When Less is More: Are Reluctant Respondents Poor Reporters?

Ting Yan

Joint Program in Survey Methodology, University of Maryland

Abt Associates, Inc

Roger Tourangeau

Joint Program in Survey Methodology, University of Maryland

Survey Research Center, University of Michigan

Zac Arens

Joint Program in Survey Methodology, University of Maryland

June 1, 2004

Abstract

Survey research has been witnessing declining response rates across a wide range of surveys. As a result, extensive resources have been invested on boosting response rates under the assumption that a higher response rate will lead to lower survey error. However, the extra attempts to reduce nonresponse are costly. Given limited budgets, survey researchers might be more confident about allocating a larger share of survey budgets to decreasing nonresponse if they could be assured that an improved response rate does in fact lower survey error. In fact, the empirical evidence in this regard is mixed. Some studies show respondents interviewed on early calls differ from those interviewed on later calls. However, other recent studies seem to demonstrate that large differences in response rate have only minor effects on cross-sectional analyses.

This study, analyzing a national RDD survey data that measures the public knowledge of and attitudes towards science, replicates the work of Curtin et al (2000) by looking at the impact of response rates on key survey variables from another topic area. In addition, this study extends Curtin et al's work (2000) by exploring the relationship between the nonresponse errors and the measurement errors in a national survey. It will focus on answering the question whether reluctant respondents have more or less measurement error.

Introduction

Declining response rates pose a threat and a challenge for survey researchers; it is an article of faith for most survey researchers that high response rates are better than low ones. Assuming that a higher response rate will lead to lower survey error (Groves 1989), survey organizations invest extensive resources on boosting response rates. But given a limited budget, survey researchers have to balance the cost of improving response rates against the potential benefit of reduced total survey error. The goal is to improve response rates in exchange for reduced nonresponse bias and lower total survey error. However, two questions follow: First of all, does the attempt to increase response rates necessarily reduce nonresponse errors? Secondly, does the reduction of nonresponse error necessarily result in a reduction of total survey error? One concern survey methodologists have regarding the second question is that the efforts to reduce nonresponse error might cause an increase in measurement error. Specifically, they ask the question: Are reluctant respondents poor reporters or good reporters? In other words, after we spend resources on pursuing and converting reluctant respondents to participate, would they bring their poor motivation into answering survey questions and give inaccurate data? If so, efforts to boost response rate may actually increase total survey error.

The literature presents mixed evidence regarding the first question. Some studies demonstrate that nonresponse does affect survey statistics and that increasing the response rate reduces nonresponse bias (Groves, Singer & Corning 2000). However, the latest findings seem to show the opposite (Keeter, Miller, Kohut, Groves & Presser 2000; Curtin, Presser & Singer 2000; Merkle & Edelman 2002). For instance, Keeter and colleagues (2000) uncovered very few significant differences in point estimates subject to different response rates. Similarly, Merkle and Edelman (2002) obtained either very little or no correlation between response rates and two exit poll error measures. Curtin, Presser and Singer's findings (2000), replicating those of the first two studies, revealed near zero effect of lower response rates on estimates from the Survey of Consumer Attitudes.

So far as the second question is concerned, not much has been done to investigate the relation between nonresponse error and measurement error, despite a few scattered findings here and there. One assumption survey researchers usually hold is that nonresponse is mostly motivational whereas measurement error is primarily cognitive; therefore, the two sources of errors may be uncorrelated. Probably, this implicit assumption underlines the standard practice of conducting separate studies of nonresponse error and measurement error. On the other hand, some researchers worry that there is a relationship between nonresponse and measurement error. Because of the motivational nature of nonresponse, they fear that converted nonrespondents might bring their low motivation to the survey interview and provide survey data of poor quality--that reluctant respondents might be poor reporters. Several studies seemed to have substantiated such a concern. Cannell and Fowler (1963) discovered that respondents who completed a self-administered questionnaire early tend to report more accurately than people who returned it late. Echoing this finding, Willimack, Schuman, Pennell and Lepkowski (1995) showed that early respondents exhibited greater response completeness than late responders. Looking at a panel study (SIPP), Bollinger and David (2001) linked missing interviews in a panel with response error through a probit analysis and found out that response error was correlated with wave nonresponse in later rounds of the panel.

