

MULTIVARIATE ANALYSIS OF NONRESPONSE IN PERSONAL VISIT SURVEYS

Robert M. Groves and Mick P. Couper (University of Michigan)
Robert M. Groves, P.O. Box 1248, Ann Arbor, MI 48106

KEY WORDS: Nonresponse, refusals

1. Introduction

Nonresponse is a feature of virtually all surveys of human populations. To the extent that full participation is not achieved in a survey, the inferential value of the sample survey method may be threatened. When researchers attempt to reduce the effects of nonresponse, they turn either to efforts to increase response rates or to improve statistical procedures for postsurvey adjustment of respondent data. Both of these strategies, we believe, are aided by an understanding of the causes of survey participation.

There is an extensive literature on survey nonresponse directed at correlates of nonresponse. Empirical studies among household surveys are based either on data from sampling frames comparing respondents and nonrespondents, comparisons to some Census data source, or on observations from interviewers or others. These correlates are reviewed in Groves (1989) and Groves, Cialdini and Couper (1992). The more frequently studied household-level correlates of survey participation include age, household size and composition, race and ethnicity, socio-economic status, home ownership and housing structure.

This literature has a number of shortcomings. First, most of the work is descriptive, addressing the question of what kinds of people tend to be nonrespondent. Second, these descriptions have focussed largely on individual attributes of nonrespondents, ignoring multivariate relationships. Third, the literature is characterized by an over-emphasis on case studies, with few (if any) cross-survey comparisons of nonresponse. Fourth, many of the studies have failed to distinguish between varieties of nonresponse (noncontact, refusal, etc.), or focus on only one component to the exclusion of others. This has led to a number of inconsistent findings in the literature (see Goyder, 1987). Finally, the literature's collective results are largely the result

of data dredging with conveniently available data, not theoretically motivated concepts.

We are currently engaged in an attempt to address these issues and are working on the specification of a mid-level theory of survey participation (see Groves, Cialdini and Couper, 1992) to guide an understanding of the processes that lead to nonresponse. Our theoretical perspective includes four sets of influences that form the foundation of the householder's reaction to a survey request: the social context, survey design, interviewer and respondent.

Some of the influences arise from relatively stable characteristics of the social environment or social context of the householder. These include prevailing social and economic conditions and public attitudes toward surveys which influence the "survey-taking climate" (Lyberg and Dean 1992). In addition, attributes of the neighborhood and the urbanicity of the residential location of the householder can influence reactions to survey requests from strangers.

The likelihood of survey participation is also directly influenced by various attributes of the households or persons sampled. These include knowledge of the topic of the survey that may determine the cognitive burdens of answering survey questions, prior experience as a survey respondent, and affective states extant at the time of the survey request.

Surveys with high cooperation generally manipulate survey design characteristics to minimize refusals. Surveys with different length of interviews, respondent selection procedures, and mode of data collection tend to vary in cooperation rates, depending on the characteristics of the population studied. Similarly, the survey design, through recruiting, training, and supervision of interviewers leads to a set of interviewers whose socio-demographic and attitudinal characteristics can influence the likelihood of cooperation of the householders they contact (see Couper and Groves, 1992a).

The influences of survey design and interviewer characteristics may or may not manifest

themselves in the interaction between interviewer and householder, depending on the nature of the conversation between the two. Whatever happens, however, we believe, is seen through the lens of the social context of the householder and the psychological states relevant to a survey request. In the brief contacts that characterize interactions between interviewers and householders, that subset of factors deemed by the householder most relevant to the decision to participate are evoked and form the basis of the judgement to participate or refuse the survey request (Groves and Couper, 1992).

The aim of this paper is relatively narrow. Within the context of the factors discussed above, we focus only on household-level effects on survey participation. We examine socio-demographic correlates of participation in the light of findings from past literature.

2. Data Collection Design

The 1990 decennial census provided a rare opportunity to obtain information on survey nonrespondents from their decennial census records. The data used here were produced by matching nonrespondent and respondent cases from six different national face-to-face household surveys to records from the decennial census: Consumer Expenditure Survey, Current Population Survey, National Health Interview Survey, National Crime Survey, National Household Survey on Drug Abuse and Survey of Census Participation. The selection of these surveys was somewhat arbitrary, and they should be treated as six separate case studies.

