THE IMPACT OF CALLBACKS ON SURVEY ESTIMATES IN AN ANNUAL RDD SURVEY

Daniel M. Merkle, D.S. Howard and Associates
Sandra L. Bauman and Paul J. Lavrakas, Northwestern University Survey Laboratory
Daniel M. Merkle, 307 N. Michigan Avenue, Chicago, IL 60601

KEY WORDS: Callbacks, Noncontacts, Nonresponse

Nonresponse rates in sample surveys have been rising since the 1950s (Steeh, 1981). This is a major concern for survey researchers because of the bias that results if respondents are different than nonrespondents. When such differences exist, the magnitude of nonresponse bias in survey estimates increases as the level of nonresponse increases. One important type of nonresponse, that of noncontacts, can be minimized by making repeated callbacks so that individuals who are more difficult to reach are included in the final sample. This paper focuses on this issue by exploring the impact of callbacks on survey estimates in an annual RDD survey.

The practice of making callbacks in sample surveys was not very common in the early days of survey research because of the large costs associated with making callbacks in personal interviews and due to the fact that many surveys were conducted using nonprobability sampling (Crespi, 1988:43). Today, the use of telephone interviewing makes conducting callbacks more cost-effective, but there are still many instances where callbacks are not made (cf. Crespi, 1988).

The failure to conduct repeated callbacks can have an impact on the representativeness of a given survey since those who are easy to contact are likely to be different than those who require more calls. Hilgard and Payne (1944) were among the first researchers to stress the importance of making callbacks. Using data from a national in-person survey, they demonstrated that certain types of people were more difficult to find at home and that leaving these individuals out of the sample would have biased the results of their survey.

Hilgard and Payne (1944) identified some characteristics of respondents who were more difficult to reach: those living in urban areas, individuals employed outside the home, and those who lived alone. Households with young children were easier to reach. Their conclusion served as an early warning to survey researchers about the importance of making callbacks in sample surveys: "People easily found at home on the first call differ significantly from those found at home only after repeated calls. The latter occur in large enough proportions to make it important for repeated calls to be made in order to represent them in sample surveys" (p. 261).

Other studies relying on data from in-person interviews have also described the characteristics of respondents who are more difficult to contact. In general, they tend to be younger, employed, male, have higher socioeconomic status, and live in urban areas (Crossley and Fink, 1951; Kish, 1965; Dunkelberg and Day, 1973; Smith, 1983).

There are fewer published studies dealing with the impact of callbacks in telephone surveys of the general public. Traugott (1987) addresses this issue using data from telephone surveys conducted in Michigan during the 1984 presidential campaign. Traugott found that the characteristics of the sample changed as more callbacks were conducted. Those who were contacted on the first call were more likely to be older, female, white, and more likely to identify with the Democratic party than those contacted during later calls.

Shaiko et al. (1991) also used data from telephone surveys to study nonresponse in pre-election polls. Their study used polls conducted in Syracuse, New York during the 1989 mayoral campaign and looked at two demographic characteristics: age and gender. Shaiko et al. found that younger respondents and males were more difficult to contact than older respondents and women.

Given the limited amount of research on the impact of callbacks in general population telephone surveys, this paper uses data from an annual RDD survey (conducted in 1990, 1991, and 1992) to examine the impact of callbacks on survey estimates and how consistent these effects are across the three years. Using a telephone survey employing essentially the same methodology at three points in time, these data allow us to begin making some generalizations about respondent characteristics that are related to contactability (i.e., the number of calls needed to complete the interview). The surveys also provide information about how different the final survey estimates would have been if only a limited number of callbacks were made.

METHOD

Unweighted data from the 1990, 1991 and 1992 Chicago Area Survey Project (CASP) are used in this study. CASP is an annual omnibus telephone survey conducted by the Northwestern University Survey Laboratory which includes questions on race relations,

quality-of-life measures, political attitudes, contemporary controversial issues, demographics and a variety of other issues. Households in Cook, DuPage and Lake counties (IL) were sampled using random digit dialing, and within each household one adult, 18 years of age or older, was randomly selected using the last-birthday technique. Approximately 1,000 interviews were conducted for each survey.

The methodology used in each of the three years was the same with two exceptions. First, up to 17 callbacks were made in 1990, while up to 40 callbacks were made in 1991 and 1992. Due to the fewer number of callbacks made in 1990, the analyses that follow are based on data from surveys completed in 1 through 17 calls. Second, there were fewer refusal conversions in 1990 (n=81) than in 1991 (n=168) and 1992 (n=185). Data from converted refusals are excluded so as to not confound the results. On average, converted refusers required significantly more calls ($p < .0001$) to complete an interview since, by definition, they were called at least one additional time for the refusal conversion attempt. For each survey, Table 1 shows the distribution of call attempts (up to 17 calls) for the data used in this study.

