

Nonresponse in Telephone Surveys: The Reporting of Outcome Measures

Richard Seltzer

Department of Political Science, Howard University*

There is no consensus about whether or not nonresponse in telephone surveys is a cause for worry. As nonresponse rates appear to be growing there is concern that non-respondents might systematically differ from respondents thus limiting our ability to infer sample characteristics to the population.

I argue that the issue is sufficiently important that data on non-response needs to be systematically collected and made publicly available. By making results available it will be possible to conduct analysis along the lines of Crespi (1988) who examined factors (type of survey, number of days in the field, etc.) that affected the accuracy of 430 pre-election surveys. If we systematically publish data on non-response we could answer whether or not non-response affects the accuracy of pre-election surveys as well as other types of surveys.

Methodology

I sent letters to ten national survey firms who conducted telephone surveys in the 1st four months of 2003 asking them to fill out a form based upon the major AAPOR categories. I also accepted reports from their CATI system or other process by which they measured survey response. I obtained data from 9 (90%) of the survey firms for 16 of their surveys. Hart-Teeter for NBC said they did not use a CATI system and relevant data were not available.

I replicated the data collection in 2004. I sent out surveys to twelve survey firms whose national pre-election survey results received publicity. I received responses from 9 of these firms. Two of the firms that did not respond were small and Opinion Dynamics Corporation said they could not send me the data because they did not have an agreement with Fox News on how to release the data.

Results

There is considerable variation depending upon which response rate, contact rate, or refusal rate is reported. For example, the average for Response Rate 1 was about half that of Response Rate 6 (16.3% v. 31.6%). Similarly, the average for Refusal Rate 1 was about half that of Refusal

Rate 3 (20.5% v. 40.3%). There are two issues which affect the basic rates. RR1, RR3 and RR5 treats partial interviews as non-respondent while in RR2, RR4, and RR6 these are treated as responded. However, the major source of the discrepancy is related to how unknown outcomes codes (always busy, no answer, call-blocking, answering machines) are treated. In RR1 and RR2 they are treated as eligible while in RR5 and RR6 they are treated as non-eligible. In RR3 and RR4 an estimate (E) is made to apportion the unknown numbers into either eligible or non-eligible.

Clearly, there is some opportunity for manipulating outcome measures. I would get very suspicious when RR1/RR2 (perhaps trying to show a high response rate) or RR5/RR6 is reported (perhaps trying to show a low response rate). Although RR3/RR4 would be more realistic – concern is warranted because (1) There is no consensus on how to determine E and (2) the high number of outcomes that are unknown is itself a basis for concern. The average unknown eligibility rate was 48.0%. I suggest that unknown eligibility rates should also be published.

The average response rate for the 26 studies was 16.1% using RR1 and 31.6% for RR6. Even the most optimistic measure shows, that on average, less than one-third of a sample responded to a survey. This is low compared to a response rate of 75.4% (RR5) for the 2002 NORC-GSS and 66.5% (RR1) for the 2002 NES. However, remember that the NORC-GSS is a face-to-face survey that is in the field for almost 4 months and uses advance letters. The 2002 NES was a telephone survey and was in the field for seven weeks and used advanced mailings with a promise of \$20 upon completion.

Discussion

My initial intent when I began this research was to determine the extent to which non-response was a problem in publicly reported surveys. Clearly, one could not answer this question without a commonly agreed upon proper measure of non-response. I naively believed that because AAPOR had established guidelines – such a measurement existed. Once I realized I was wrong, I extended the focus of my research

to also study how operationalizing non-response measures affect reported non-response.

AAPOR's publication and dissemination of *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys* (2004) is a welcome contribution to the development of higher standards for the survey industry. However, four problems need to be addressed.

I was unable to identify any response rate that was reported for the 26 surveys discussed in this article when the survey results were first released. Many argue that such reporting would likely confuse readers. However, as Martin argued in her 2004 Presidential Address at AAPOR – there is a need for transparency in our methods or otherwise it begins to appear that we are trying to hide something. If readers can be trusted to understand “margin of error” they can also be trusted to evaluate non-response rates. AAPOR-CASRO-NCPP should require that outcome measures be reported in publicly released polls and/or made readily available on each organization's website. In addition, AAPOR-CASRO-NCPP should help to collate outcome measures for all surveys reported by its members. In essence, I suggest a compromise: requiring that when surveys results are published the reader is also given a website or other contact point where additional information about the survey could be obtained such as the full questionnaire and frequency distribution, contact information about the sponsor and survey firm, documents to help the reader interpret the results and evaluate the methodology, and detailed information about sampling and outcome measures (PEW now release these data on its website). This would go a long way toward the transparency that Martin (2004) called for and still not inundate the casual reader in his or her initial reading of the results.

