SURVEY TOPIC INVOLVEMENT AND NONRESPONSE BIAS[1]

Brian A. Kojetin (BLS), Eugene Borgida and Mark Snyder (University of Minnesota)
Brian A. Kojetin, Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Washington DC 20212

Key words: Nonresponse, recycling, mail survey

The response rates of surveys conducted by most private and academic organizations have steadily declined the past few decades (Goyder, 1987). The response rate of government surveys appear to have been less effected by this general decline, although there does also appear to be some impact even on surveys with high response rates (Groves, 1989). For example, the Current Population Survey (CPS) has consistently maintained a response rate of 95%; however, the proportion of refusals has increased during the past several years (DeMaio, Marquis, McDonald, Moore, Sedlacek, & Straf, 1986).

Survey response rates are often interpreted as an index of the quality of the data (Tucker, 1987). However, bias in survey estimates is due not simply to the response rates, but also to differences between respondents and nonrespondents on relevant survey variables. Nonresponse bias in a linear statistic can be expressed as:

$$(n_{nr} / n) (y_{nr} - y_r)$$

where:

n_{nr} = the number of nonrespondent cases in the sample;
n = the full sample size
y_{nr} = linear statistic for variable y based on nonrespondent cases in the sample;
y_r = the same linear statistic for variable y based on the respondent cases in the sample

In the above expression, the first term, (n_{nr}/n), represents the nonresponse rate, and the second term, $(y_{nr}-y_r)$, represents the difference between the nonrespondents and the respondents on some linear statistic of interest (Groves & Cooper, 1992).

It is the nature of nonresponse; however, that the linear statistic for nonrespondent cases is rarely available. The individual characteristics of nonrespondents have most often been inferred from survey studies using panels or successive waves to increase participation. Later responders are then compared to the initial responders to infer the characteristics of people who did not respond at all. The assumption underlying this design is that the people who initially "resist" and later "convert" and respond are similar to those who refuse to participate. Researchers rarely have much external information available on nonrespondents, but they have been able to obtain some information under some special circumstances. For example, people's involvement with the survey topic or sponsoring organization has studied most frequently by the particular organization to which they belong, which has information on all members. Donald (1960) found that the interest and involvement of members of the League of Women Voters was related to the promptness (in terms of wave) of their response to an organizational survey. Similarly, Goyder (1987) found faculty were more likely to respond to a university survey on computer use if they were higher in rank and were more frequent users of computers. Recently, Groves and Couper (1992) have been able to match decennial census data to nonrespondent cases from seven national surveys and are exploring the available characteristics of nonrespondents to each of these surveys.

Bias in the sample due to nonresponse can influence not only estimates of population parameters but also may effect the interrelationships among survey variables and even the prediction of other criteria from variables on the survey. For example, Goudy (1976) surveyed a rather homogeneous sample who had been interviewed two years earlier. Although Goudy found differences in the distributions, means, and standard deviations of some of the earlier interview items according to which wave the person responded, these differences proved to be quite small when they were cumulated with the prior waves. Goudy (1976) also examined correlations

[1]The views expressed in this paper are those of the authors and do not necessarily represent those of the Bureau of Labor Statistics.

among the interview items and used regressions to predict survey responses from the prior interview items as a function of which wave the survey was returned. He found that "Only slight differences were evident when bivariate relationships were tested at various wave points; however, several relationships were examined simultaneously in the multiple regression, and the interaction of small differences in the correlation matrix produced larger changes in the regression equation" (p. 368). Nevertheless, Goudy concluded that the effect of nonresponse on variable relationships had been exaggerated.

The Present Study

We conducted a survey on people's attitudes and knowledge about recycling to obtain estimates of population parameters for those variables and also to predict recycling behavior from those factors. In the present investigation, we sent a recycling survey to a random sample of registered voters in one metropolitan county with a stratified oversampling of one city within the county that had a well-established recycling program. Although we knew nothing about most nonrespondents to our survey, we had some limited information about the prior recycling behavior of people from one city, and we obtained records of their actual participation in that community's recycling program for six months following the survey. These data afforded us the opportunity to examine nonresponse bias in terms of actual behavior directly relevant to our survey. Furthermore, we employed these behavioral data to weight the questionnaire data from respondents and, thus, examine the influence of nonresponse on the prediction of recycling behavior from the survey questionnaire items.

