NONRESPONSE AND ITS EFFECTS IN A FOLLOWUP TELEPHONE SURVEY OF LOW INCOME WOMEN

William D. Kalsbeek, Todd A. Durham, Survey Research Unit
William D. Kalsbeek, Survey Research Unit, 730 Airport Road, Suite 107, Chapel Hill, NC 27599-2400

Key Words: Nonresponse bias, panel survey, cost-efficiency, survey design

1. INTRODUCTION

The bias implications of unit and item nonresponse have been long recognized and are conceptually alike (e.g., Hansen and Hurwitz, 1946; Politz and Simmons, 1949), but research continues on the effects of each level and ways to measure and deal with the problem at each level. One area of needed research is in assessing the biasing effect of nonresponse in longitudinal studies, particular for those studies dealing with populations with unique survey response patterns (e.g., persons with low income). Findings presented here represent the early stages of research aimed at addressing these two needs within a specific survey environment. We build from conceptual frameworks discussed by Kish (1965) and Groves (1989) to develop bias measures that enable us to assess the following for longitudinal studies involving cohort samples:

(1) Pattern of response rates during baseline and followup phases of the study;
(2) Total population and population subgroup estimates of total nonresponse bias due to all sources of attrition; and
(3) Total population estimates of components of nonresponse bias due to various classifications of the bias sources defined by reasons for nonresponse at individual rounds of data collection and by round of data collection.

Findings for this study were obtained from call disposition and substantive data for a recent telephone followup survey on breastfeeding of about 2,100 Women, Infants, and Children (WIC) program participants in North Carolina. The list sample for this survey was selected from an available WIC program administrative file whose contents include demographic, medical, and breastfeeding information.

2. METHODS

Direct assessment of nonresponse bias in sample surveys requires substantive survey data for both respondents and nonrespondents. Multiple rounds of data gathering in a longitudinal study make it possible for some substantive data from the early rounds to be available for a portion of the nonrespondents at the later rounds.

2.1 1990-92 WIC Breastfeeding Followup Survey

A followup telephone survey on breastfeeding behavior by North Carolina participants in the WIC program was undertaken by the University of North Carolina at Chapel Hill Survey Research Unit under contract to the North Carolina Department of Environment, Health and Natural Resources between July 1990 and March 1992. A stratified simple random sample of participants due to give birth in late 1990 was selected and contacted three times: first; just prior to the scheduled birth of the child (baseline); second, approximately three months postpartum (first followup); and finally, approximately one year postpartum (second followup). All reasonable efforts were made in the two followup rounds to contact for interview all respondents from the prior round.

Baseline Sample Design --- A proportionate stratified list sample (Kish, 1965, Section 3.4) was selected for the baseline round of the WIC study. Using the North Carolina WIC certification file as the sampling frame, a sample of 2,148 women was chosen from those who at the time of sampling were shown to be pregnant, eighteen years or older, had indicated telephone numbers where they say they could be reached, and had a due date between October 11, 1990 and January 3, 1991. There were 5,377 eligible enrollees from which the sample was selected. Findings presented below exclude our experience with those less than 18 years of age, since for them initial contact and the baseline interview had to be done by WIC program staff, rather than SRU interviewers.

Baseline Survey Data Collection --- Design specification for this study followed diligent but fairly standard procedures for locating and soliciting participation from sample members. Because of the nature of the survey, however, only female SRU interviewers were used. Interviewers were specially trained in ways to improve location and solicitation rates. During the pretest it was observed that many of the respondent telephone numbers on the WIC master file belonged to the neighbors or relatives or were no longer in service. Much of the interviewer training was therefore devoted to handling these situations. Additionally, a small group of interviewers with the best solicitation rates was used for converting initial refusals and other reluctant members of the sample. One other common feature of the location and solicitation plan for this survey was the use of an

advance mailing designed to arrive just before the start of data collection.

Interviewing sessions were conducted seven days a week, during afternoon and evening hours. Callbacks were automatically scheduled by the CATI system for "busy signals," "no answers," and "answering machines." Most numbers were called up to six times before being dropped from the study. Some difficult-to-schedule and affirmative-but-chronically-unavailable numbers were called more than six times.

