.6%	8.9%
Impaired	8.9%	8.7%	8.7% *	− 0.1%	5.8%
Fecund	63.7%	62.1%	63.6% *	− 2.4%	1.4%
Population (000's)	40,722	24,761	40,799		
Contraceptive Choice					
Female sterilization	25.8%	24.7%	24.6%	0.2%	3.8%
Male sterilization	10.8%	12.7%	11.5% *	− 11.3%	6.1%
Pill	33.1%	31.7%	33.6% *	− 2.8%	3.1%
IUD	1.9%	1.6%	1.6%	3.2%	17.3%
Diaphragm	6.6%	7.2%	6.9% *	− 5.2%	8.0%
Condom	13.4%	13.2%	12.6%	4.1%	5.8%
Other	8.4%	8.9%	9.2%	3.2%	6.9%
Population (000's)	26,680	16,947	27,285		
Douches					
All education levels	37.3%	34.9%	37.5% *	− 5.9%	2.4%
Population (000's)	38,353	23,602	38,719		
0–11 years education	56.5%	59.9%	60.3%	0.7%	4.2%
Population (000's)	5,604	2,435	4,975		
12 years education	46.5%	45.7%	47.3% *	0.0%	3.2%
Population (000's)	13,924	8,830	14,566		
13+ years education	24.8%	22.2%	24.0% *	− 7.3%	4.7%
Population (000's)	18,825	12,338	19,178		
Ever treated for PID (Pelvic Inflammatory Disease) – All levels of lifetime sexual partners					
	12.7%	12.2%	12.9% *	− 2.4%	4.8%
Population (000's)	38,436	23,650	38,043		
1–3 partners	8.5%	8.7%	9.3%	7.1%	7.8%
Population (000's)	21,037	13,179	20,945		
4–9 partners	15.8%	15.0%	15.4% *	− 2.5%	8.0%
Population (000's)	10,905	6,532	11,123		
10 or more partners	21.1%	19.5%	20.7% *	− 5.7%	9.2%
Population (000's)	5,530	3,482	5,976		
# Partners unknown	18.8%	19.6%	19.4% *	− 1.1%	26.8%
Population (000's)	963	457	758		

*Estimate improved by nonresponse adjustment and post-stratification.

longitudinal components. The multi-purpose aspect is very important for most public use files. Nevertheless, the multi-purpose adjustment leads to a greater variance in weights than would be needed for some analyses. There should probably be a section in the manual for the public use tape where users are instructed how to create adjustment cells for a particular substantive variable. Perhaps a SAS software file could also be included to repeat the post-stratification for a new set of nonresponse-adjusted weights.

9. 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., and 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.
- 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.
- Kish, L. (1967). Survey Sampling. 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.
- 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. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.

Received September 1991
Revised May 1992