8-2006

Examination of the Relationship between Nonresponse and Measurement Error in a Validation Study of Alumni

Kristen Olson
University of Nebraska - Lincoln, kolson5@unl.edu

Courtney Kennedy
Pew Research Center for the People & the Press

Proceedings of the American Statistical Association, Survey Research Methodology Section, 2006

Examination of the Relationship between Nonresponse and Measurement Error in a Validation Study of Alumni

Kristen Olson,[1] Courtney Kennedy [2]
Program in Survey Methodology, University of Michigan [1]
Joint Program in Survey Methodology, Pew Research Center for the People & the Press [2]

Abstract

While the individual components of total survey error have been well documented in the literature, relatively little is known about the intersection of these error sources. In particular, there is scant empirical work on the interplay between nonresponse error and measurement error – despite the potentially significant implications for data quality as well as techniques used to recruit respondents. In this paper we investigate the connection between these two error sources using data from a survey of University of Maryland alumni. The availability of administrative records for seven items on the survey instrument (donations, membership in the alumni association, and multiple measures of academic performance) make this dataset particularly well-suited for this type of analysis. We evaluate several causal models related to the nonresponse / measurement error nexus. These models predict differential effects for particular subgroups of the population: recent versus older graduates and alumni who demonstrated low versus high academic achievement.

Keywords: Nonresponse error, Measurement error

1. Introduction

Are respondents who are the least likely to respond to a survey request also the most likely to provide lousy data? This question has been present practically since the beginning of survey research. The underlying scientific hypothesis is that survey practitioners are making a tradeoff of nonresponse error for measurement error. Few empirical investigations of this question exist, but the importance of the tradeoff has never been more important. To compensate for declining response rates, survey organizations are investing more money in respondent recruitment efforts. The question remains as to whether this extra effort induces more error than it remedies.

Bringing in respondents who were difficult to contact or who had previously refused to participate in the survey may decrease nonresponse bias and increase the sample representativeness. However, if these respondents are also those who are the most likely to give faulty answers, and the increase in measurement error exceeds the reduction in nonresponse bias, then the mean square error of those estimates will have increased, relative to not having included those respondents in the sample. Knowledge of this relationship could drive design decisions.

Previous research suggests that there is an association between the level of effort exerted to recruit the respondent and item nonresponse, weaker associations with measurement error when compared to records, and unclear associations for measurement error as assessed by covariance structures (Cannell and Fowler 1963; Jones and Lang 1982; Rodgers and Herzog 1987; Green 1991; Martin 1994; Lin and Schaeffer 1995; Yan, Tourangeau and Arens 2004; Ygge and Arnetz 2004; Stang and Jöckel 2005). These articles do not look at the joint effects of reduction of nonresponse rates and potential increases in measurement error. For a mean square error model, it is important to look at both.

This paper investigates the connection between these two error sources using data from a survey of University of Maryland (UMD) alumni. A key feature of the study is that validation data are available for seven items on the survey instrument (donations, membership in the alumni association, and multiple measures of academic performance). The validation data allow us to study the magnitude of measurement error by comparing respondent reports against the true values. We are also able to study nonresponse bias by comparing the true values of the respondents and nonrespondents in the sample.

1.1 Common Causes for Nonresponse and Measurement Error

Measurement error will be related to nonresponse propensity and nonresponse bias when correlates of the survey variables predict both nonresponse propensity and measurement error. This is a “common cause” model for the nonresponse / measurement error nexus. The connection between these error sources is likely to depend on the type of nonresponse because the different types of nonresponse result from different causes (Groves and Couper 1998). Measurement error in surveys frequently results from memory or retrieval failures or when social desirability norms are activated. If the survey variables are correlates of measurement error and are also correlates of nonresponse propensity, then nonresponse

Papers presented at the Joint Statistical Meetings, Seattle, Washington, August 6-10, 2006.

propensity will be related to both measurement error and indicative of nonresponse bias.

Three "common causes" are posited in the alumni survey: academic performance, alumni involvement, and recency of graduation. With the first two common causes, academic performance and alumni involvement, we hypothesize that the lower academic performers and the uninvolved alumni will be less amenable to the survey request. We also expect social desirability pressure to induce misreporting among these groups will also experience social desirability pressure to misreport these variables. In particular, those low achievers who cooperate will overreport their grade point average and underreport poor grades due to social desirability pressure. Similarly, inactive alumni may falsely report being an alumni association member or having donated money to the University. We predict that we will also observe more measurement error among the low cooperation propensity groups because those who have low propensities due to poor academic performance or uninvolvement will be more likely to misreport on academic measures.