TABLE 1

Number Of Calls For Completions By Year

NO. OF CALLS	1990 (n=945)	1991 (n=812)	1992 (n=841)
1	28.4 %	20.6 %	38.0 %
2	21.4	19.0	16.1
3	12.1	11.3	11.2
4	9.5	9.5	8.4
5	6.5	7.3	5.5
6	5.1	6.4	5.1
7	5.0	4.9	3.4
8	3.4	3.7	3.6
9	3.1	2.5	1.4
10	2.5	2.8	2.1
11	0.7	2.5	1.2
12	0.8	1.8	1.2
13	0.4	2.0	1.0
14	0.8	1.6	1.0
15	0.2	2.2	0.5
16	0.0	1.1	0.2
17	0.1	0.9	0.1

ANALYSES

First, in order to determine the magnitude of the impact of call attempts on survey estimates, we computed the correlation between the number of calls needed to complete the interview and responses to each survey question (cf. Rosenthal, 1991). A square root transformation was performed on the number of calls to reduce the skew (Norusis, 1988: Chapter 17). (Information on question wording and how the variables were coded can be obtained from the first author.)

TABLE 2

Relationship Between Number Of Calls And Demographic Characteristics By Year[a]

	1990 (n=945)	1991 (n=812)	1992 (n=841)
Age	-.18***	-.14***	-.12***
Race	-.12***	-.04	-.03
Education	.11**	-.03	.09*
Employmt status	.15***	.14***	.17***
Spouse employmt	.05	-.04	-.03
Household income	.01	-.01	.05
Gender	.03	.12**	.08*

[a]Table entries are correlation coefficients. Sample sizes vary because of missing data.

*p<.05; **p<.01; ***p<.001; All significance tests are two-tailed.

Table 2 presents the correlational results for the various demographic characteristics. Two of the variables, age and employment status, are reliably related to the number of calls in each of the three years: younger respondents and those employed full-time require more calls to complete an interview. In addition, the coefficients for education and gender are each significant in two of the three years indicating that males and respondents with more education are somewhat more difficult to reach. In general, the findings for age, employment status, gender and education are consistent with past research. In addition, race is reliably related to number of calls in one of the three years (1990). When the data for the three years are combined, the correlation for race reaches statistical significance ($r=-.06$, $p<.01$) indicating that it takes slightly fewer calls to interview white respondents. (In a multiple regression using the combined 1990, 1991 and 1992 data, age, employment status, gender and race were all reliable predictors of the number of calls -- data not shown.) On the other hand, household income and spouse's employment status are not related to number of calls needed to complete the interview at the bivariate or multivariate level.

TABLE 3

Relationship Between Number Of Calls And Various Survey Questions By Year[a]

QUESTION	1990 (n=945)	1991 (n=812)	1992 (n=841)
Years of residence?	-.13***	-.04	-.09*
Likelihood of moving?	.05	.09*	.03
Recycling pickups?	-.02	-.06	-.04
Regularly recycle?	.01	-.08*	-.01
Satisfaction with neighborhood?	.00	.02	.01
How safe feel in neighborhood?	.05	.06	.01
Condition of streets and sidewalks?	.01	.05	-.01
Is home burglary a problem?	-.00	-.02	-.04
Sale/use of illegal drugs a problem?	.03	.03	.02
Quality of public schools a problem?	.09*	-.01	-.01
Is robbery/assault a problem?	-.04	-.01	.01
Poor care of property a problem?	.02	-.01	-.01
Quality of police services a problem?	.01	-.02	-.00
Satisfied with standard of living?	-.05	.02	.03
Ever worry about paying mortgage/rent?	-.01	.05	-.03
Anyone you could live with in emerg.?	.02	.04	.06
Worry about making ends meet?	.01	.02	-.06
Outlook better/worse a year from now?	.13***	.09*	.05
Own VCR?	-.04	-.03	-.00
Own home computer?	-.00	.02	-.06
All HH members have health insurance?	-.01	-.00	.06
Times seen/talked to a doctor in last year?	-.03	-.06	-.01
Days spent ill or injured in past year?	.00	-.01	.01
Amt. of housing discrim. for Hispanics?	.05	-.01	-.02
Amt. of housing discrim. for Blacks?	.05	.02	-.02
Amt. of housing discrim. for Asians?	.08*	.01	.03
Woman should be able to get legal abortion	.02	-.00	-.01
Dealth penalty should be available for use	.05	-.00	.04
Number of TV sets?	.04	-.02	-.01
Any TVs connected to cable?	.05	-.03	.03
Hours of TV watched per day?	-.11***	-.10**	-.11**
Registered to vote?	-.00	.05	.01
Rent or own home?	-.06	-.05	-.05
Apartment or house?	-.05	-.02	-.06
Hours per month volunteer work?	.03	.03	.00
Number of adults in HH?	.03	-.02	.01
Number of children in HH?	.03	.01	-.01
Number of phone lines?	.03	.01	.10**

