Are Nonrespondents Necessarily Bad Reporters? Using Imputation Techniques to Investigate Measurement Error of Nonrespondents in an Alumni Survey

Mandi Yu[1], Ting Yan[2]
[1]Program in Survey Methodology, University of Michigan, 426 Thompson St, 4050, Ann Arbor, MI, 48104
[2]National Opinion Research Center (NORC), 1155 East 60th Street, Chicago, IL, 606137

Abstract

Declining response rates in household surveys speak to the increasing difficulty in recruiting sampled persons to respond to a survey request. Survey organizations are spending increasingly more effort on contacting and persuading sampled persons with a low response propensity in an attempt to increase response rates and to minimize nonresponse error. A major concern, however, is that, if sampled persons with a low response propensity turned out to be bad reporters, the costly extra recruitment effort could reduce nonresponse error at the expense of increasing measurement error. However, this concern is difficult to address since the measurement error property of nonrespondents is unknown. This paper treats this issue as a missing data problem and imputes for values of survey reports for nonrespondents. The results showed that the relationship between response propensity and measurement error depends on the cause of nonresponse. The report quality was not worse for nonrespondents due to non-contact. However, nonrespondents due to refusal tend to produce less stable reports than cooperative respondents.

KEY WORDS: Measurement Error, Nonresponse bias, Multiple Imputation

1. Introduction

Household surveys have been experiencing a falling response rate over the past few decades (Atrostic, Bates et al. 2001; De Leeuw and de Heer 2002; Curtin, Presser et al. 2005). The declining response rates speak to the increasing difficulty that survey organizations have in contacting sampled persons and persuading them to participate in the survey once contacted. The danger with lower response rates is the presence of nonresponse bias if sampled persons with low response propensities systematically differ from those with higher response propensities with regard to survey variables of interest. To minimize potential nonresponse bias, survey organizations invest extensive resources on contacting and obtaining cooperation from nonrespondents in order to boost response rates. However, these additional efforts for recruiting reluctant individuals tend to be expensive; Mason, Lesser, and Traugott (2002) shows that, in a voting survey, the cost for interviewing one reluctant respondent was 1.6 times as many as the cost of interviewing a respondent without refusal conversion effort. The same study also reports that the refusal conversion efforts took one third of their budget but provided only a quarter of cases for the sample.

In addition to cost concerns, there has been a data quality concern about persuading sampled persons with low response propensities. The concern that people who are hard to recruit might be worse reporters than the cooperative respondents dates back to the 1960s (Cannell and Fowler 1963). If there existed a negative correlation between response propensity and measurement error, then the costly extra recruitment effort could reduce nonresponse error at the expense of increasing measurement error. Therefore, it is critical to have a better understanding of the link between response propensity and measurement error. Such an understanding will help survey organizations to make better trade-offs between nonresponse and measurement error and to make judicial decisions on when to stop the field period.

1.1 Theoretical Link between Nonresponse and Measurement Error

The research on the relationship between nonresponse and measurement error has increasingly caught the attention from survey researchers (see Olson (2006), for the latest work). At least three models can be posited to explain the nature of the relation between nonresponse and measurement error. The first model – the independence model – assumes that the two sources of error may be uncorrelated. The assumption is that nonresponse is mostly motivational whereas measurement error is primarily cognitive; therefore, increasing an individual's response propensity wouldn't affect his/her accuracy in reporting. By contrast, the common cause model posits that there exist some common factors that affect both sample persons' propensity to participate in a survey request and their measurement error property. 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 that person to underreport on, say, his/her drug use even if he/she 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 third model—the intervening cause model—posits that response propensities change some internal state of sample persons that subsequently affect their behaviour as respondents. For instance, the decision to participate in a survey, induced either by extra monetary incentives or successful refusal conversion efforts, affects people's willingness to provide accurate data, which subsequently affects their data quality.