There are, however, no studies yet that tackle the nature of the causal relationship between nonresponse and measurement error, assuming such a relationship exists. Two competing causal models have been proposed. The first model posits a common cause, in which the same factors affect both nonresponse and measurement error. For instance, social desirability concerns related to the survey topic might influence a sample person's decision to participate in a survey. The same concerns might cause the respondent to underreport on, say, their drug use even if the respondent were persuaded to respond. Social desirability concerns are the common cause, in this case, that produces a spurious relationship between nonresponse and measurement error. Partialling out the common cause will reduce or even eliminate the relationship. The second model—the intervening cause model—posits that response propensities change some internal state of the respondents that subsequently affect their behavior as respondents. For instance, the decision to participate in a survey, induced either by extra monetary incentives or successful refusal conversion efforts, affects the respondents' willingness to provide accurate data, which subsequently affects the respondent's data quality. At present, no research has been done to determine which one of these two competing models is the right one.

This paper attempts to examine these two questions together but with emphasis on the second research question—the relation between nonresponse and response quality. The stochastic nonresponse model assumes that every respondent has a propensity to respond (response propensity *p*) that varies across respondents (Groves 1989). Similarly, every respondent has a propensity to be a nonrespondent, which is 1-*p* and can be called the nonresponse propensity. To better capture nonresponse and response propensity, I adopted Groves and Couper's survey participation framework, which distinguishes three major types of nonresponse--nonresponse due to noncontact, nonresponse due to refusal, and nonresponse due to other problems, such as respondents' inability to respond because of mental or physical impairment (Groves & Couper, 1998). This paper accordingly distinguishes several response propensities—the propensity to be a noncontact nonrespondent, the propensity to be a refusal nonrespondent, and lastly, the propensity to be a late respondent, which presumably combines the effects of nonresponse due to noncontact and nonresponse due to refusal. I examine the effects of nonresponse on survey statistics and on measurement errors.

To answer the first research question—i.e., the impact of nonresponse on nonresponse error, univariate distributions of key study variables will be constructed and plotted against the response rate. In addition, I will compare the distributions for respondents and for noncontact nonrespondents, for amenable respondents who never refused and for refusal nonrespondents, and for early respondents and for late respondents. I hypothesize that the univariate distributions for key variables do not change much as the response rate changes. Consistent with the studies by Keeter et al (2000) and Curtin et al (2000), I hypothesize that there won't be many or large significant differences between respondents and noncontacts, between respondents and refusals, or between early respondents and late respondents.

To answer the second research question—i.e., the relation between nonresponse and measurement errors, I develop three sets of response propensity scores. One set of propensity scores contrasts easy-to-contact respondents with noncontacts. The second set compares amenable respondents with reluctant respondents and the last set combines the effects of both noncontact and reluctance by looking at early respondents versus late respondents.

Furthermore, I use five response effects as proxy measures of response quality. These five response effects are considered as reflecting measurement errors in the survey research literature because they are the result of respondents' satisficing rather than optimizing (Krosnick 1991). They are acquiescence, non-differentiated answers, selection of scale extremes, no-opinion responses, and selection of middle answers. I will examine the relationship between each set of response propensity scores and each of the measurement error indicators. I predict that the response propensity scores are independent of the response quality indexes (measurement error indicators) according to the independent process model.

The Study

The study was a national telephone survey, conducted by Schulman, Ronca & Bucuvalas, Inc. (SRBI) on behalf of the Joint Program in Survey Methodology and the National Science Foundation (NSF). NSF has conducted a nationwide telephone survey to measure public knowledge of and attitudes toward science for over thirty years. The 2003 JPSM Practicum Survey was designed to gain a better understanding of the effect of different modes of administering the survey on NSF's measurements of Americans' knowledge of science. In addition, some new measures of scientific knowledge were included in order to explore the utility of these measures for future NSF surveys. Responses were obtained via a self-administered Internet questionnaire as well as through the traditional mode of a telephone interviewer-administered questionnaire. Data collection was carried out from July through September of 2003.