2. Study Design

The Alumni Survey was conducted August 8 through September 9, 2005 with a sample of alumni of the University of Maryland. The Alumni Survey was conducted as part of the Survey Practicum at the Joint Program in Survey Methodology. The respondents came from a stratified[1] random sample of 20,000 alumni who graduated from the University between 1989 and 2002. The sample was drawn from records kept by the Office of the Registrar and matched to the Alumni Office's records. The Alumni Office had contact information for just over half of the persons in the stratified sample, and ultimately 1,559 alumni completed the main interview (40% response rate, AAPOR RR3).

All respondents were initially recruited over the telephone and asked to participate in a survey of University of Maryland alumni. Upon completion of a short screener, respondents were assigned randomly to complete the main interview through interactive voice response (IVR), computer-assisted telephone interviewing (CATI) or on the Web. Respondents in the IVR and CATI conditions simply stayed on the line. Web respondents were offered a $20 incentive and asked for their email address so that the field house could send them a link to the survey website and a unique password. Follow-up emails were sent to nonrespondents after several days. Although the mode experiment may influence the relationship between nonresponse and measurement error, we do not analyze the nonresponse / measurement error relationship by mode in this paper. Findings from the mode experiment are reported elsewhere (Kreuter, Presser and Tourangeau, 2006).

Administrative records for five questions on the survey instrument were obtained from the Office of the Registrar. These questions included the respondent's grade point average (GPA), year of graduation, whether he/she had received certain letter grades (D,F, or W), and whether he/she graduated with honors. In addition, the Alumni Office provided administrative records to validate a question about whether the respondent is a dues-paying member of the Alumni Association and a question about whether the respondent ever donated to the University. While this alumni population limits external validity, the strength of the study arises because the validation data permits joint examination of two error sources.

3. Results

3.1 Overall Univariate Analyses

Table 1 reports the response rates among subgroups defined by the sampled alumni's true scores on the validated items. Separate rates of response are shown for each stage of recruitment – overall survey completion, contact, cooperation with the survey request, and item-level response using two measures, break-offs and item nonresponse. We anticipated little difference in contact rates across the subgroups, but greater difference in refusal rates.

Five out of the seven response rate comparisons are significantly different. Alumni who had higher grades, graduated with honors, graduated before 2000, are an alumni association member, and donated to the University are all significantly more likely to participate with the study request. Despite the fact that the number of noncontacts in the study (n=1,398) is more than double the number of refusers (n=595), most of the response rate differences arise because of differences in cooperation rates rather than contact rates. Contact rates are significantly different only for "A" students versus "C" students and for those who had versus had not donated to the University.

The largest differences in refusal rates are between alumni association member and non-members and those who had donated to the University and those who had not. The survey introduction indicated that this was a survey of University of Maryland alumni. This finding, therefore, is consistent with the idea that features of the survey design made salient in the survey request will disproportionately draw in those who are positively inclined toward that feature (Groves, Singer, and Corning

[1] The sample was stratified by year of graduation.

Table 1. Response Rates for Each Stage of Recruitment

	Overall Response Rate		Contact Rate		Cooperation Rate among the contacted	
	%	SE	%	SE	%	SE
Academic performance measures						
With D or F	31.5	0.9	70.4	0.8	44.7	1.1
Not with D or F	32.8	1.1	70.5	1.1	46.5	1.4
With W	31.8	0.8	70.0	0.8	45.4	1.0
Not with W	32.4	1.2	71.5	1.2	45.3	1.6
"A" students	35.3*	1.6	72.8*	1.5	48.5*	1.9
“B” students	31.7	0.8	70.4	0.8	45.1	1.1
“C” students	28.8	1.7	67.6	1.7	42.5	2.2
Honors student	36.6*	2.2	71.2	2.1	51.4*	2.8
Not an honors student	31.5	0.7	70.3	0.7	44.7	0.9
Alumni Measures						
Graduated 2000, 2001, 2002	29.1*	1.3	71.5	1.3	40.6**	1.7
Graduated before 2000	32.9	0.8	70.1	0.8	47.0	1.0
UMD Alumni Member	49.4****	2.4	72.8	2.1	67.9****	2.6
Not a UMD Alumni Member	30.2	0.7	70.2	0.7	43.0	0.9
Donated to UMD	41.2****	1.3	73.4**	1.2	56.1****	1.6
Did not donate to UMD	28.1	0.8	69.2	0.8	40.6	1.0

Note: +p<.10, *p<.05,**p<.01,***p<.001,****p<.0001

2000; Groves, Presser and Dipko 2004). Honors students and “A” students are also significantly more likely to cooperate with the survey request, but these contrasts are not as large as the differences created by the alumni association and donation subgroups.