From each of these surveys a probability sample of respondent and nonrespondent cases was selected. Sample addresses were then matched to decennial census records in order to extract census data for the selected cases. All matching was done at the address level. For details of the survey-census match operation, see Couper and Groves (1992b).

The data at our disposal have a number of limitations. First, these are all high response-rate surveys. Response rates for these six surveys range from 82% to 97%, making the detection of effects on response rates difficult. Second, given the selection of these six surveys, generalization to other surveys and organizations is limited. Third,

the information we have on nonrespondent (and respondent) households is limited to key demographic indicators collected in the decennial census. Nevertheless, the data allow us to examine household-level correlates of various components of nonresponse in a multivariate context, using data from six different surveys.

In this paper we have pooled the six surveys into a single dataset, containing first-wave cases only from panel surveys. We use three participation rates. The response rate is defined as (interviews)/(all eligible units). The contact rate is defined as (interviews + refusals + other noninterviews)/(all eligible units). Finally, the cooperation rate is defined as (interviews)/(interviews + refusals). Cooperation is thus defined as response, given contact.

The multivariate modeling is done using logistic regression with a variety of binary outcome variables coded as 1=success (response, contact, or interview respectively) and 0=failure (nonresponse, noncontact or refusal). Standard error estimates and statistical tests are calculated using Taylor Series approximations, reflecting stratification and clustering of the survey designs (using SUDAAN, Shah et al. 1991). Data are weighted to reflect different probabilities of selection, both in the original studies, and in the survey-census match sample.

3. Results

Correlates of Nonresponse

We begin with an examination of bivariate correlates of various components of nonresponse. In Table 1 we present the three rates by various household characteristics for the pooled data set.

The indicator for tenure suggests that home owners are more likely to respond than renters. However, this appears largely due to difficulties contacting renters, as cooperation rates do not differ significantly for the two groups. The marginal effects of tenure, therefore, might be affected by controlling on household size. Socio-economic status (as measured by house value among owners and monthly rent among renters) appears negatively associated with all three participation rates. This tends to support Smith's (1983) finding that higher income households tend to be disproportionately nonrespondent.

Response rates are significantly lower for

households in large multi-unit structures than in single family homes. Again, this appears largely due to a noncontact problem. Once contacted, apartment dwellers and those in single-family units show no differences in cooperation. This finding supports the common speculation of field interviewers that "getting into the structure" is the main problem but that cooperation given contact is no more difficult in such structures.

Turning to household size, the bivariate results support the literature. All types of nonresponse appear to decrease with increasing household size, with the biggest differences being between single-person households and those with two persons.

Given that we have household-level rather than person-level data, we created a variable to distinguish households with all younger persons from those with all older persons. This shows that households where all persons are under 30 have higher response rates than other households. It appears that these households are significantly more difficult to find at home, but once contacted, are more cooperative. Contrary to expectation, elderly households (those where all members are over 70) do not have response rates that are much lower than other households.

The presence of children in households show a positive effect on survey participation for all measures (response, contact and cooperation). Finally, we find no significant response rate differences for race/ethnicity, although it appears that cooperation rates for White nonHispanics are significantly lower than that for all other groups.

Multivariate Analysis of Cooperation, Given Contact

It is clear that many of the variables in Table 1 are themselves correlated with one another. For example, those in large multi-unit structures may be more likely to be renters than owners, elderly households may be disproportionately single-person households, and so on. Multivariate analysis is needed to disentangle the effects of these variables.

The results of a lengthy process of sequential model-fitting are produced in Table 2. Although the fitted model also includes urbanicity effects, only the household-level variables are presented and discussed here. We have focussed our efforts on the specification of a logistic regression model predicting cooperation (given contact), and this is

presented in the first column in Table 2. Generally, the results confirm those from the bivariate analyses, and support those reported in the literature reviewed earlier.

The finding for household age composition is again a surprise. The marginal effect for households with all persons under 30 is significant and positive, even after controlling for presence of children in the household. Similarly, households where all members are over 70 are also significantly *more* likely to cooperate with a survey request, relative to those households where one or more members are between the ages of 30 and 70. As these are all surveys conducted or sponsored by the federal government, one explanation may be that government sponsorship might simultaneously increase the positive effects of civic duty among the elderly and decrease the fear of victimization.