Method

Sample

The survey questionnaire was mailed to a random sample of all residents who were registered voters in Hennepin County, Minnesota. 1583 questionnaires were sent and the overall response rate was 71%. Approximately 27% percent of the total were sent to one city (St. Louis Park) in Hennepin County, which has had an active, incentive-based recycling program operating for several years. Residents who recycle are given credit on their utility bills. Our sampling of this city was stratified by the resident's participation in the community's recycling program six months prior to the survey. Specifically, we randomly selected half of the community sample from four districts in the city from those people who previously had been infrequent recyclers, and the remaining half who had been more frequent recyclers.

Procedure

The questionnaire was titled the Minnesota Recycling Survey and was accompanied by a letter from the Minnesota Center for Survey Research which is associated with the University of Minnesota. The person in the household most familiar with recycling issues was requested to complete and return the questionnaire. The instrument consisted of items asking about their current recycling behavior, their attitudes toward recycling, their knowledge and awareness of recycling programs, and their reasons for recycling or not recycling. One week after the original questionnaire was mailed out, a reminder card was sent. About 10 days after the reminder card was sent, another copy of the questionnaire with an accompanying letter was mailed to all potential respondents who had failed to return it previously.

Measures

Recycling Attitudes. To measure people's general attitudes towards recycling, we created a scale composed of a series of items inquiring about people's beliefs, opinions, and feelings about recycling. For example, "The disposal of solid waste is fast becoming the number one threat to a clean environment." Respondents indicated how much they agreed or disagreed with each item on a 4 point Likert scale.

Awareness and Knowledge about Recycling. To determine how well-informed these people were about recycling, we asked them eight True-False test questions that tapped their knowledge of recycling programs and government policies concerning recycling. For example, "All colors of glass are recyclable" (True). Their awareness of their community's recycling programs was assessed by a single item that asked the survey respondents if they were aware of recycling programs in their area.

Intentions to Participate in Recycling. To measure the degree to which people intended to

participate in their community's recycling program, we gave them a list of recyclable materials and asked them how likely they would be to recycle each of those items at least once a month over the next six months on a scale from 0 (extremely unlikely) to 4 (extremely likely). A scale was created summing the responses to three items: newspapers, glass, and cans.

Frequency of Recycling. To assess the level at which people were involved in recycling, we created a frequency of recycling scale composed of items that inquired how often they said they recycled three different items: newspapers, cans, and glass on a scale from 0 (never) to 4 (more than once a month). Thus, people's scores on this scale could range from 0 if they never recycled newspapers, cans, or glass, to 12 if they recycled all three more than once a month. For people living in one city, St. Louis Park, we also obtained a classification of their recycling behavior six months prior to the survey, either frequent recycler or infrequent recycler. In addition, their *actual* frequency of recycling for the six months following the survey was also obtained. The city picked up recyclables twice a month, so these scores could range from 0 (no recycling) to 12 (twice a month for six months).

Results

Wave Analyses

Demographic Differences. Forty-nine percent of the respondents returned their survey within one week of its mailing and before our first reminder. An additional 27% percent responded within ten days of our first reminder and 24% replied after our final reminder. There were few demographic differences among people who responded during these three different waves. No differences were observed in age or sex or income level of the respondents across the three waves, F's (2, 1024) < 1.73, all p's > .10. However, there were significant differences among respondents of each wave in their educational backgrounds and home ownership. Specifically, people who returned the survey in the first wave were more educated than those who returned it in the third wave, F (2, 1028) = 6.04, $p < .01$, with Tukey HSD follow-up test. Homeowners were also more likely to return their surveys sooner than renters, X^2 (2) = 10.06, $p <$ .01.

Predictor Variable Mean Differences. To examine differences by response wave in variables predicted to be relevant to people's recycling behavior, we conducted a multivariate analysis of variance with attitudes, intentions, awareness, and knowledge of recycling as the dependent variables and wave as the independent variable. There was a significant multivariate effect for wave mF (10, 1706) = 5.24, $p < .001$. The results of a series of follow-up univariate analyses of variance are shown in Table 1 along with the means of each of the recycling predictor variables by wave. There were significant effects of wave on each of the predictor variables. Post hoc Tukey HSD tests revealed that people returning their surveys during the third wave had less positive attitudes toward recycling, less awareness of their community's recycling program, less knowledge about recycling, and also reported that they recycled less often than people who returned their surveys during either of the first 2 waves (respondents to the first two waves did not differ significantly from each other). Respondents who returned their surveys in the third wave had lower intentions to recycle than people from the first wave, but their intentions were not different from the second wave respondents. As shown in the bottom half of Table 1, this pattern of results was very similar when the sample was restricted to only St. Louis Park.