3.2 Translating Response Rate Differences into Nonresponse Bias

Next we translate the differential response rates are now translated into nonresponse bias on the survey estimates. Table 2 reports the means on the validated items for various subgroups. The means are based on the record data, except for the far right column which presents the means based on the survey responses.

Nonresponse bias clearly exists for these statistics, the magnitude of which also clearly varies across statistics. Differences between the true scores for the total sample (column I) and the survey respondents (column II) illustrate the overall magnitude of the nonresponse bias. As expected from Table 1, the questions about alumni association membership and donating to the University have the largest nonresponse bias. While 37% of the respondents had donated to the University, just 29% of the total sample did this. Similarly, 14% of the respondents were alumni association members, which compares to a 9% membership rate among the total sample. Survey respondents also were more likely to have earned higher grades than nonrespondents.

The two nonresponse mechanisms can be compared using column V for the noncontacts and column VII for the refusers. The means for the noncontacts and the refusers are similar on the grade-related questions. According to the administrative records, 18% of the noncontacts were “A” students compared with 19% of the refusers. The alumni participation measures differ between the noncontacts and refusals. By contrast, those who refused were less likely to be donors or alumni association members than the noncontacts, translating into lower means for the refusers than the noncontacts.

3.3 The Effects of Differential Measurement Error

The characteristics of those who break-off the survey and those who fail to answer a single item are quite different. As a result, the impact on statistics from break-offs (not shown) and item nonresponse (VIII) differ considerably. The break-off rate is quite high, given the transition to the two self-administered modes, and has a marked effect on the survey statistics. The total number of item nonrespondents is small, so the mean for the item respondents is not tremendously different from that of the cooperators.

Table 2. Record Means on Validated Items for Nonresponse Analysis Groups

	Total Sample I.	All R's[1] II.	All NR's III.	Con-tacts IV.	Non-contacts V.	Coop-erators VI.	Refus-als VII.	Item NR's VIII.	Item R's IX.	Survey Responses X.
	%	%	%	%	%	%	%	%	%	%
Received D or F	62.1	61.0	62.6	62.1	62.2	61.1	62.8	94.4	60.7	45.5
	(0.7)	*(1.0)*	*(0.9)*	*(0.8)*	*(1.3)*	*(1.3)*	*(1.1)*	*(5.6)*	*(1.5)*	*(1.5)*
Received W	69.6	69.1	69.8	69.2	70.7	69.2	69.1	80	68.1	47.9
	(0.7)	*(1.0)*	*(0.8)*	*(0.8)*	*(1.2)*	*(1.2)*	*(1.1)*	*(6.9)*	*(1.4)*	*(1.5)*
G.P.A.	3.0	3.1	3.0	3.0	3.0	3.1	3.0	3,0	3.1	3.2
	(0.0)	*(0.0)*	*(0.0)*	*(0.0)*	*(0.0)*	*(0.0)*	*(0.0)*	*(0.04)*	*(0.0)*	*(0.0)*
A Grades	19.3	21.0	18.4	20.0	17.7	21.3	18.8	16.8	22.3	28.0
	(0.6)	*(1.0)*	*(0.7)*	*(0.7)*	*(1.0)*	*(1.1)*	*(0.9)*	*(3.4)*	*(1.3)*	*(1.4)*
B Grades	65.4	65.1	65.6	65.4	65.5	64.9	65.7	68.9	65.6	67.8
	(0.7)	*(1.0)*	*(0.8)*	*(0.8)*	*(1.3)*	*(1.2)*	*(1.1)*	*(4.3)*	*(1.5)*	*(1.5)*
C Grades	15.3	13.9	16	14.7	16.7	13.8	15.5	14.3	12.1	4.2
	(0.5)	*(1.0)*	*(0.7)*	*(0.6)*	*(1.0)*	*(0.9)*	*(0.9)*	*(3.2)*	*(1.1)*	*(0.7)*
Graduated w/ honors	9.8	11.2	9.2	10.0	9.6	11.3	8.9	0	12.3	17.2
	(0.4)	*(1.0)*	*(0.5)*	*(0.5)*	*(0.8)*	*(0.8)*	*(0.7)*	*(0.0)*	*(1.0)*	*(1.2)*
Alumni member	9.3	14.2	6.9	9.7	8.6	14.4	5.7	15.4	15.9	23.2
	(0.4)	*(1.0)*	*(0.5)*	*(0.5)*	*(0.8)*	*(0.9)*	*(0.5)*	*(7.2)*	*(1.1)*	*(1.3)*
Ever donated to UMD	29.4	37.4	25.4	30.7	26.5	37.9	24.6	42.9	40.6	41.4
	(0.6)	*(1.0)*	*(0.8)*	*(0.8)*	*(1.2)*	*(1.3)*	*(1.0)*	*(8.5)*	*(1.5)*	*(1.5)*
Recent graduate	24.8	22.6	25.9	25.2	23.9	22.6	27.4	32.1	22.6	22.5
	(0.6)	*(1.0)*	*(0.8)*	*(0.8)*	*(1.1)*	*(1.1)*	*(1.1)*	*(9.0)*	*(1.3)*	*(1.3)*
Number of cases	4,726	1,559	3,167	3,328	1,398	1,003	595	min= 18	min= 1,054	min= 1,054