A national random digit dial (RDD) 3+ list-assisted sample of 12,900 phone numbers was obtained from Survey Sampling, Inc. (SSI) for the study. Under this sample design, a systematic sample of blocks with at least three residential listings was selected with equal probability across all eligible blocks, stratified by county. Once a block was selected, a two-digit number from 00-99 was systematically selected and appended to the exchange and block, to form a 10-digit telephone number. The sample was pre-screened to identify not-in-service or business numbers. The target population for the random sample was the adult (18 or older) non-institutional, non-military population of the United States in the contiguous 48 states. No provisions were made for non-English speaking respondents, or respondents using a telephone device for the deaf, so these portions of the population are not represented in this survey.

All sampled households were called in an attempt to conduct a screening interview with the adult household member with the most recent birthday. The screener asked whether the selected adult used the Internet at home or at work. Based on responses to these questions, respondents were assigned to a main interview survey condition. For those without Internet access, the main interviews were conducted by telephone. Persons with Internet access were randomly assigned to complete the main interview by telephone or via the Internet. Up to 20 call attempts were made to maximize contact and response rates with sampled households.

The final response rate (unweighted) for the screener was 42.5 percent (AAPOR Response Rate 3). The overall main survey response rate was 74.2 percent, leading to a combined final response rate of 31.5 percent. For the telephone main survey, the response rate was 97.7 percent (conditioning on completion of the screener), compared to the Internet main survey response rate of 51.6 percent. A total of 1,548 completed interviews were obtained, including 530 from Internet users interviewed by telephone, 546 from Internet users who completed the Internet survey, and 472 non-Internet users interviewed by telephone.

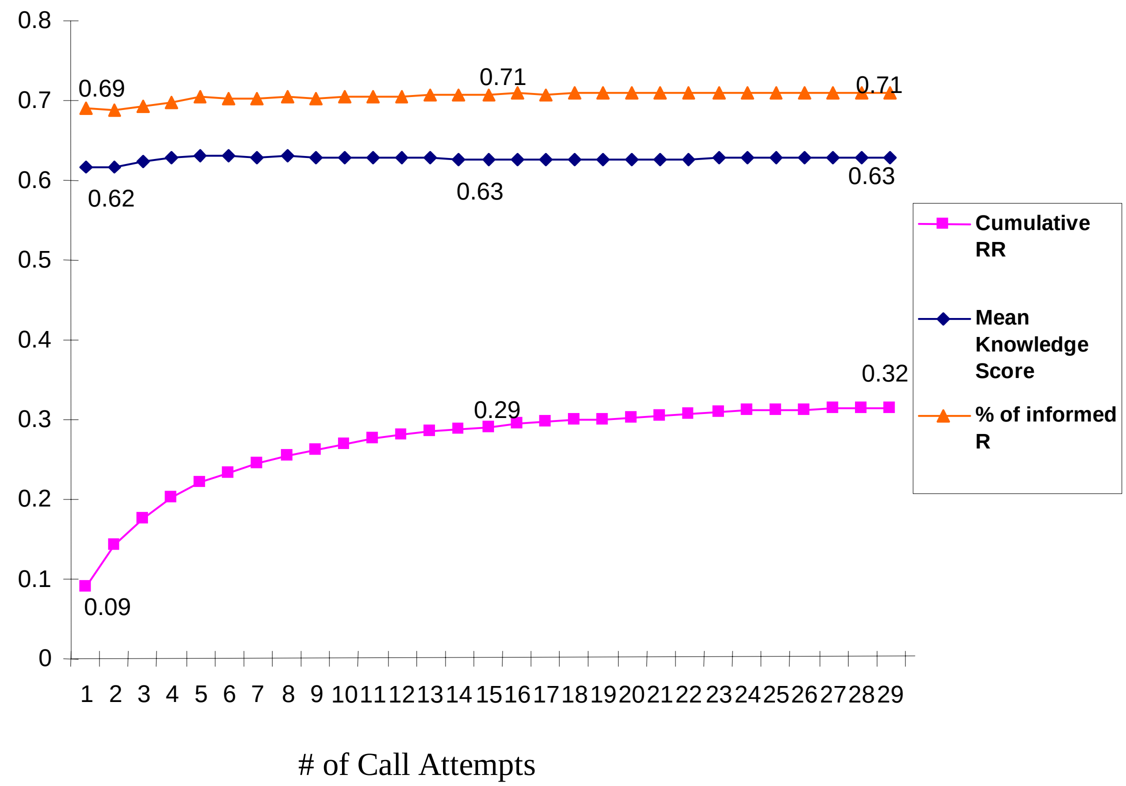
The duration of interviews varied substantially by survey condition. For respondents assigned to the telephone survey, the median screening interview length was 2.1 minutes, compared to a median length of 4.7 minutes for those assigned to the Internet survey (the additional time reflected the need to request and record an email or physical mailing address to convey the web address and PIN for the web surveys). The main interview lengths also differed by mode. The median length for telephone surveys was 19.7 minutes, while for Internet surveys the median was 25.7 minutes, as Internet respondents were free to complete the survey over the course of several hours or days.