Multivariate Analysis of Contact with Sample Households

An important question for postsurvey nonresponse adjustment is whether the same influences on refusal propensity apply to noncontact propensity. We note that increasing overall response rate by efforts in the field usually acts to reduce the noncontact rate of a survey more than the refusal rate. This is because increasing the number of callbacks is a more cost efficient way to reduce nonresponse in contrast to refusal conversion. However, if different influences apply to noncontact and refusals, different optimal postsurvey adjustment models should apply for high and low response rate surveys.

Table 2 is a simple test of whether the same model applies to both dependent variables. After the cooperation model was constructed consistent with the theoretical structure, we used the same model specification for the dependent variable measuring contact propensity.

It is interesting to note that we find no effect for type of dwelling unit on cooperation in these surveys. However, examining the contact-rate model in Table 2, we find that large multi-unit structures (those with 10 or more apartments) have a significant negative effect on contact. This supports our interviewer reports that gaining access to such structures and finding their residents at home are the biggest problems; once such

persons are contacted they are no less likely to cooperate with the survey request than other households.

We also note that whereas the effects for house value and monthly rent are relatively small in the cooperation rate model, these coefficients are both significantly negative in the contact rate model. This suggests lifestyle differences across socio-economic status (as measured by these two variables). Greater difficulty is experienced finding people in higher cost homes and apartments at home, but once contacted they show little difference in cooperation.

4. Summary and Conclusions

This paper has focussed on one particular component of a broader theory of survey participation. We believe such a theory should acknowledge that participation has many causes, many of which may interact in producing a final decision. Understanding the social context in which requests for participation take place, the nature of the request (as reflected in survey design differences), the role of the interviewer, other social psychological attributes of the sample household, and the nature of the interaction between householder and interviewer are, we believe, all critical in understanding the factors that lead to such a decision.

These findings clearly illustrate the importance of differentiating among the different components of nonresponse. The cooperation and contact models in Table 2 differ in a number of important respects. This may in part explain some of the contradictory findings in the literature on correlates of nonresponse. It also suggests that the form of postsurvey adjustment models for noncontact would differ from those for refusals. For example, surveys with very high contact rates will move sample units in multi-unit structures from "noncontact" status to either "interview" or "refusal" status. The cooperation model suggests they will cooperate or be refused at about the same rate as other households. This means the overall relationship between nonresponse and structure type will vary by the level of the noncontact rate. Surveys with high contact rates will find no relationship between response rate and structure; those with low contact rates will find a relationship. Specification of postsurvey

adjustment models need to take into account such differences.

References

- Couper, M.P. and Groves, R.M. (1992a). "The role of the interviewer in survey participation." Survey Methodology, 18 (2): 263-271.
- Couper, M.P. and Groves, R.M. (1992b). Survey-census match procedures. Washington, DC: Bureau of the Census (unpublished memorandum).
- Goyder, J. (1987). The Silent Minority: Nonrespondents on Sample Surveys. Boulder, CO: Westview Press.
- Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley and Sons.
- Groves, R.M., Cialdini, R.B., and Couper, M.P. (1992). "Understanding the decision to participate in a survey." Public Opinion Quarterly, 56 (4): 475-495.
- Groves, R.M. and Couper, M.P. (1992). "Respondent-interviewer interactions in survey introductions." Paper presented at the Third International Workshop on Household Survey Nonresponse, The Hague.
- Lyberg, L. and Dean, P. (1992). "Methods for reducing nonresponse rates: A review." Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Petersburg, FL.
- Shah, B.V., Barnwell, B.G., Hunt, P.N. and LaVange, L.M. (1991), SUDAAN User's Manual, Release 5.50. Research Triangle Park, NC: Research Triangle Institute.
- Smith, T.W. 1983. "The hidden 25 percent: An analysis of nonresponse on the 1980 General Social Survey." Public Opinion Quarterly, 47: 386-404.

Acknowledgements

This work is supported by a funding consortium of the Bureau of the Census, Bureau of Labor Statistics, National Center for Health Statistics and Substance Abuse and Mental Health Services Administration. Contents of the paper are the responsibility of the authors and do not necessarily represent those of the above agencies.