Behavioral Analyses

The wave during which the respondents returned their questionnaire was marginally related to their previous classification as a frequent or infrequent recycler, X^2 (2) = 4.78, $p < .10$. Infrequent recyclers were more likely to return their surveys during the third wave rather than during the first two waves. There was also a significant relationship between the promptness of response to the survey (wave) and their subsequent recycling behavior, $r = -.19$, $p < .01$, indicating that people who returned their surveys more quickly recycled more frequently during the six months after the survey.

An examination of Figure 1 further shows that the distribution of recycling behavior was significantly different for people who *did* and *did not* return the survey, X^2 (12) = 90.62, $p < .001$, with respondents being much more likely than

respondents to recycle more frequently, which would result in biased parameter estimates of recycling behavior. In this case, the actual mean number of times recycled in six months in this sample based on the survey respondents was 8.41, but was only 4.82 for nonrespondents, resulting in a nonresponse bias of: (.29)(4.82 - 8.41) = -1.06.

Predicting Recycling Behavior. The substantive purpose of our investigation was to predict people's recycling behavior, so we further examined how the restriction of range on this variable influenced the results of our analyses across the different waves. As can be seen in Table 2, there is a fair degree of variability in how strongly the variables of attitudes, intentions, knowledge and awareness predicted behavior at each wave. In a similar manner to Goudy's (1976) results, these differences are less noticeable when responses are cumulated across waves. However, there are still some substantial differences even at the cumulative level. For example, the importance of attitudes and intentions in predicting behavior is well documented in social psychological research (e.g., see Ajzen, 1987), but it is not until the addition of wave 3 data that attitudes and intentions become significant predictors of behavior.

Simultaneously entering all of these predictors into a regression equation gives further insight into the importance of these variables in predicting behavior. At wave 1, wave 2, and both together, awareness and knowledge contributed significantly to the prediction of behavior with intention playing no role. With the addition of the third wave respondents, however, intention emerges as the most important predictor of behavior.

Weighted Correlation and Regression Analyses. Knowing our nonrespondents behavior unfortunately does not tell us their awareness, knowledge, attitudes, or intentions. However, if we assume for a moment that people who have the same behavior have *similar* attitudes, intentions, and levels of knowledge and awareness, then we can roughly estimate the effect of nonresponse bias on the relationships among our predictors and behavior. We conducted this analysis by weighting our the number of respondents at each level of recycling behavior to reflect the actual proportion of the *total* number of people at each level. Thus, those people who recycled less frequently were weighted more heavily than those who recycled more frequently to compensate for the people who did not respond and also recycled less frequently. We then conducted correlational and regression analyses on the weighted sample. The zero-order correlations of attitudes, intentions, and knowledge with behavior all increased in magnitude from their unweighted levels. Regressing behavior on the predictors led to similar increases in overall prediction with an additional 5.9% of the variance being explained.

Conclusions

The present study examined the amount of nonresponse bias in a community survey on recycling and its effect on the interrelations among important variables. Significant differences in theoretically relevant variables such as knowledge about recycling, awareness of recycling programs, attitudes about recycling, intentions to recycle, and reported recycling behavior were found across waves of responses to the survey. Further, there were clear actual behavioral differences among respondents of different waves and between respondents and nonrespondents in their frequency of recycling during the six months following the survey. These differences lead to biased estimates of frequency of behavior, and one may presume also, people's attitudes, intentions, knowledge about recycling, and awareness of recycling.

Further analyses showed that this nonresponse bias had some influence on the relationships among the predictors, attitudes, intentions, knowledge, and awareness with behavior. Specifically, analyses of the first two waves of respondents to the survey considerably underestimated the influence of intentions on behavior.

The results of the present study have implications for researchers using survey data for substantive theory testing. People who are not as involved with, aware of, or knowledgeable about the survey topic were less likely to return the survey. However, for theory testing, these people are extremely important as a comparison group to those who are more involved, aware, and knowledgeable. Thus, it appears that the variables that hold a central place in surveys are also important considerations that respondents use to self-select themselves for participation. This self-selection or nonresponse bias may lead to biased

parameter estimates and attenuated relationships with other variables.

References

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed.) *Advances in Experimental Social Psychology*, vol. 20. New York: Academic.

DeMaio, T., Marquis, K., McDonald, S., Moore, J., Sedlacek, D., & Straf C. B. (1986). Cognitive and motivational bases of census and survey response. *Proceedings of the Second Annual Research Conference,* U. S. Bureau of the Census, Washington, D. C.