The empirical findings on the relation between nonresponse and measurement error is at best mixed. While some studies supported a causal link between response propensity and measurement error by showing that early respondents provided more accurate data or greater response completeness than late respondents (Cannell and Fowler 1963; Willimack, Schuman et al. 1995), other studies failed to find support for the hypothesis that the more difficult and reluctant sample persons produced less accurate or less useful data than those who were easier to recruit or more amenable (Green 1991; Yan, Tourangeau et al. 2004; Olson 2006; Olson and Kennedy 2006). Olson (2006) demonstrated that the relationship between nonresponse propensity and measurement error is statistic-specific; within one study, she found evidence for a negative correlation between nonresponse and measurement error on only some statistics, but not all statistics of the study.

1.2 Challenges In Studying the Relation of Nonresponse and Measurement Error

At present, research on the nature of the relation between nonresponse and measurement error is subject to two inherent difficulties. First, true values are needed to examine measurement error properties. However, true values are hard to get in most of the cases (why do we need to conduct a survey if the true values are already known?) and nonexistent in other cases (the existence of true values for attitudinal measures is a controversy). Second, by definitions, nonrespondents don't participate in the survey and don't provide answers to survey questions; thus, their measurement error properties can not be evaluated even in the presence of true values. Because of these inherent problems, most of the empirical studies cited earlier employed a level-of-effort analysis approach. Respondents who were hard-to-contact (e.g., those respondents who needed extra call attempts in telephone surveys) are treated as proxy for non-contact nonrespondents and respondents who initially refused to the survey request but were later converted are treated as proxy for refusal nonrespondents. The implicit assumption of this approach that respondents who are hard to recruit are more like nonrespondents is open to debate.

Using a dataset of rich validation information, this paper proposes to employ the imputation techniques to solve the second difficulty. That is, we treat nonresponse due to non-contact and refusal as a regular missing data problem and fill in the missing values with draws from one nonparametric imputation model – the alternating conditional expectations (ACE) regression technique (Breiman and Friedman 1985). This paper investigates the feasibility of using imputation techniques to evaluate measurement error of nonrespondents. We first impute the report values for non-contact nonrespondents and refusal nonrespondents and then evaluate the measurement error properties of these derived values against the reported values from respondents.

2. Methods

2.1. Data set-up for Imputation

Let Y^* be the survey variable of interest. Let Y be the validation record information for this variable and X be the validation auxiliary variables. (Y, X) are available for all sampled units.

Table 1: Missing Patterns and Imputation Models

			Survey data	Auxiliary Data	
			Y^*	Y	X
	Easy-Cooperative Respondents	N1			
Imputation model 1	Hard-to-contact Respondents	N2	Observed	Observed	Observed
	Non-contact	N3	Missing	Observed	Observed
Imputation model 2	Reluctant Respondents	N4	Observed	Observed	Observed
	Refusals and Item missing Respondents	N5	Missing	Observed	Observed

Observed: ▤ Missing: ▥

We impute for values of survey reports for non-contact nonrespondents and refusal nonrespondents separately since nonresponse due to non-contact and nonresponse due to refusal are considered as two major different dimensions of survey nonresponse (Groves and Couper 1998). To set up the imputation model for non-contact nonrespondents, we split the full sample into easy-cooperative respondents (N1), hard-to-contact respondents (N2), and non-contact respondents (N3). As shown in Table 1, we fit an imputation model to generate multiple imputations on Y^* for non-contact nonrespondents using data components (N2+N3), relying on the assumption that hard-to-contact respondents are treated as a proxy for non-contact people. Similarly, to impute for refusal nonrespondents, we separate respondents who refused once but were later converted to participate in the survey (N4), from refusal nonrespondents who failed to complete the survey (N5). Data components (N4+N5) are used to impute for refusal nonrespondents, again treating reluctant respondents as a proxy of refusals.

2.2. Imputation Model

The alternating conditional expectations (ACE) regression based imputation is used in this paper because of its advantages over linear regression based imputation models. First, ACE protects model misspecification by making few assumptions about the normal distributions and functions between response and predictor variables. It fully explores and explains the complex relationships between response and predictor variables. In addition, the imputed values are always within the observed data range, protecting the multiple imputation inference from being unintentionally affected by extreme values.