The ability to “fudge” the results by utilizing different outcome measures is cause for concern. AAPOR-CASRO-NCPP should develop a consensus on which measure(s) are preferable and mandate the use of that measure(s) unless the researcher can articulate why another measure is preferable. Furthermore, all measures should be reported on the organization's website similar to what is displayed in Table 1. I admit that I cannot give a recommendation on which measure is preferable. RR3 was the most frequently reported measure in POQ articles I examined and in my mind makes the most intuitive sense.

However, this requires an estimate for E – which is beyond the ability of most pollsters to estimate. AAPOR-CASRO-NCPP should publish suggested values of E which could be used when pollsters are not confident developing an estimate on their own.

There needs to be greater standardization on how to code outcome measures. For example, at what point does one determine that a callback is actually a refusal? McCarthy (2003) points to a related issue: how to code an outcome may be affected by whether one uses the most recent or final disposition of a call. Survey firms and developers of CATI software should be encouraged to further standardize the outcome measures.

Surveys that have a large number of calls that are no answer, busy, answering machine, caller ID, etc is also a cause for concern. Therefore, I suggest that AAPOR-CASRO-NCPP should mandate that unknown eligibility rates be published.

I began with a question about whether non-response affects the quality of data? I cannot answer this question. However, if survey firms routinely published on their web site information on non-response rates as well as other variables tested by Crespi (1998) then his research could also be extended to examine non-response as a possible source of error.

References

- American Association for Public Opinion Research. 2004. *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys*, Second Edition. Ann Arbor, Michigan: AAPOR.
- Crespi, Irving. 1988. *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation
- Martin, Elizabeth, 2004. “Unfinished Business,” *Presidential Address* at the American Association for Public Opinion Research, Phoenix.
- McCarthy, Christopher. 2003. “Differences in Response Rates Using Most Recent Versus Final Dispositions in Telephone Surveys.” *Public Opinion Quarterly*, 67:396-406.

*This is note is derived from a far longer paper. Readers who wish to obtain the full version can contact the author at rseltzer@howard.edu.