Donald, M. N. (1960). Implications of nonresponse for the interpretation of mail questionnaire data. *Public Opinion Quarterly, 24*, 99-114.

Goudy, W. J. (1976). Nonresponse effects on relationships between variables. *Public Opinion Quarterly, 40*, 360-369.

Goyder, J. (1987). *The silent minority*. Boulder, CO: Westview Press.

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

Groves, R. M. & Couper, M. (Aug., 1992). *Correlates of nonresponse in personal visit surveys*. Paper presented at the Joint Statistical Meetings, Boston, Massachusetts.

Tucker, C. (1987). Toward the development of an ordered set of data quality categories. *Proceedings of the Section on Survey Research Methods American Statistical Association,* 558-563.

Figure 1. Distribution of Post-Survey Recycling Behavior for Survey Respondents and Nonrespondents

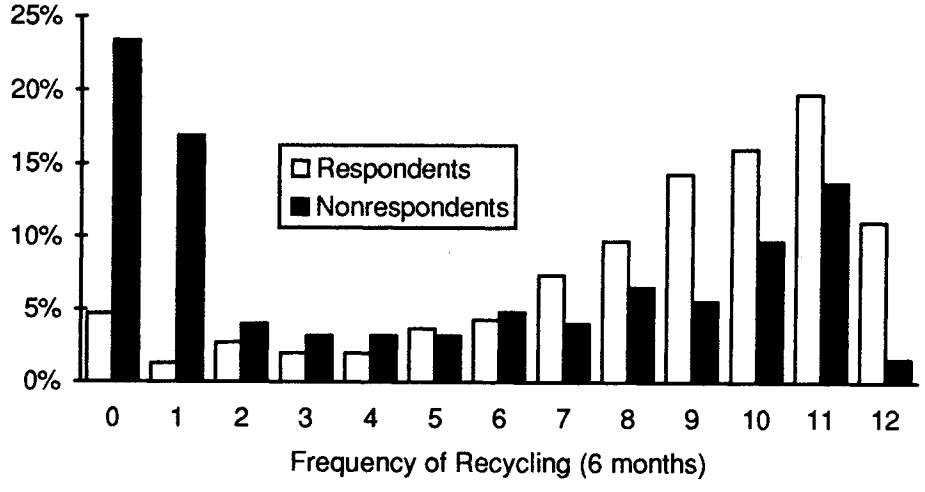

Table 1. Means and F values of Predictor Variables and Self-reported Behavior at each Wave.

Variable	Wave 1	Wave 2	Wave 3	F
Attitudes	4.28_a	4.27_a	4.04_b	16.00**
Intentions	10.03_a	9.45_{ab}	8.75_b	9.96**
Awareness	$.78_a$	$.78_a$	$.62_b$	12.17**
Knowledge	$.51_a$	$.51_a$	$.43_b$	9.27**
Self-reported Behavior	6.73_a	6.13_a	4.47_b	19.69**
St. Louis Park Only				
Attitudes	4.43_a	4.43_a	4.15_b	8.78**
Intentions	11.56_a	11.42_a	10.38_b	10.41**
Awareness	.99	.97	.94	2.51#
Knowledge	$.64_a$	$.61_{ab}$	$.56_b$	3.47*
Self-reported Behavior	10.39_a	9.28_b	7.89_c	13.49**

Note: Means with different subscripts are significantly different at $p < .05$ by Tukey HSD test.
** $p < .001$ * $p < .05$ # $p < .10$

Table 2. Correlations and Regression Weights for Predictor Variables with Recycling Behavior by Wave.

Zero-order correlations					
Variable	Wave 1	Wave 2	Wave 3	Waves 1-2	Waves 1-3
Attitudes	.10	.00	.15	.06	.13*
Intentions	.14	.10	.44**	.13#	.30**
Awareness	.28**	.19#	- .02	.23**	.13*
Knowledge	.22**	.28**	.25*	.25**	.27**
Regression weights					
Attitudes	-.01	-.07	-.10	-.03	-.03
Intentions	.11	.04	.45**	.09	.26**
Awareness	.26**	.13	.04	.19**	.12*
Knowledge	.17*	.25*	.13	.20**	.20**
R	.351	.316	.463	.321	.376
R^2	.123	.100	.214	.103	.141

Note: ** $p < .01$ * $p < .05$ # $p < .10$