Imputations for non-contact nonrespondents and refusal nonrespondents are conducted independently using the ACE method. We carry out 20 multiple imputations for each imputation model to tolerate large missing data and to incorporate the uncertainty in estimating the ACE models. A Bayesian bootstrap sample for complete data of the same size is created and the ACE transformation functions are estimated based on the Bayesian bootstrap sample. Predictions on Y^* for nonrespondents are obtained conditional on their covariates and the estimated transformation functions plus random draws from observed ACE regression residuals.

Specifically, each imputation is created in the following three steps:

1) Creating Bayesian bootstrap samples of size r or a multiple of r to simulate the population.
2) Estimating the transformation functions ϕ , θ and obtains the residual z using complete data based on the following model

$$\psi(Y_i^*) = \alpha + \varphi(Y_i) + \sum_{j=1}^{p-1} \phi_i(X_{ij}) + z_i$$

$$z_i \sim N(0,1); i = 1,2,...,r; j = 1,2,..., p-1$$

where,

(Y^*, Y) denotes the reported Y values and the Y values from the administrative data respectively;

r : the sample size of hard-to-contact respondents (N2 in Table 1) in the imputation model for non-contact nonrespondents, and sample size of reluctant respondents (N4 in Table 1) in the imputation model for refusal nonrespondents.

p : the number of covariates entered in the model.

3) Estimating Y^* for nonrespondents conditional on their characteristics on Y and X :

$$Y_i^* = \psi^{-1}\left(\varphi^{-1}Y_i + \sum_{j=1}^{p-1} \phi_i^{-1} X_{ij} + \hat{z}_i\right); i = r+1, r+2, ..., n$$

where $\hat{z}_i$ is the random draw from estimated regression residual $\hat{Z}$ (with length r).

Repeat step 1)-3) $M = 20$ times independently to generate multiple imputations for the missing data.

2.3. Multiple Imputation Inference

For each imputed data $D_l, l = 1,2,...,M$ where M is the total number of multiple imputations, a scalar estimand θ, which may be a function of (Y, Y^*), is estimated with a point estimate $\hat{\theta}_l$ and its associated variance $\hat{v}_l$. Under the assumption described in Rubin (1987), θ can be estimated by $\overline{\theta} = \sum_{l=1}^{M} \hat{\theta}_l / M$ with variance $T = (1+1/M)b + \overline{v}$, where the between imputation variance $b = \sum_{l=1}^{M} \left(\hat{\theta}_l - \overline{\theta}\right)^2 / (M-1)$ and the within imputation variance $\overline{v} = \sum_{l=1}^{M} \hat{v}_l / M$.

3. Data

The dataset used in this paper is from the 2005 JPSM (Joint Program in Survey Methodology) practicum study. The study is a survey of University of Maryland alumni. The sample was drawn from the 55,320 graduates who received undergraduate degrees from the University of Maryland from 1989 to 2002, as reflected in the records maintained by the Office of the Registrar. The survey interviews were conducted in August and early September 2005. The Registrar's records were used to select a random sample of 20,000 graduates, stratified by graduation year. Of these 20,000 graduates, 10,325 could be matched to Alumni Association records containing telephone contact information. After clean-up and pretest, 7,957 phone numbers were eligible for the survey. Call attempts were made to 7,591 numbers. With 24 deceased alumni the denominator of the AAPOR response rate shrinks down to 7,567.

The alumni were initially contacted by telephone and administered a brief set of screening questions to verify that the interviewer has reached the correct person. A total of 1,501 alumni completed the screener and were randomly assigned to one of the three modes of data collection (CATI, IVR, and web). For those without access to the web, the random assignment was restricted to CATI versus IVR. 1,094 alumni started the main questionnaire and 1,003 completed interviews were obtained. Treating the completion of the screener and the completion of the main questionnaire as two independent events, a response rate of 13.25% (AAPOR RR1) is obtained, as shown in table 2.

Table 3 displays the actual sample sizes for each data component listed in Table 1. We adopted the same operationalization as in Lynn et al. (2000) and defined hard-to-contact respondents as those respondents who took 7 or more calls to complete the interview. Easy cooperative respondents, by contrast, are those who readily answered the survey request and completed the interview with 6 or fewer number of calls. Reluctant respondents are those who initially refused to participate and required refusal conversion efforts to complete the interview.