[1] "R's" is used here as an abbreviation for respondents. Similarly, "NR's" denoted nonresponders.

The most striking contrast in Table 2 is the difference between the true scores for the respondents (II) and their responses in the survey (X). In terms of magnitude, the measurement error was somewhat more egregious on the grade questions than the items concerning alumni-related activities. For instance, although there are more students who ever received a "D" than did not receive a "D" according to the academic records, the survey reports suggest the reverse.

3.4 Survey Effort and Error Reduction

Next we assess the effect of additional survey effort on nonresponse and measurement error through examination of call history records. Table 3 presents the means for the responding sample based on 1-2, 3-4, 5-9, and 10 or more call attempts. The means are cumulative and computed from the records rather than the survey responses because we seek here to isolate nonresponse error. The far right column reports the true scores for the 4,726 eligible alumni in the initial sample. On the academic variables, we see that additional calling reduces nonresponse bias. As more calls are attempted, the mean for the responding

Table 3. Mean of True Scores for Responding Sample after Additional Calling

	Responding sample mean based on…				**(Target)**
	2 call attempts	*4 call attempts*	*9 call attempts*	*10+ call attempts*	True score for entire sample
	%	%	%	%	%
Received D or F	59.3	60.5	60.7	61.1	62.1
	(2.1)	*(1.6)*	*(1.4)*	*(1.3)*	*(0.7)*
Received W	65.0	66.8	68.5	69.2	69.6
	(2.0)	*(1.6)*	*(1.3)*	*(1.2)*	*(0.7)*
Graduated w/ honors	12.8	12.3	11.8	11.3	9.8
	(1.4)	*(1.1)*	*(0.9)*	*(0.8)*	*(0.4)*
G.P.A.	3.10	3.08	3.06	3.06	3.03
	(0.02)	*(0.02)*	*(0.01)*	*(0.01)*	*(0.01)*
A Grades	23.8	22.7	21.7	21.1	19.1
	(1.8)	*(1.4)*	*(1.1)*	*(1.1)*	*(0.6)*
B Grades	64.3	64.3	64.7	64.9	65.2
	(2.0)	*(1.6)*	*(1.3)*	*(1.2)*	*(0.7)*
C Grades	11.9	13.0	13.6	14.0	15.7
	(1.4)	*(1.1)*	*(1.0)*	*(0.9)*	*(0.5)*
UMD Alumni member	15.1	15.0	14.5	14.4	9.3
	(1.5)	*(1.2)*	*(1.0)*	*(0.9)*	*(0.4)*
Ever donated to UMD	38.6	37.6	37.9	37.9	29.4
	(2.1)	*(1.6)*	*(1.4)*	*(1.2)*	*(0.6)*
Recent graduate	20.2	21.7	22.0	22.6	24.8
	(1.7)	*(1.4)*	*(1.2)*	*(1.1)*	*(0.7)*
Number of cases	555	880	1293	1510	4,726

cases converges toward the mean for the initial sample. For example, 65 percent of the alumni responding on the first or second call attempt had received a "W," but after nine calls 69 percent of the responding sample had received a "W," which is within sampling error of the true value for the entire sample (70%).