Table 1 Participation rates by household-level variables

	Response Rate		Contact Rate		Cooperation Rate	
Tenure:						
Rent	90.9	(0.5)	96.5	(0.3)	95.4	(0.3)
Own	92.4	(0.3)	98.0	(0.1)	94.9	(0.3)
		*		**		ns
House value (among owners):						
Lowest	94.3	(0.5)	98.7	(0.2)	96.1	(0.4)
Lower	92.3	(0.6)	98.0	(0.3)	94.8	(0.5)
Higher	93.1	(0.6)	98.3	(0.3)	95.3	(0.5)
Highest	90.6	(0.7)	97.2	(0.3)	94.0	(0.6)
		**		**		*
Monthly rent (among renters):						
Lowest	93.6	(0.7)	97.7	(0.4)	96.7	(0.5)
Lower	92.3	(0.8)	96.9	(0.4)	96.1	(0.6)
Higher	89.9	(1.1)	95.5	(0.7)	95.0	(0.7)
Highest	88.0	(1.3)	95.5	(0.7)	94.1	(0.7)
		**		*		*
Units in structure:						
Single family/mobile home	92.7	(0.3)	98.1	(0.1)	95.2	(0.2)
10+ apartments	86.2	(1.1)	93.7	(0.7)	94.0	(0.7)
Other	91.1	(0.7)	96.7	(0.3)	95.1	(0.5)
		**		**		ns
Household size:						
One	88.6	(0.7)	95.9	(0.3)	93.7	(0.5)
Two	92.1	(0.4)	97.7	(0.2)	95.0	(0.4)
Three	93.0	(0.5)	98.2	(0.2)	95.4	(0.4)
Four	93.4	(0.6)	98.5	(0.2)	95.3	(0.5)
Five or more	94.4	(0.5)	98.3	(0.3)	96.9	(0.4)
		**		**		*
Household age:						
All under 30	93.6	(0.5)	96.9	(0.4)	97.5	(0.3)
All over 70	91.4	(0.3)	97.3	(0.1)	94.6	(0.3)
Other	91.7	(0.9)	98.3	(0.3)	95.1	(0.7)
		**		**		**
Children in household:						
Yes	94.0	(0.3)	98.4	(0.1)	96.0	(0.3)
No	90.7	(0.4)	97.0	(0.2)	94.5	(0.3)
		**		**		**
Race/ethnicity:						
Hispanic	93.8	(0.9)	97.5	(0.5)	96.9	(0.6)
White nonHispanic	91.8	(0.3)	97.7	(0.1)	94.7	(0.2)
Black nonHispanic	91.9	(0.7)	96.8	(0.4)	95.8	(0.5)
Other	91.4	(1.1)	97.3	(0.5)	96.6	(0.7)
		ns		ns		*

Chi² tests: ns p > .05, * p < .05, ** p < .01, standard errors in parentheses

Table 2 Effect of Household Variables on Cooperation and Contact

Independent Variables	Cooperation Rate		Contact Rate	
	Coeff	(SE)	Coeff	(SE)
Reference Person				
Black nonHispanic	0.34 *	(0.15)	-0.22	(0.16)
Hispanic	0.55 *	(0.23)	-0.11	(0.21)
Single Person Household	-0.33 *	(0.14)	-0.59 **	(0.13)
Single Family Unit/ Mobile Home	0.33	(0.18)	0.39 **	(0.15)
Large Multi-Unit Structure	0.035	(0.18)	-0.39 *	(0.17)
Children in Household	0.23 *	(0.11)	0.36 **	(0.11)
All Persons Under 30	0.82 **	(0.15)	-0.03	(0.15)
All Persons Over 70	0.38 *	(0.19)	0.61 **	(0.19)
Owner Occupied	-0.29	(0.21)	-0.21	(0.18)
Monthly Rent for Renters ¹	-0.051	(0.032)	-0.091 **	(0.029)
House Value for Owners ²	-0.013 *	(0.0057)	-0.029 **	(0.0050)

* $p < .05$, ** $p < .01$

¹ Measured in units of \$100

² Measured in units of \$10,000