[a]Table entries are correlation coefficients. Sample sizes vary because of missing data.

*p<.05; **p<.01; ***p<.001; All significance tests are two-tailed.

Table 3 presents bivariate correlations for 38 questions concerning various behaviors, opinions and perceptions. Only three of the 38 variables are reliably related to the number of calls in at least two of the three years. The first item in Table 3, years of residence (i.e., the length of time one has lived in the same city or suburb), is negatively related to the number of calls in 1990 and 1992. In addition, respondents with a better financial outlook for the following year are somewhat more difficult to contact in two of the three years. Finally, amount of television viewing is negatively related to number of calls in each of the three years.

However, the most striking finding in Table 3 is the general absence of significant correlations for most of the variables studied. In addition, it is important to point out that even the statistically significant correlations in Tables 2 and 3 are quite small in magnitude. For example, the strongest relationship (r = -.18 for age in 1990) accounts for only 3% of the variance. Therefore, the next logical question concerns the practical implications of these correlations.

We answer this question by comparing survey estimates based on three calls to those based on all calls. This allows us to determine the extent to which conducting numerous callbacks changes survey estimates relative to the three call situation. We chose three calls instead of the more typical one call because, for many survey researchers, the important question isn't whether or not callbacks should be made but rather how many callbacks should be made.

For the demographic variables, the difference between making three calls and all calls ranges from less than one percentage point to 7.5 percentage points (see Table 4). As we would expect given the

TABLE 4

Comparison Of Select Estimates For Three Calls Vs. All Calls by Year
(Demographic Characteristics)

	1990 (n=945)			1991 (n=812)			1992 (n=841)		
	1-3 Calls	All Calls	Difference	1-3 Calls	All Calls	Difference	1-3 Calls	All Calls	Difference
AGE									
> 50 yrs.	33.0%	28.5%	**4.5%**	31.2%	25.7%	**5.5%**	32.2%	30.6%	**1.6%**
RACE									
White	79.6	75.7	**3.9**	74.3	72.3	**2.0**	74.1	73.5	**0.6**
EDUCATION									
College grad	32.0	34.9	**-2.9**	40.4	39.8	**0.6**	40.1	41.1	**-1.0**
EMPLOYMENT									
Full-time	50.7	55.3	**-4.6**	54.2	61.1	**-6.9**	54.3	58.1	**-3.8**
SPOUSE'S EMPLOYMENT									
Full-time	34.6	31.9	**2.7**	31.0	29.4	**1.6**	29.0	27.5	**1.5**
INCOME									
<$20,001	22.9	22.9	**0.0**	23.2	21.8	**1.4**	23.5	22.8	**0.7**
GENDER									
Female	60.8	59.6	**1.2**	63.2	55.7	**7.5**	60.8	58.4	**2.4**

TABLE 5

Summary Of Percentage Point Differences Between Three Calls And All Calls By Year

PERCENTAGE POINT DIFFERENCE	TOTAL (n=135)	YEAR 1990 (n=45)	1991 (n=45)	1992 (n=45)
Less than 1 percentage point	47.4 %	48.9 %	37.8 %	55.6 %
1-1.99 percentage points	31.9	24.4	35.6	35.6
2-2.99 percentage points	8.9	8.9	13.3	4.4
3-3.99 percentage points	6.7	11.1	4.4	4.4
4 percentage points or more	5.2	6.7	8.9	0.0
MEAN	1.39	1.48[a]	1.69[b]	.99[ab]
S.D.	1.38	1.36	1.71	.89

NOTE. This table summarizes the differences between three calls and all calls for the 45 survey items analyzed in this paper. The mean and standard deviations are calculated using absolute values of the differences. Means with the same superscript are significantly different at $p<.05$.

correlations in Table 2, the largest percentage point differences across the three years are for employment status and age. Taking employment status as an example, if survey results were based only on the first three calls, rather than all calls, the proportion of people employed full-time would have been underestimated by 4.6 percentage points in 1990, 6.9 percentage points in 1991, and 3.8 percentage points in 1992. A similar analysis of the 38 other survey items indicated that the difference between estimates based on three calls and all calls was quite small in most cases (data not shown).