On the alumni activity questions, however, the respondents brought in through additional calling do little to bridge the nonresponse gap. Across all levels of calling effort, the proportion of respondents who ever donated to the University is between 38 and 39 percent, but only 29 percent of the initial sample had ever donated A similar disparity is observed for alumni association membership.

The addition of the converted refusers did little to make the respondents more representative of the entire sample. Refusal conversion was successful in 12% of the cases in which it was attempted, which translated into 73 completed interviews. Alumni association members formed 14.5 percent of the respondents who never refused. Adding the converted refusers lowered the figure to 14.4 percent, still far from the 9.3 membership rate for the entire sample.

We conducted similar analyses to examine the effect of effort on measurement error. Surprisingly, there was no evidence of a relationship between the number of call attempts needed to obtain the interview and the level of measurement error across each of the validated items. As an example we consider the question of ever having received a "W." The difference between the mean of the survey responses and the true mean for those answering was roughly 20 percentage points both for respondents requiring just one or two call attempts and those requiring 10 or more attempts.

Similarly, no relationship exists between refusing the survey request and the propensity to misreport. Table 4 shows the measurement error on each validated item for those who never refused and those who refused once or more but eventually cooperated. On most of the validated items, the measurement error among the converted refusers appears to be less than or comparable to the measurement error among those who never refused. The small number of converted refusers limits the power of this analysis. For instance, the question about ever having donated to the University appears to be an exception, but since only 53 of the 73 converted refusers gave a substantive response, the 8 percentage point deviation from zero measurement error is attributable to a single case.

Table 4. Measurement Error by Number of Refusal Conversions[1]

	(Survey Mean - True Mean) for…	
	Respondents who never refused	*Respondents who refused at least once*
Received D or F	-16	-14
Received W	-20	-14
Graduated w/ honors	+5	+7
A Grades	+3	-1
B Grades	+7	+8
C Grades	-5	-4
UMD alumni member	+7	+6
Ever donated to UMD	+1	-8
Recent graduate	+1	-2
Minimum N	872	47

[1]Note small sample sizes for respondents who refused at least once.

These results indicate that the recruitment of respondents through high levels of effort did not adversely change aggregate levels of measurement error on the validated items in the alumni survey. We also examined respondent-level measurement error indicators, paralleling analyses reported above. This micro-level analysis mirrored those at the statistic level – we found no relationship between level of effort and rates of misreporting.

3.5 Multivariate Examination of the Relationship Between Nonresponse Propensity and Measurement Error

To understand the combined effects of academic performance, involvement in the University as an alumnus, and outcomes from the survey call attempts, we estimate three logistic regression models. These logistic regression models will be referred to as "response propensity" models (Little 1986; Groves and Couper 1998) and estimate the likelihood that the respondent will participate with the survey request, be contacted, and cooperate after contact. We create five groups – response propensity classes – from the predicted probabilities in each model. In a well-specified model, respondents and nonrespondents will be equivalent on the characteristics of interest within each class, and likelihood of survey participation will vary across the classes.

Each multivariate model includes the variables examined above and additional control variables. A log-transformed count of the total number of calls made to the case and a binary indicator for whether the case ever refused are added to the models. Additionally, gender, race, and state of current residence are included in the models.

The multivariate models replicate the bivariate analyses for the alumni measures – alumni association membership and donating to the University are significantly positive predictors of survey participation overall, and cooperation in particular. Only donating to the University is predictive of contact after controlling for the rest of the variables in the model. The academic performance measures are not significant predictors of overall response or cooperation after controlling for the alumni measures, graduation year, call outcomes, and demographics, but "A" students are more likely to be contacted than "C" students.

The total number of calls and the refusal indicator are highly significant predictors of ultimate outcome. The more calls to a case, the less likely the alumnus was to be contacted or to participate. This is not surprising. Additionally, Maryland residence is a significant negative predictor of response, contact, and cooperation. Gender and race have no relationship with the likelihood to be contacted or cooperate with the survey request.

From these models, five response propensity classes were created, ordered from lowest estimated propensity to highest estimated propensity.