Findings

1. Response rate and key study variables

Since the purpose of the study is to measure public knowledge of and attitudes toward science, one key survey variable is the science knowledge score. The survey included some 23 knowledge questions; we examined knowledge scores as the proportion of correct answers to those 23 questions. A key determinant of nonresponse and measurement error is public interest in and attitudes towards science and technology. Unfortunately the public interest was not directly measured in this study. Therefore, I examined a proxy measure that asks respondents about how well informed they are about science and technology, assuming a positive correlation between people's interest in science and informedness about science. I coded the level of informedness as 1 if respondents answer either 'very informed' or 'somewhat informed' to the relevant three questions and 0 if otherwise. Figure1 exhibits the plot of the mean knowledge score and the percentage of respondents who consider themselves informed about science by the cumulative response rate.

Fig.1. Knowledge Score, % of Informed Respondents and Cumulative Response Rate by # of Call Attempts



The figure clearly demonstrates that as the number of call attempts increases, the mean knowledge score doesn't vary much; neither does the percentage of informed respondents. Therefore, the extra call attempts increased the response rate but didn't change the aggregate value of the two key variables very much.

2. Univariate distribution of key study variables by respondents and by different types of nonrespondents

I divided the sample into respondents and nonrespondents in three different ways. The first contrasted easy-to-contact respondents with noncontacts: I classified respondents who completed the interview with less than 7 call attempts as easy-to-contact and those who completed after 7 or more call attempts as hard-to-contact group. I also distinguished respondents who completed without refusing (the "amenable" group) from those who ever refused at any stage ("refusals"). Last, I classified those who completed after two months into the field as late respondents and the rest as early respondents. I used the same criteria to construct the three sets of response propensity scores.

Figure 2 shows the box plot of knowledge scores under different conditions (mean knowledge scores are shown in Table 1 below.) The left graph contrasts the box plots of knowledge scores for early (left) and late respondents (right). The middle graph displays the plots for the easy-to-contact (left) and the hard-to-contact (right) groups. The right graph shows the contrast between the amenable group (left) and the refusal group (right). It is apparent that the distributions are quite similar across groups except for refusal respondents who had somewhat higher knowledge scores than the amenable group. The hypothesis that the univariate distribution of key study variables doesn't vary much across the three contrasts is supported.