Followup Survey Data Collection Protocol --- Much of the calling protocol in the followup surveys was essentially the same as in the baseline. Efforts to locate and obtain a response during the two followup rounds differed from the baseline in two ways. First, call scheduling was done manually during followup using a "paper and pencil" method to record dispositions, keep a call history and make next-call decisions. Secondly, the intensity of location and solicitation efforts was substantially greater during the two followup rounds than during baseline. Three methods were used, separately or in conjunction, during the followup rounds to interview potential respondents: confirming the existing telephone information or obtaining a new telephone number through friends/relatives and/or directory assistance, allowing more call attempts to sample members, and enlisting the help of WIC staff in the local health department to locate the woman.

2.2 Estimation of Nonresponse Indicators

Bias Components --- Application of the estimation methodology discussed below was to data tied to the attrition at each of three data gathering rounds of the WIC Followup Survey. A commonly used model for nonresponse bias associated with linear statistics (e.g., Kish, 1965; Cochran, 1977; Groves 1989; and Lessler and Kalsbeek, 1992), along with an extension to the case of multiple components, will serve as the starting point for our work. For estimating a population proportion (P), this model views bias as arising out of the difference between the expected value of the estimate (p_r) based on respondent data only (i.e., $E(p_r)=P_r$) and P, which can be expressed as,

$$\text{Bias}(p_r) = P_r - P = W_{nr}[P_r - P_{nr}], \qquad (1)$$

the product of the rate of nonresponse (W_{nr}) and difference between P_r and the proportion among nonrespondents (P_{nr}). Thus, for example, nonresponse bias for the rate of intent to breastfeed can be estimated by finding suitable direct estimates of P_r and P (also W_{nr} and P_{nr}, if assessment of the roles of response rates and respondent-nonrespondent differences is sought). When data from a survey with a complex sampling design are used, simple weighted estimates (accounting for differential sampling probabilities) and associated variances can be obtained following the same fundamental estimation concepts as existing computer software (e.g., SUDAAN; see Shah, 1991) which utilizes a Taylor series approximation for error deviations as the approach to variance estimation.

When interested in isolating components of nonresponse bias attributable to specific reasons for nonresponse (i.e., attrition at a particular round or for specific reasons during a given round), Groves (1989, Section 4.5.3) suggests a model based on the total bias expression above, which again can be used as a starting point. If

$$B_i = W_{nr,i}[P_r - P_{nr,i}] \qquad (2)$$

is the component of bias due to some reason (i) for nonresponse, where $W_{nr,i}$ is the proportion of the population who would fail to respond for reason (i), and $P_{nr,i}$ is the proportion applicable to nonrespondents for reason (i), then the total bias is the sum of component biases among all I reasons; i.e.,

$$B = \sum_{i}^{I} B_i \qquad (3)$$

3. FINDINGS

The findings in Table A reveal somewhat differing patterns of response outcomes in each round of the WIC study. The five general response categories follow those suggested by Lessler and Kalsbeek (1992). Eligibility in each of the three rounds was determined by pregnancy status of the WIC program participant and (in the two followup rounds) by the post-partum status of the child. The relatively low baseline response rate (61.7%) was due mainly to difficulty in locating sample members. The other noteworthy finding from the baseline was the low (0.7%) refusal rate, due perhaps to the strong support of the WIC program in our solicitation efforts and the familiarity of WIC participants with filling out forms and completing interviews. Round-specific response rates for the two followups were each approximately 80 percent, with differences due mainly to variation in the percent of WIC participants not solicited. The greater difficulty in trying to locate women for the second followup is probably due to the longer period between the two followups (generally 9-11 months) than between the baseline and the first followup (mostly 4-6 months). Similar to the baseline round, low refusal rates were observed for both the first and second followups (0.7% and 0.9%, respectively).

Subgroup-specific response rates for the WIC study are presented in Table B, along with findings from logistic regression with response to second followup as the dependent variable. Subgroups in this

table were defined by participant response to questions asked of them at the time they enrolled in the WIC program (i.e., earlier in their pregnancy than the data for the baseline round). Cumulative response rates in Table B suggest that rural participants (43.5% response rate) are more likely to respond than their urban counterparts and that women 30 and older (44.4%) are more likely to respond than women in the younger age groups. Interestingly, the middle age group, the 25-29 year olds had a lower cumulative response rate (34.4%) than the youngest age group, the 18-24 year olds (38.9%). Findings suggest a lower response rate for program participants who were migrant workers at the time of enrollment in the program, although the number of migrants in the sample was small. Interestingly, there was little difference in response rates by race, indicating perhaps that racial differences in response propensity diminish when controlled by socio-economic status (i.e., low for WIC program participants). Subgroup response rates are higher for the two followup rounds, generally ranging between 75% and 85%, with the same demographic imbalances as seen in the baseline. Nonresponse in the last two rounds of data collection thus exacerbates the demographic imbalance problems noted in the baseline and amplifies the bias potential with each successive round with the compounding of nonresponse.