The question remains as to whether those who have a lower response propensity provide answers that are more filled with measurement error than those with high response propensities. The answer, for this alumni population and these variables under the given recruitment protocol, seems to be no. Almost all associations between any of the outcomes, whether break-off, item nonresponse, or the survey report matching the record and the estimated response propensity, whether overall or by subgroup, are not significant. While some of the associations are significant, they are not consistent (e.g., all for one variable, one type of measurement error, or one type of nonresponse). We ran models that included the main effect for each variable corresponding to the measurement error and an interaction effect between the response propensity classes and the variable of interest. Few models had any significant associations; of 216 models that were run, only 15 had either a significant ($p <$

Table 5. Response Propensity Models for Overall Response, Contactability, and Cooperation

	Predicting response=1		Predicting contact=1		Predicting cooperation=1	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Intercept	0.74****	0.16	1.98****	0.16	1.15****	0.19
Academic Measures						
Received D or F	0.08	0.09	0.13	0.08	0.05	0.10
Received W	0.05	0.08	-0.04	0.08	0.08	0.09
A Grades	0.18	0.16	0.33*	0.16	0.13	0.19
B Grades	0.09	0.11	0.12	0.1	0.08	0.12
Graduated w/ honors	0.12	0.15	-0.15	0.15	0.16	0.18
Alumni Measures						
Recent graduate	-0.17*	0.08	0.08	0.08	-0.26**	0.09
UMD Alumni Member	0.47***	0.13	-0.11	0.13	0.72****	0.16
Ever donated to UMD	0.50****	0.09	0.24**	0.08	0.48****	0.10
Additional Variables						
Female=1	-0.07	0.07	-0.09	0.07	-0.05	0.08
Nonwhite=1	-0.08	0.08	-0.05	0.08	-0.1	0.09
State=MD	-0.28***	0.08	-0.17*	0.08	-0.19*	0.09
State=VA, DC	0.19	0.14	-0.15	0.14	0.43*	0.17
Log (# call attempts)	-0.91****	0.04	-0.65****	0.04	-0.78****	0.05
Ever Refused	-1.53****	0.13	n/a	n/a	-2.12****	0.14

+p<.1, *p<.05, **p<.01, ***p<.001, ****p<.0001

.10) main effect for the propensity classes or the interaction effect including the propensity classes. This is fewer than would be expected by chance.

This is not to say that there are no subgroup differences in measurement error – as shown above, measurement error varies by whether the respondent has or does not have the characteristic. However, measurement error, independent of the indicator, does not appear to be related to response propensity. Figure 1 illustrates this lack of relationship for the various validation measures by overall response propensity stratum.

4. Conclusions

Using validation data and call records, we find evidence of both nonresponse bias and measurement error in the survey of UMD alumni. The relationship between these two error sources – either in terms of a statistic being highly sensitive to both nonresponse bias and measurement error bias or nonresponse propensity related to measurement error – however, is not uniform across statistics and in some cases appears to be nonexistent. Bivariate analysis shows that alumni involvement is the most powerful predictor of response and that social desirability pressure to report good academic marks likely is the primary mechanism of measurement error for the variables examined here.

Support is lower for nonresponse / measurement error linkages than we expected based on presumed common correlates. We find evidence that relatively low academic achievers are less amenable to the survey request, but the differences may not be practically significant. As expected, lower academic achievers are more likely to misreport values for the academic items. Alumni engagement is associated with higher cooperation propensity and reporting errors on the alumni activity items were in the expected direction, but the magnitude of this measurement errors is lower than for the academic performance measures.

Finally, the relationship between level of effort and nonresponse bias and measurement error bias was not consistent across error sources. Additional call attempts improved sample representativeness, but did not completely remove nonresponse bias. Contrary to the common belief, no support was found for the hypothesis that the more difficult respondents to recruit or more reluctant respondents gave less useful data than respondents who were easier to recruit or more amenable.

5. Discussion

This paper has clearly shown that nonresponse bias exists and that measurement error exists, but that the nexus of the two is not as clear cut. When will nonresponse propensity be related to measurement error? One possible scenario is that low nonresponse propensity is also a cause of measurement error. This was investigated here in great detail, and little support was found for the hypothesis.