Fig. 2. Box plots of knowledge scores across different groups

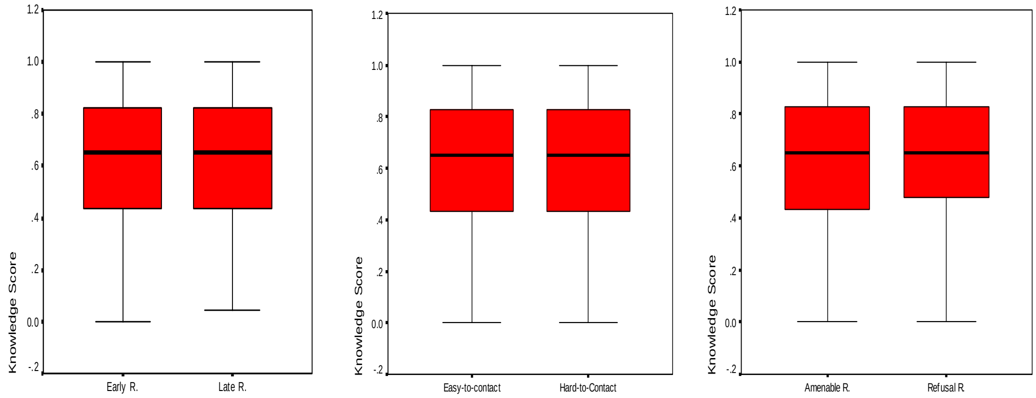
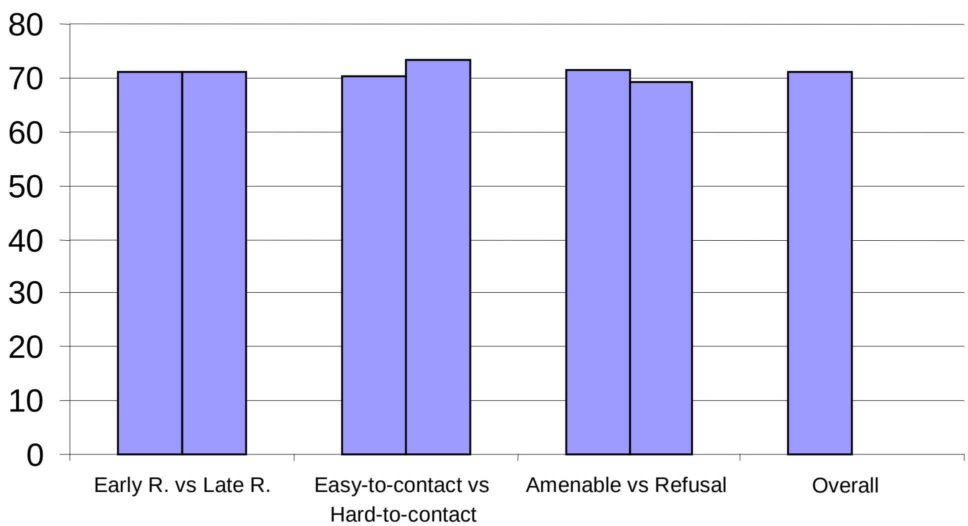


Table 1: Mean knowledge score (% correct) across conditions

	Early R.	Late R.	Easy-to-contact R.	Hard-to-contact R.	Amenable R.	Refusal R.	All
Means (N)	62.9% (1403)	61.2% (145)	62.9% (1207)	62.3% (341)	62.4% (1125)	63.8% (423)	62.8% (1548)

The same trend is observed in Figure 3, which presents the proportion of informed respondents across the different groups. The results also support the first hypothesis. Multiple callbacks and other measures to boost response rate didn't bring in respondents who differed much from those who were in without much effort.

Fig.3. % of Informed Respondents Across Groups



3. Response propensity scores and Measurement error indicators

The main analyses examine the relations between the propensity scores and measurement error indicators. Propensity scores were estimated through a logistic regression model:

$\text{Log}(p/1\text{-}p)=b_0 + \underline{bX}$

where p=pr(R=1|$\underline{X}$) is the propensity score for being a respondent (R=1) given a matrix $\underline{X}$ of covariates. As pointed out in the previous section, the nonreponse propensity is equal to 1-p and is used in later analysis.

The right side matrix includes age, gender, race, education, number of college-level science courses, web access, number of children, employment state, level of urbanicity (urban/suburb/rural), Census region, mode of interview, and knowledge score. The 12 covariates have different effects on the different nonresponse propensities. Covariates such as number of children, employment state, urbanicity and Census region influence a sampled person's at-home pattern and hence contactability; covariates such as age, education, gender and number of college level science courses may determine a sample respondent's knowledge and interest in the survey topic (science and technology) and accordingly affect their decision to refuse or to agree to the survey request. Including all 12 covariate variables in the three logistic regression models does run the risk of overfitting a model. Nonetheless, Rubin and Thomas (1996) suggest including all covariates, even if some are not statistically significant in predicting the outcome variables, unless they are unrelated to the treatment outcomes or inappropriate for the model. Table 2 lists the point-estimates for the covariates in the three models as well as the statistics for goodness of fit of the models.