Table 2**.** The Response Rate and the Number of Completes by Mode

	Total	***Percent***
Alumni eligible & number dialed	7,560	100
Screener completion	1,501	19.84
Initially assigned	1,501	100
Started main questionnaire	1,094	72.88
Number of completes	1,003	66.82
Response Rates (AAPOR RR1)	13.25 (=66.82*19.84)	

Table 3. Actual Sample Sizes for Each Data Component Listed in Table 1

		Data Components (as in Table 1)	*Total*	***Percent***
Total Sample in this Study			7,425	100
Imputation Model 1	Easy-cooperative respondents	N1	619	8.34
	Hard-to-contact respondents	N2	224	3.02
	Non-contact nonrespondents	N3	5,245	70.64
Sample size used for imputation 1		N2+N3	5,469	73.66
Missing fraction for Imputation 1		N2/(N2+N3)		96
Imputation Model 2	Reluctant respondents	N4	51	0.69
	Refusal nonrespondents	N5	1,286	17.32
Sample size used for imputation 2		N4+N5	1,337	18.01
Missing fraction for imputation 2		N4/(N4+N5)		96

Non-contact nonrespondents are defined as those sampled persons who were never contacted and refusal nonrespondents are those who were contacted but showed resistance to the survey request and didn't complete the interview. There are 96 respondents who completed the survey but had missing data for Y. We treated them as refusal nonrespondents for the purpose of imputation (they were grouped under N5 in Table 3).

Given the nature of the survey, we have the luxury of registrar data including student's GPA at school. The registrar data are treated as true values that allow us to evaluate the measurement error properties. The survey questionnaire contained questions asking alumni to report their GPAs. We chose the survey reports on GPA as the variable of interest in this paper. The covariates include the true GPA scores obtained from the registrar data as well as age, gender, race, graduation year, degree types, alumni membership, whether ever donate to the school and other administrative academic performance measures such as whether students graduated with honor, whether students failed, dropped or withdrawn a class and whether ever on probation. The covariate variables are available for all sampled students whether they completed the survey or not.

4. Results

4.1. Observed GPA Reports

Figure 1 displays the histograms of reported GPA for the three sample groups with differential response propensities: easy-cooperative respondents, hard-to-contact respondents and reluctant respondents. The mean GPA's are not significantly different from each other based on the two-sample t tests. The histograms also suggest that it is reasonable to treat the GPA scores as a continuous variable. Besides, the ACE imputation method tolerates normality violation to a certain extent. We felt reasonably assured about our imputations.