In a logistic model of the probability of responding to all three rounds of data collection, race, population density, age and migrant status were tested as potential predictors of response using SAS Proc Logistic (*SAS/STAT* ®, 1989). Results of this test indicate that only population density ($p<.0001$) and age, in three categories, ($p<.02$) are significant predictors of response. A similar logistic regression analysis was carried out for survey response subsequent to the baseline round. For this analysis, another independent variable was added to the four outlined above, a dichotomous variable indicating whether or not the participant's pre-baseline and baseline intentions were consistent. This consistency measure was not found to be a significant predictor of response. Results for the other independent variables in this second model were the same as the first model.

For pre-baseline, baseline and first followup measures, estimates of total bias and bias components by round and by reason for nonresponse are presented in Table C. The most important message from these findings is the smallness (in absolute size) of total nonresponse bias estimates for the three key substantive breastfeeding measures. Total bias estimates of -0.23 %, -0.08 %, and 0.81 % for the pre-baseline, baseline, and first followup measures are quite small relative to the percentage of respondents with the corresponding attribute (63.2%, 63.4%, and 73.1%, respectively). For the pre-baseline measure of intent not to breastfeed, the bias component largest in absolute magnitude is attributable to nonresponse for all reasons combined at baseline (-0.27%). The smallest (in absolute magnitude) for the pre-baseline measure is attributable to nonresponse for all reasons combined at the first followup (0.15%) These results are consistent with the observed response rates for the baseline and first followup rounds of data collection (61.7% and 82.5%, respectively). For the baseline measure of intent not to breastfeed, the largest bias component in absolute magnitude is associated with nonresponse at the first followup (-0.14 %). In a somewhat different pattern than that observed for the pre-baseline measure, the largest bias components in absolute size for all rounds combined are attributable to nonresponse due to participants not being solicited (0.29%) and being ineligible (-0.19). Similarly, the predominant bias component for the first followup measure is associated with survey participants not being solicited (0.66%), with only a minimal contribution from those unwilling to participate (-0.04%).

Total nonresponse bias estimates for population subgroups, presented in Table D, are somewhat larger than estimates obtained for the entire population, although as with the corresponding total population bias measures, they are relatively small in magnitude. Subgroup estimates (in absolute size) range from 0.11% to -2.13%. Bias estimates for the pre-baseline measure do not indicate any racial differences, however population density and age subgroups exhibit some variation. The urban subgroup has an estimate of total bias of -1.21%, while the rural subgroup has an estimate of 0.11%. Within the age subgroups, a pattern of positive total bias for the youngest and oldest age groups and negative bias for the middle age group is indicated. This pattern persists for all three key measures. Compared to the pre-baseline measure, racial differences in total bias are more pronounced for the baseline measure of intent (-1.21% for whites and 1.07% for non-whites), and for the first followup measure (0.25% for whites and 1.26% for non-whites).

4. DISCUSSION

Over three rounds of data collection in the 1990-1992 WIC Breastfeeding Survey, a cumulative response rate of 38.8% was observed. Surprisingly, estimates of total nonresponse bias were relatively small, even for other measures, i.e., attitudes and opinions about breastfeeding, not presented in this text. We may be able to attribute the size of these biases to two sources: a small number of refusals at each round

and for all rounds combined, and to the relative homogeneity of the population under study, primarily disadvantaged women. Few patterns of nonresponse were observed, either with respect to direction of bias or reason for nonresponse, probably due to the smallness of bias estimates, not deviating far from zero. An investigation of potential predictors of response yielded two significant predictors of response propensity: age and degree of urbanization. There may be some indication of an age effect with respect to nonresponse bias, suggesting possibilities for future research. The youngest and oldest age groups had higher cumulative response rates and positive total nonresponse bias estimates for all three measures and the middle age group had the lowest cumulative response rate and negative total nonresponse bias estimates for the three measures. A pattern relating response and nonresponse bias is not as evident for the population density subgroup, though it was shown to be highly predictive ($p<.0001$) of response propensity.