However, model misspecification may have led to the null result. Level of effort clearly was not related to measurement error or to nonresponse bias. Figure 2 plots the relative nonresponse bias due to noncooperation by the relative measurement error bias for each of the

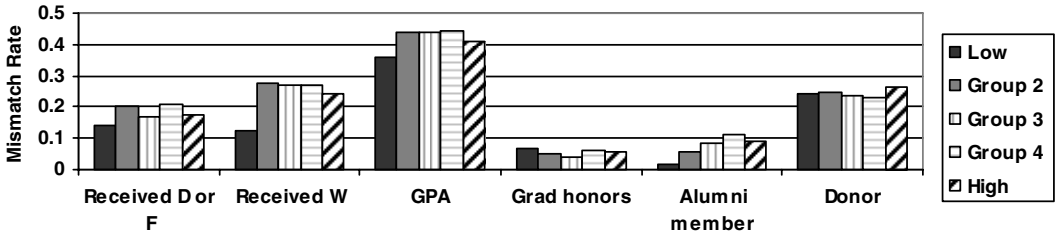

Figure 1. Discrepancy Rate between Survey Report and Records across Response Propensity Strata

validation items. This figure shows that some of the statistics that are the most sensitive to nonresponse bias are also the most sensitive to measurement error, but that the relationship is variable-specific and not clearly linear. The level of effort data might obscure the relationship between nonresponse propensity and measurement error. Future analyses will explore this in greater detail.

Acknowledgements

The authors would like to thank Principal Investigators Roger Tourangeau, Stanley Presser, Frauke Kreuter, and class instructor Katherine Abraham for use of the Alumni Survey data for these analyses. The authors would also like to acknowledge the cooperation from the University of Maryland registrar and alumni association for providing the records.

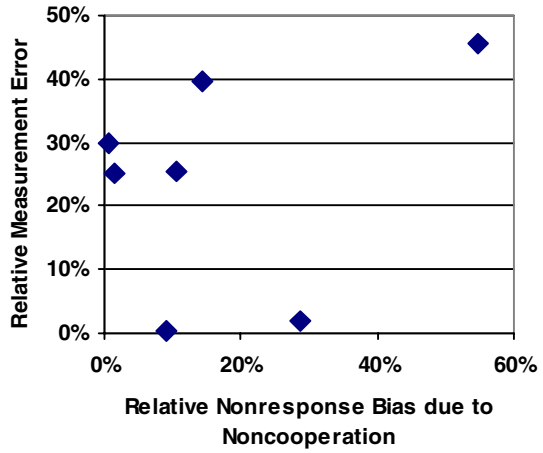

Figure 2. Scatterplot of Relative Noncooperation Bias and Relative Measurement Error

References

Armenakis, A. A. and Lett, W. L. (1982), "Sponsorship and Follow-up Effects on Response Quality of Mail Surveys." *Journal of Business Research*, 10:251-262.

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

Green, K. E. (1991), "Reluctant Respondents: Differences Between Early, Late and Nonresponders To a Mail Survey." *Journal of Experimental Education*, 59:268-276.

Groves, R. M. and Couper, M. (1998), *Nonresponse in Household Interview Surveys*, John Wiley & Sons, New York.

Groves, R. M., Presser, S., and Dipko, S. (2004), "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly*, 68:2-31.

Jones, W. H. and Lang, J. R. (1982), "Reliability and Validity Effects under Mail Survey Conditions." *Journal of Business Research*, 10:339-353.

Kreuter, F., Presser, S. and Tourangeau, R. (2006), "Comparing Social Desirability Bias in CATI, IVR and Web Surveys." Paper presented at the American Association for Public Opinion Research, Montreal.

Lin, I., and Schaeffer, N. (1995), "Using Survey Participants to Estimate the Impact of Nonparticipation." *Public Opinion Quarterly*, 59: 236-258.

Little, R. (1986), "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review*, 54: 139-157.

Martin, C. L. (1994). "The Impact of Topic Interest on Mail Survey Response Behaviour." *Journal of the Market Research Society*, 36:327-338.

Rodgers, W. L. and Herzog, A. R. (1987), "Covariances of Measurement Errors in Survey Responses." *Journal of Official Statistics*, 3:403-418.

Stang, A. and Jöckel, K. (2005), "Letter to the Editor: The Authors Reply." *American Journal of Epidemiology*, 161:403.

Yan, T., Tourangeau, R., and Arens, Z. (2004), "When Less is More: Are Reluctant Respondents Poor Reporters?" in Proceedings of the Annual Meetings of the American Statistical Association. Toronto.

Ygge, B. and Arnetz, J. E. (2004), "A Study of Nonresponse in a Questionnaire Survey of Parents' Views of Paediatric Care." *Journal of Nursing Management,* 12:5-12.