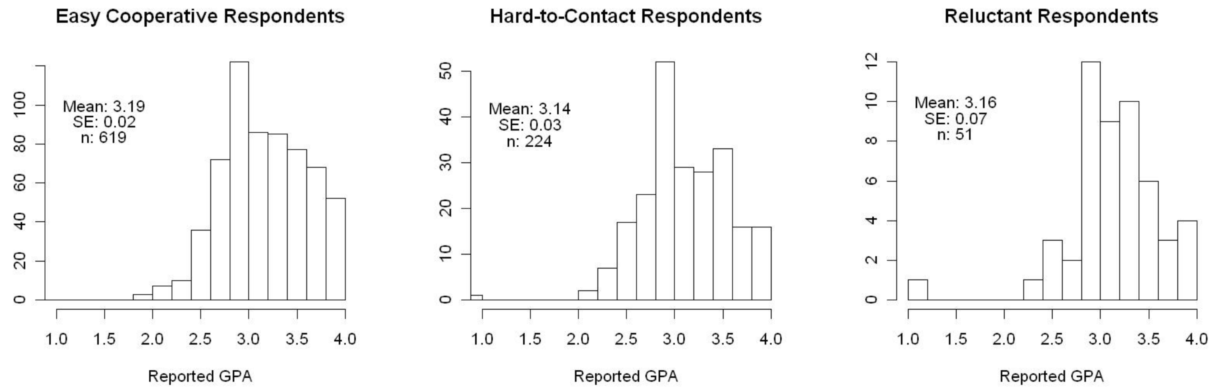

Figure 1: Reported GPA for Respondents with Different Response Propensity

4.2. Evaluating the Reporting Quality for Respondents

The reporting quality of a survey question can be evaluated in terms of both bias and variance. Measurement bias exists if reported values systematically overstate or understate the true value. The presence of the systematic errors biases almost all statistical estimates. Random error, instead, is a non-systematic measurement error which averages to zero over repeated measures. Random error doesn't necessarily bias the estimation of descriptive statistics such as means and sums, but it reduces estimation precision by inflating the measurement variance. The presence of measurement variance biases most regression analysis (Biemer 1991).

Under the simple measurement error model setup, the reported values on a variable can be decomposed into

$$X_{i,obs} = u_{i,tru} + e_i; i = 1, 2, ..., n$$

where

$X_{i,obs}$ = response obtained for the i-th person

$u_{i,tru}$ = true value for the i-th person

e_i =deviation for the i-th person from its true value

Therefore, the measurement bias is estimated by the mean of the raw measurement error, $\theta_{raw} = \sum_{i=1}^{n} e_i / n$ and measurement variance is defined as the mean of absolute measurement error $\theta_{abs} = \sum_{i=1}^{n} |e_i| / n$.

Table 4 shows the means of raw measurement errors and absolute measurement errors for easy-cooperative respondents, hard-to-contact respondents and reluctant respondents. All respondents, regardless of their overall survey response propensities, tend to over report their GPAs. The positive biases are significantly different from 0 for easy-cooperative and hard-to-contact respondents. In terms of the means of absolute measurement errors, there is no difference among these groups. However, there is a tendency for reluctant and hard-to-contact respondents to have larger measurement variance than easy-cooperative respondents in their reports of GPA. We next examine whether the same trend of is present in reporting quality between nonrespondents and respondents.

Table 4: Mean of Measurement Error and Absolute Measurement Error

	Easy-cooperative Rs	Hard-to-contact Rs	Reluctant Rs
Raw Measurement Error (SE)	.09(.01)	.09(.02)	.07(.06)
Absolute Measurement Error (SE)	.15 (.01)	.17(.02)	.21(.05)
N	619	224	51

4.3. Evaluating Reporting Quality for Nonrespondents

Using the Multiple imputation inference techniques described in the earlier section, the point and interval estimates for θ_{raw} and θ_{abs} for the non-contact and refusal nonrespondents are obtained. Consistent with results shown in table 4, all nonrespondents also tend to over-report their GPA regardless of their contact propensity or cooperation propensity. In addition, respondents and nonrespondents do not differ significantly in the extent of their overreport. Table 5 demonstrates that refusal nonrespondents are associated with a higher variance in reporting GPAs than respondents and the difference is significant. Non-contact nonrespondents also tend to be more unstable in reporting than easy-cooperative respondents; however the differences are not statistically significant.

Table 5: Mean of Reporting Error and Absolute Reporting Error

	Easy-cooperative Rs	Refusal NRs	Non-Contact NRs
Raw Measurement Error (S.E.)	.09 (.01)	.09(.06)	.10(.04)
Relative Efficiency	-	20.55	86.56
Absolute Measurement Error (S.E.)	.15(.01)	.35(.04) ***	.21(.04)
Relative Efficiency	-	18.36	181.79
Sample size	619	-	-

The results on both the raw and absolute measurement error suggest that refusal nonrespondents and non-contact nonrespondents do not seem to be reporting less accurately than easy-cooperative respondents. However, nonrespondents do give less stable reports than easy-cooperative respondents in terms of measurement variance.

5. Discussions

Our analyses showed that the reporting quality of GPA is quite high for all respondents regardless of their cooperation and contact propensities in terms of both measurement bias and measurement variance. The bias ratio (the mean ratio of measurement bias over the true GPA) is expected to be roughly around 3%, which may not be significant enough to make a difference in policy makings.