The findings from this analysis of nonresponse bias have several implications. Perhaps the most important is the interaction of survey content and the decision to participate or not and the role this interaction has on nonresponse bias. We might expect that an individual's decision to not participate in a survey would have some relationship with his or her opinions or attitudes about the subject matter of the survey. Under this hypothesis, minimizing the rate of those who refuse or who are chronically inaccessible would limit the potential nonresponse bias. Additionally, the findings in this research may indicate that intensive data collection efforts, i.e., multiple call attempts, subject tracking, and refusal conversions, are worthwhile in minimizing nonresponse bias. It is also possible that in populations with unique survey response patterns, as in our case with typically lower response, a higher level of homogeneity exists, minimizing the degree of aggregate difference between respondents and nonrespondents and thus on the amount of nonresponse bias. Future research could include a study of the relationship between substantive survey measures and response probabilities (or other proxy measures), and the cost of steps to minimize nonresponse to the changes in nonresponse bias that they effect.

REFERENCES

Burgess, R.D. (1989). Major Issues and Implications of Tracing Survey Respondents. In *Panel Surveys*, eds., D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh. New York: John Wiley & Sons.

Chapman, David, L. Bailey L., and L. Kasprzyk (1986). Nonresponse Adjustment Proceedings Procedures at the U.A. Bureau of the Census, *Survey Methodology*, **12**, 161-180.

CASRO Task Force on Completion Rates (1982). *On the Definition of Response Rates*. Special Report. New York: Council of American Survey Research Organizations.

Cochran, W.G. (1977). *Sampling Techniques*. 3rd edition, New York: Wiley.

Freeman, D.S., S. Thorton A, and D. Camburn (1980). Maintaining Response Rates in Longitudinal Studies. *Sociological Methods and Research*, **9**, 87-98.

Groves, R.M. and Kahn, R.L. (1979), Surveys by Telephone: A National Comparison with Personal Interviews, Academic Press, New York.

Groves, R.M. (1989), Survey Errors and Survey Costs, Wiley and Sons, New York.

Hansen, Morris H., and William N. Hurwitz (1946). The Problem of Nonresponse in Sample Surveys. *Journal of the American Statistical Association*, **41**, 516-529.

Hausman, J.A. and A.D. Wis (1979). Attrition bias in Experimental and Panel Data: the Gary Income Maintenance Experiment. *Econometrica*, **47**, 455-473.

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

Kish, Leslie (1965). *Survey Sampling, New York: Wiley.*

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

Lessler, J.T. and W.D. Kalsbeek (1992). Nonsampling Errors in Surveys. New York: Wiley and Sons.

Politz, Alfred N. and Willard R. Simmons (1949). "An Attempt to Get "Not-at-Homes' into the Sample Without call-Backs." *Journal of the American Statistical Association*, **44**, 9-31.

SAS Institute Inc., *SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 2*, Cary, NC: SAS Institute Inc., 1989.

Shah, B.V. (1991), "Software for Survey Data Analysis (SUDAAN), Version 6.1, Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC.

A more detailed version of this paper is available upon request from the authors.

Table A: Comparative Profile of Final Disposition Codes by Round for the 1990-1992 WIC Survey

Outcome	Baseline (%)		First Followup (%)		Second Followup(%)	
Not Solicited	36.7		16.3		21.1	
No Answer		1.9		0.7		0.3
Busy		0.2		0.0		0.0
Answering Machine		0.9		0.3		0.4
Wrong # / Can't locate		23.1		11.7		18.1
Callback other phone		5.0		2.0		0.9
Callback resp. phone		5.6		1.7		1.3
Solicited / Unable	0.3		0.4		0.3	
Language / Medical Problem		0.3		0.4		0.0
Unreachable for duration		0.0		0.0		0.3
Solicited / Unwilling	0.7		0.7		0.9	
First Refusal		0.3		0.2		0.2
Second Refusal		0.4		0.5		0.7
Other reasons	0.4		0.0		0.0	
Cutoff		0.4		0.0		0.0
Complete	55.7		81.5		77.2	
Ineligible	6.1		1.0		0.5	
Lost baby / Adopted		6.1		1.0		0.5
Total	100.0	100.0	100.0	100.0	100.0	100.0
Sample Size	2090		1165		950	
Response Rate: Lower, Upper Bounds (%)	(59.4, 97.4)		(82.4, 98.7)		(77.6, 98.4)	
Response Rate Best Estimate (%)[CASRO]	61.7		82.5		77.6	