Table 2: Logistic Regression Models (Parameter Estimates and SE)

Table 2: Logistics Models (Bhat and SE)	Noncontact Propensity Model		Refusal Propensity Model		Late Propensity Model	
	Estimate	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
Age	-0.18	0.17	-0.25	0.16	-0.17	0.25
Gender	-0.10	0.13	0.04	0.13	-0.29	0.19
Education	0.24*	0.12	0.00	0.11	0.06	0.17
# science courses	0.01	0.01	-0.01	0.01	0.02	0.02
Race	0.01	0.05	-0.02	0.05	-0.04	0.08
Web access	0.36**	0.19	0.61*	0.18	0.96	0.26*
# of children	-0.01	0.04	0.07**	0.03	-0.03	0.06
Knowledge Score	-0.86*	0.38	-0.04	0.36	-0.23	0.55
Urban vs other	-0.10	0.19	-0.04	0.18	-0.07	0.27
Suburb vs other	0.12	0.16	0.11	0.15	0.16	0.23
NE vs other	-0.08	0.20	0.08	0.19	0.04	0.29
MW vs other	-0.46*	0.20	-0.35**	0.18	-0.26	0.29
South vs other	-0.05	0.18	-0.20	0.17	-0.01	0.26
Mode	0.11	0.15	0.96*	0.14	0.12	0.23
Employment Status	0.51*	0.17	0.05	0.15	0.21	0.24
Likelihood Ratio	33.942 (p=.0004)		69.248 (p<.0001)		22.695 (p=.091)	

*statistically significant predictors at p<.05 **marginally statistically significant at p<.10

Two simple methods are used to relate propensity scores to measurement error indicators. One method is to regress the propensity scores on the five response quality measures. I used a set of 10 questions on people's attitudes towards different science and technology developments to assess response quality. The acquiescence measure was the percentage of agreeing or supporting answers within the 10 questions. The non-differentiation measure was the number of identical answers among the 10 questions. In the same way, the measure of the selection of scale extremes, no-opinion responses and mid-point answers were counts of the number of extreme answers, no-opinion answers, and middle answers within the ten questions.

Table 4 displays the simple regression coefficients of the propensity scores on measurement error indicators.

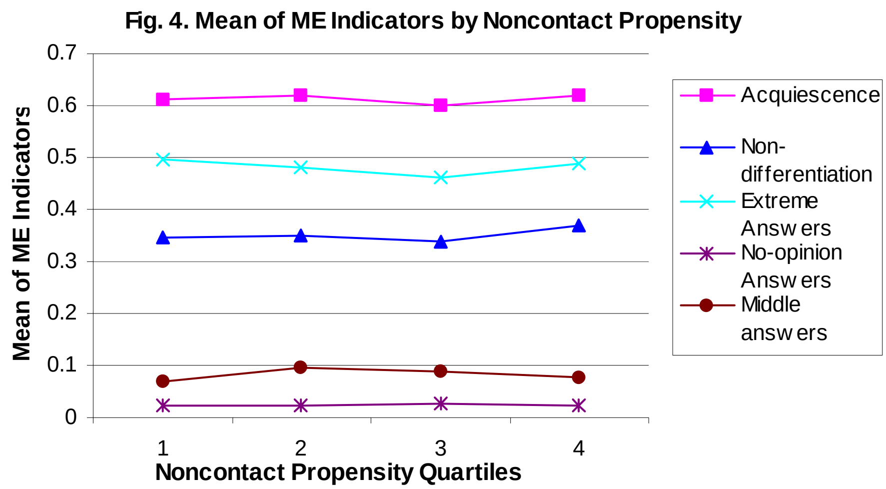
Table 4: Regression Coefficients on Measurement Error Indicators