Furthermore, the measurement bias of the imputed responses for both types of nonrespondents was not significantly worse than those of respondents. There is evidence, however, that responses from nonrespondents are less stable than those of respondents. Specifically, refusal nonrespondents are prone to larger measurement variance than easy-cooperative respondents. Our results seem to suggest that the refusal nonrespondents might be less stable reporters, but noncontact nonrespondents are not worse than their counterparts who are recruited without extensive efforts. Therefore, the relationship between response propensities and measurement error is dependent on the underlying nonresponse mechanism. However, we would like to caution the readers about the generalizability and robustness of our findings. The dataset used in this paper is from an alumni survey of college graduates. Thus, our results might not apply to surveys of general population. In addition, the *Y* variable of interest is GPA scores and it is not a question frequently asked in surveys of general population.

This paper demonstrates an example of using imputation method to study the relation between nonresponse and measure error. We also show the flexibility and power of the ACE imputation method for variables that violate normality assumption. We believe that this method is useful for both researchers who are interested in the theoretical link between nonresponse and measurement error and also for practitioners who need to make trade-off decisions. Future research is needed (and is strongly encouraged) to advance this line of study.

Acknowledgements

We are indebted to the Principle Investigators of the 2005 JPSM Practicum study (Profs. Roger Tourangeau, Stanley Presser, and Frauke Kreuter), who allowed us to use the practicum data for this paper.

References

Atrostic, B. K., N. Bates, et al. (2001). "Nonresponse in U.S. government household surveys: Consistent measures, recent trends, and new insights " Journal of Official Statistics **17**(2): 209-226

Biemer, P. (1991). Measurement errors in surveys. New York, Wiley.

Breiman, L. and J. H. Friedman (1985). "Estimating Optimal Transformations for Multiple-Regression and Correlation." Journal of the American Statistical Association **80**(391): 580-598.

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**(2): 250-264.

Curtin, R., S. Presser, et al. (2005). "Changes in telephone survey nonresponse over the past quarter century." Public Opinion Quarterly **69**(1): 87-98.

De Leeuw, E. and W. de Heer (2002). Trends in household survey nonresponse: A longitudinal and international comparison. Survey Nonresponse. R. M. Groves. New York, John Wiley.

Green, K. E. (1991). "Reluctant respondents: differences between early, late and nonrespondents to a mail survey." Journal of Experimental Education **56**: 268-276.

Groves, R. M. (2002). Survey nonresponse. New York, Wiley.

Groves, R. M. (2006). "Nonresponse rates and nonresponse bias in household surveys." Public Opinion Quarterly **70**(5): 646-675.

Groves, R. M. and M. P. Couper (1998). Nonresponse in household interview surveys. New York, Wiley.

Little, R. J. A. and D. B. Rubin (2002). Statistical analysis with missing data. Hoboken, N.J., Wiley.

Lynn, P. and P. Clarke (2002). "Separating Refusal Bias and Non-Contact Bias: Evidence from UK National Surveys." The Statistician **51**(3): 319-333.

Mason, R., V. Lesser, et al. (2002). Effect of item nonresponse on nonresponse error and inference. Survey nonresponse. R. M. Groves. New York, John Wiley

Olson, K. (2006). An Investigation of the Nexus of Survey Nonresponse Error and Measurement Error. Nonresponse Workshop. Omaha, NE.

Olson, K. and C. Kennedy (2006). Examination of the Relationship between Nonresponse and Measurement Error in a Validation Study of Alumni. AAPOR.

Rubin, D. B. (1981). "The Bayesian Bootstrap." Annals of Statistics **9**(1): 130-134.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York, Wiley & Sons.

Tourangeau, R., L. J. Rips, et al. (2000). The psychology of survey response. Cambridge ; New York, Cambridge University Press.

Willimack, D. K., H. Schuman, et al. (1995). "Effects of a Prepaid Nonmonetary Incentive on Response Rates and Response Quality in a Face-to-Face Survey." Public Opinion Quarterly **59**(1): 78-92.

Yan, T., R. Tourangeau, et al. (2004). When Less is More: Are Reluctant Respondents Poor Reporters? The Annual Meetings of the American Statistical Association, Toronto, CA.