Table B: Response Rates (RR) and Sample Sizes (n) by Demographic Subgroup for the 1990-1992 WIC Survey

	Baseline		First Followup		First Followup	Second Followup		All Rounds
Subgroup	n	RR (%)	n	RR (%)	Cum. RR (%)	n	RR (%)	Cum. RR (%)
Total	2090	61.7	1165	82.5	50.3	950	77.6	38.8
Race								
White	1026	61.7	572	82.5	50.5	467	77.4	38.8
Non-white	1063	61.6	593	82.6	50.2	483	78.1	38.9
Population Density**								
Urban	1001	57.1	517	78.4	44.6	403	75.9	33.8
Rural	1088	65.9	648	85.9	55.6	547	79.0	43.5
Age*								
18-24	446	63.1	256	83.1	52.2	211	75.8	38.9
25-29	926	58.8	488	79.7	46.3	384	74.7	34.4
30+	717	64.5	422	85.5	54.4	355	81.9	44.4
Migrant worker?								
Yes	18	38.9	7	57.1	22.2	4	75.0	16.7
No	2071	61.9	1158	82.7	50.6	946	77.7	39.0
Intend to breastfeed								
Yes	434	60.3	241	79.9	47.8	191	74.3	35.5
No	1324	61.8	735	82.5	50.3	597	77.7	39.0
Maybe	331	63.2	189	86.2	54.2	162	81.5	42.8
Consistent intent?								
Yes	---	---	840	81.7	---	676	76.6	---
No	---	---	325	84.5	---	273	80.1	---

* $p < .02$; ** $p < .0001$; None of these subgroup indicators produced significant effects in a logistic model with response to the second followup interview as the dependent variable.

Table C: Bias Components by Round and Reason for Pre-baseline, Baseline, and First Followup Measures

Round	Reason	Pre-baseline Intent (%)		Baseline Intent (%)		First Followup Accounting (%)	
Baseline	Ineligible	-0.25		---		---	
	Not Solicited	-0.20		---		---	
	Solicited/Unable	0.07		---		---	
	Solicited/Unwilling	0.17		---		---	
	Other reasons	-0.06		---		---	
	All		-0.27		---		---
First Followup	Ineligible	-0.07		-0.20		---	
	Not Solicited	-0.05		0.05		---	
	Solicited/Unable	-0.04		-0.07		---	
	Solicited/Unwilling	0.05		0.09		---	
	All		-0.11		-0.14		---
Second Followup	Ineligible	0.10		0.01		0.17	
	Not Solicited	0.16		0.25		0.66	
	Solicited/Unable	-0.01		-0.01		0.02	
	Solicited/Unwilling	-0.11		-0.20		-0.04	
	All		0.15		0.06		0.81
All Rounds	Ineligible	-0.21		-0.19		0.17	
	Not Solicited	-0.09		0.29		0.66	
	Solicited/Unable	0.02		-0.08		0.02	
	Solicited/Unwilling	0.11		-0.10		-0.04	
	Other reasons	-0.06		---		---	
	All		-0.23		-0.08		0.81

Table D: Total Bias by Demographic Subgroup for Pre-baseline, Baseline, and First Followup Measures

Subgroup		Pre-baseline Intent %	Baseline Intent %	First Followup Accounting %
Race	White	-0.21	-1.21	0.25
	Non-white	-0.27	1.07	1.26
Population Density **	Urban	-1.21	1.53	-0.31
	Rural	0.11	-1.26	1.56
Age *	18-24	1.53	1.39	2.06
	25-29 [1]	-2.13	-1.21	(-0.00)
	30+	1.34	1.27	1.45
Migrant worker?	Yes	27.78	28.57	25.00
	No	-0.44	-0.12	0.71
Consistent Intent?	Yes	---	0.81	1.55
	No	---	-0.72	-0.73

* $p < .02$

** $p < .0001$

All other predictors of nonresponse, i.e., race, migrant status, intent to breastfeed, and consistent intent, were not found to be significant in the logistic model with response to second followup as the dependent variable.

[1] Total bias for the first followup accounting of 25-29 year olds is -0.00364.