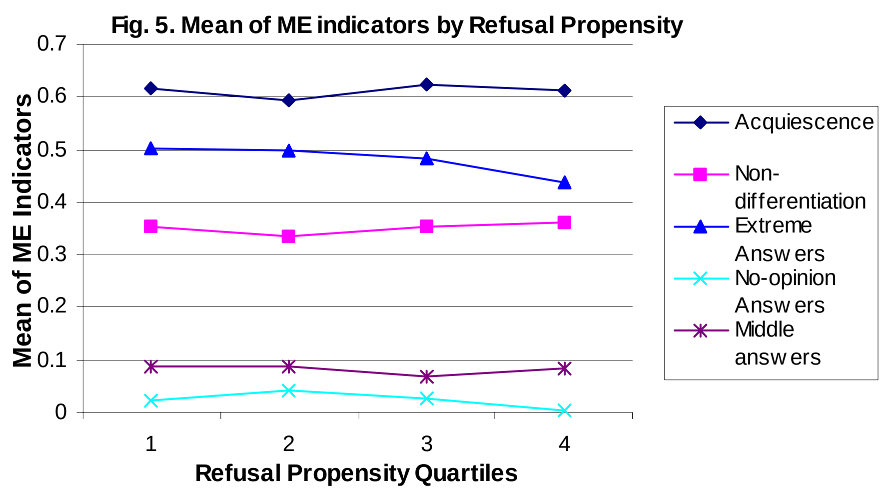
	Acquie-scence	Non-differentiation	Extreme Answers	No-opinion Responses	Mid-point answers
Noncontact Propensity	-0.04	0.13**	-0.06	-0.01	0.02
Easy vs. hard to contact	-0.02*	0.02*	-0.01	0.00	0.00
Refusal Propensity	0.01	0.04	-0.28*	-0.10*	-0.04
Amenable R. vs Refusal	0.00	0.01	-0.03*	-0.01*	-0.00
Late Propensity	-0.17	0.18	0.09	0.26*	-0.30*
Early vs Late	-0.01	0.01	-0.02	0.00	0.01

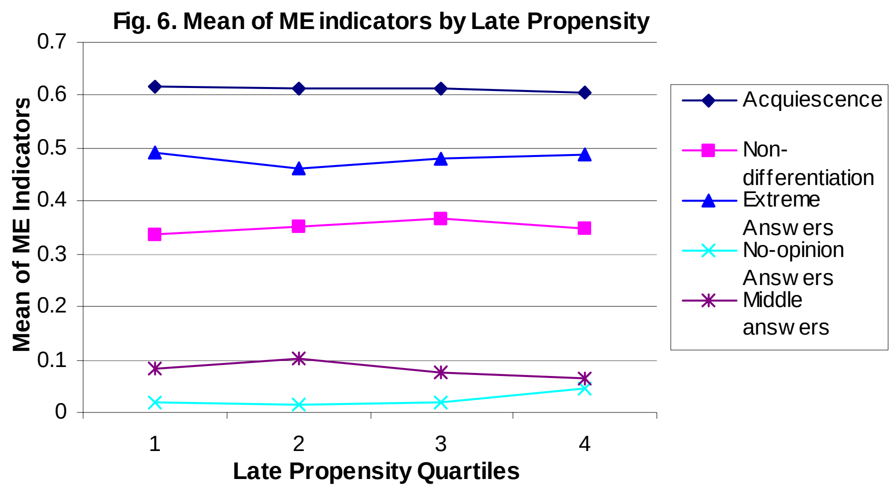
*statistically significant at p<.05 **marginally statistically significant at p<.10

A few inferences can be drawn from the table. First of all, in most of the cases (10 out of 15 coefficients), the propensity scores are not significantly related to the measurement error indicators. Second, among the five significant coefficients, three are also apparent when I examined the differences between the two respondent groups that the propensity score models. For the other two cases where there seemed to be a relationship between propensity scores and measurement error indicators, however, the relationship is not observed if the simple regression was done on the group variable (i.e., where a 0-1 indicator replaces the propensity score on the right side of the regression models). The propensity scores seemed to be more sensitive to reveal relationships with measurement errors than the dichotomous variables they model. Third, the table shows that different nonresponse propensities are related to different measures of response quality. Acquiescence seems to be the only type of measurement error unrelated to nonresponse propensities. Furthermore, noncontact respondents seemed to be more likely to give nondifferentiated responses whereas refusal respondents tended to give less extreme answers and fewer no-opinion responses. The late respondents, by contrast, produced more non-opinion responses but fewer midpoint answers.