OUTCOME MEASURES

Firm	StartDates	#Days	RR1	RR2	RR3	RR4	RR	RR6	CP1	CP	CP	CP	RF1	RF2	RF3	CN1	CN2	CN3	E	Unk
Start of Iraqi War																				
ABC/W Post	3/20/03	1	14.1	14.3	20.8	21.2	40.2	40.9	49.1	49.9	52.8	53.7	12.3	18.3	35.3	28.6	42.4	81.9	49.9	65.0
ABC/W Post	2/6/03	4	23.9	24.4	32.0	32.6	49.7	50.6	56.0	57.1	61.3	62.5	14.6	19.6	30.4	42.7	57.1	88.6	51.3	51.8
AP	4/2/03	4	6.5	6.7	10.4	10.8	20.9	21.6	21.0	21.7	22.4	23.1	22.3	35.8	71.7	30.9	49.6	99.3	45.4	68.9
CBS/NYT	3/4/03	2	12.8	12.8	16.8	16.8	27.8	27.8	36.1	36.1	39.0	39.0	20.0	26.4	43.5	35.5	46.7	77.0	55.4	53.9
CBS/NYT	3/20/03	5	19.4	19.4	23.3	23.3	34.7	34.7	41.2	41.2	44.7	44.7	24.0	28.9	43.0	47.1	56.6	84.3	62.1	44.1
Gallup	3/20/03	1	16.1	16.1	22.6	22.6	41.3	41.3	56.3	56.3	61.0	61.0	10.3	14.5	26.5	28.5	40.2	73.5	52.5	61.2
Gallup	1/3/03	3	16.5	16.5	19.1	19.1	24.8	24.8	32.8	32.8	39.9	39.9	24.8	28.7	37.2	50.4	58.2	75.6	59.5	33.3
LAT	4/2/03	2	8.5	8.5	12.5	12.5	18.1	18.1	45.1	45.1	48.6	48.6	8.9	13.2	19.1	18.8	27.8	40.1	38.9	53.2
LAT	1/30/03	4	10.1	10.1	15.3	15.3	21.0	21.0	37.8	37.8	40.8	40.8	14.6	22.2	30.5	26.7	40.4	55.5	34.5	52.0
Newsweek/PSRA	2/6/03	2	23.2	23.4	27.5	27.7	35.8	35.9	40.8	41.0	41.8	42.0	32.3	38.2	49.6	57.0	67.5	87.6	55.4	35.0
PEW	3/20/03	5	28.2	30.5	33.4	36.2	42.0	45.6	46.4	50.3	48.8	52.9	27.2	32.3	40.6	60.7	72.0	90.5	52.3	33.0
PEW	3/13/03	4	26.2	27.8	28.2	29.9	31.8	33.7	35.9	38.0	37.8	40.1	41.5	44.7	50.3	73.1	78.6	88.5	59.6	17.4
Time/Harris	3/27/03	1	4.9	4.9	6.8	6.8	11.8	11.8	17.8	17.8	18.7	18.7	21.2	29.4	51.2	27.4	37.9	66.1	52.4	58.6
Time/Harris	2/19/03	2	4.6	4.6	5.9	5.9	10.3	10.3	17.6	17.6	18.6	18.6	20.0	25.7	45.0	25.8	33.2	58.2	60.1	55.6
Wirthlin	4/4/03	4	7.5	9.8	10.4	13.6	17.5	22.9	20.7	27.1	21.8	28.5	24.6	34.0	57.4	36.3	50.0	84.5	51.9	57.1
Wirthlin	4/25/03	4	7.6	9.9	10.2	13.3	16.1	20.9	18.4	24.0	19.5	25.4	29.1	39.0	61.4	41.4	55.5	87.3	51.6	52.6
	Sub Average	3.0	14.4	15.0	18.5	19.2	27.7	28.9	35.8	37.1	38.6	40.0	21.7	28.2	43.3	39.4	50.9	77.4	52.1	49.5
Pre-election - 2004																				
ABC/W Post	10/1/04	33	20.1	20.9	28.6	29.7	47.0	48.8	50.5	52.4	53.4	55.4	16.8	23.9	39.2	39.9	56.7	93.1	48.2	57.1
CBS/NYT	10/28/04	5	10.4	10.4	13.8	13.8	29.2	29.2	35.3	35.3	37.4	37.4	17.3	23.1	48.8	29.4	39.1	82.7	61.3	64.5
Gallup	10/29/04	3	25.3	25.3	30.9	30.9	39.2	39.2	61.1	61.1	69.7	69.7	11.0	13.4	17.0	41.5	50.6	64.2	48.8	35.4
Harris	10/29/04	4	8.4	8.4	9.6	9.6	14.4	14.4	25.6	25.6	27.2	27.2	22.5	25.7	38.5	32.9	37.5	56.2	70.4	41.5
LAT	10/21/04	4	14.6	14.6	21.7	21.7	28.8	28.8	49.2	49.2	52.6	52.6	13.2	19.6	26.0	29.6	44.2	58.6	33.4	49.5
Marist	11/1/04	1	36.4	36.4	39.3	39.3	79.5	79.5	87.0	87.0	88.1	88.1	4.9	5.3	10.7	41.8	45.1	91.4	86.5	54.2
Newsweek/PSRA	10/28/04	2	17.6	17.6	18.9	18.9	23.9	23.9	38.6	38.6	43.6	43.6	22.8	24.5	31.0	45.5	48.9	61.9	73.7	26.5
PEW/PDS	10/27/04	4	24.5	25.6	29.5	30.9	38.0	39.8	44.3	46.4	47.9	50.2	25.4	30.7	39.5	55.2	66.6	85.8	52.0	35.7
PEW/SRBI	10/27/04	4	19.1	19.9	23.6	24.5	28.4	29.5	41.1	42.7	44.4	46.2	23.2	28.6	34.4	46.5	57.4	69.1	41.9	32.7
Zogby	11/1/04	2	10.8	10.8	14.8	14.8	26.3	26.3	26.3	26.3	27.4	27.4	28.7	39.3	69.7	41.1	56.4	100.0	54.0	58.9
	Sub Average	6.2	18.7	19.0	23.1	23.4	35.5	35.9	45.9	46.5	49.2	49.8	18.6	23.4	35.5	40.3	50.3	76.3	57.0	45.6
	Average	4.2	16.1	16.5	20.2	20.8	30.7	31.6	39.7	40.7	42.7	43.7	20.5	26.3	40.3	39.8	50.6	77.0	54.0	48.0

#Days - Number of days survey in the field

RR - Response Rate
 CP - Cooperation Rate
 RF - Refusal Rate

CN - Contact Rate
 E - Estimated proportion of unknown eligibility that are eligible based upon eligibility status that is known
 Unk - Unknown Eligibility Rate