I also examined the “coarsened” response propensities, classifying cases with similar response propensities into four propensity groups (lowest propensity quartile, second lowest quartile, second highest quartile, and highest quartile). Then the means of the various indicators were compared across the propensity groups. The next three figures display the plots.







The figures tell pretty much the same story about the relationship between response propensities and measurement error as the earlier analysis did. They also disclose the presence of curvilinearity in the relationship between nonresponse propensities and some measurement errors, which disconfirms both the common cause model and intervening cause model.

Discussion

This paper provides further empirical evidence for the relationship between nonresponse and response quality. Consistent with the findings by Curtin et al (2000), Keeter et al (2000), and Merkle and Edelman (2002), extra attempts to boost response rate didn't bring in respondents who differed sharply from the early respondents; the higher response rate resulting from added callbacks and refusal conversion efforts didn't change the distribution of key study variables very much (see Fig.1.) In addition, the study supported the independent process model that hypothesizes little or no relationship between response propensities and measurement error. In cases where we do see a non-zero relationship, however, the correlations are in the desired direction, indicating better data quality for reluctant respondents. Overall, the study seemed to contradict two traditional beliefs held by survey researchers, suggesting that 1) higher response rates do not necessarily reduce nonresponse error, and 2) reluctant respondents are not necessarily poorer reporters than easy respondents.

However, given the size of the study (1548 completed interviews in total), there are inevitably limitations. Firstly, the propensity scores are determined through call attempts and call history. But such models might be weak at capturing the response propensities. It would be a breakthrough if response propensity models can be established through other kinds of measures. Secondly, only five response effects are examined as indicators of response quality. We have no single overall response quality measure based on the accuracy of the answers. Accordingly, future studies should utilize other indicators of response quality. Thirdly, only ten question items were used in constructing indictors of response quality and all of them were attitude items. As a result, the external validity of the findings is low. Future research should focus on the impact of response propensities using the overall item-missing rate and/or overall response quality measures.

Reference

Bollinger, C, and M. David. 2001. "Estimation With Response Error and Non-response: Food Stamp Participation in the *SIPP." Journal of Business and Economic Statistics* 19: 129-142.

Cannell, C. F., and F. J. Fowler. 1963. "Comparison of a Self-Enumerative Procedure and a Personal Interview: A Validity Study." *Public Opinion Quarterly* 27: 250-264.

Curtin, R., S, Presser, and E. Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64: 413-428.

Groves, R. 1989. *Survey errors and survey cost*. New York: John Wiley and Sons.

Groves, R., and M, Couper. 1998. *Nonresponse in household interview surveys*. New York: John Wiley and Sons.

Groves, R., E. Singer, and A. Corning. 2000. "Leverage Salience Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64: 299-308.

Keeter, S., C. Miller, A. Kohut, R. Groves, and S. Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64: 125-148.

Krosnick, J. A. 1991. "Response Strategies for Coping With the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology 5*: 213-236.

Merkle, D., and M. Edelman. 2002. Nonresponse in Exit Polls: A Comprehensive Review." In *Survey nonresponse*, ed. R. Groves, D. Dillman, J. Eltinge, and R. Little. New York: John Wiley and Sons.

Rosenbaum, P. and D. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79: 516-524.

Rubin, D. B. and N. Thomas. 1996. "Matching Using Estimated Propensity Scores: Relating Theory to Practice." *Biometrics* 52: 254-268.

Willimack, D. K., H., Schuman, B., Pennell, and J. M. Lepkowski. 1995. "Effects of A Prepaid Nonmonetary Incentive on Response Rates and Response Quality in a Face-to-Face Survey." *Public Opinion Quarterly* 59: 78-92.