Table 5 summarizes the differences between three calls and all calls for each question presented in Table 4 and for the 38 other survey measures. Overall, for the 135 questions across the three years, approximately half of the estimates changed by less than one percentage point when the results from calls 4-17 were added to calls 1-3. In addition, over three-fourths of the estimates changed by less than two percentage points when comparing three calls with all calls. Five percent of the estimates changed by four percentage points or more, with the largest change being 7.5 percentage points.

These results vary somewhat when broken down by year (see Table 5). Specifically, the average difference between three calls and all calls was somewhat larger in 1990 (M = 1.48) and 1991 (M = 1.69) than in 1992 (M = .99). Even after taking into account the few characteristics that differed between the three surveys (e.g., the number of refusal conversions, the proportion of interviews completed in three calls of less), the reason for these differences is not readily apparent.

CONCLUSION

This research found that some demographic variables are reliably related to the number of calls needed to complete an interview. In general, the results indicate that in telephone surveys it takes more calls to interview younger respondents, those employed full-time, men, those with more education and nonwhite respondents. These findings are fairly consistent with past research using data from in-person and telephone surveys. However, unlike previous research, household income was not reliably related to the number of calls. In a further analysis, only three of 38 other survey measures were reliably related to the number of calls in at least two of the three years.

The results of this study also indicate that in most cases survey estimates based on a limited number of calls (i.e., up to three) were not very different than those based on numerous calls. Conducting three versus numerous calls changed survey estimates by less than two percentage points for most variables. For some demographic characteristics (e.g., employment status) and other variables related to contactability (e.g.,

television viewing) these differences were somewhat larger, but not larger than 7.5 percentage points. However, it should be noted that the interpretation of these differences as large or small really depends on the purpose of the survey and how the data will ultimately be used.

It is important to keep in mind that the focus of this paper was on the impact of callbacks on survey estimates, and it therefore does not provide information about bias due to refusals or noncontacts (i.e., people who were never reached even after repeated callbacks). Although conducting numerous callbacks is an easy way to increase one's response rate in a telephone survey by decreasing the number of noncontacts, in some cases this strategy may actually increase nonresponse bias if refusals are ignored or pursued to a lesser extent (cf. Groves, 1989:194; Groves and Lyberg, 1988:209; Stinchcombe et al., 1981; Wilcox, 1977:593). This is because, in some situations, those who are easier to contact are also more likely to refuse (e.g., Merkle, Bauman and Lavrakas, 1991; Shaiko et al., 1991; Wilcox, 1977). By conducting a large number of callbacks and ignoring refusal conversions, the potential bias due to noncontacts decreases, but any bias due to refusals remains and may actually increase.

Therefore, from the standpoint of minimizing total survey error (cf. Lavrakas, 1993), the findings in this study suggest that, in some cases, it may be desirable to conduct somewhat fewer callbacks and shift these resources to another area such as converting refusals. The goal of future research should be to investigate further how the biases due to noncontacts and refusals interact while exploring optimal ways of allocating resources in order to minimize both of these potential sources of bias in the most efficient way.

REFERENCES

Crespi, I. *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation, 1988.

Crossley, H.M. and R. Fink. Response and nonresponse in a probability sample. *International Journal of Opinion and Attitude Research*, 1951, 5, 1-19.

Dunkelberg, W.C. and G.S. Day. Nonresponse bias and callbacks in sample surveys. *Journal of Marketing Research*, 1973, 10, 160-8.

Groves, R.M. *Survey Errors and Survey Costs*. New York: John Wiley and Sons, 1989.

Groves, R.M. and L.E. Lyberg. An overview of nonresponse issues in telephone surveys. In R.M. Groves, et al. (Eds.), *Telephone Survey Methodology*. New York: John Wiley and Sons, 1988.

Hilgard, E.R. and S.L. Payne. Those not a home: Riddle for pollsters. *Public Opinion Quarterly*, 1944, 8, 254-61.

Kish. L. *Survey Sampling*. New York: John Wiley, 1965.

Lavrakas, P.J. *Telephone Survey Methods: Sampling, Selection and Supervision*. 2nd Edition. Newbury Park: Sage Publications, 1993.

Merkle, D.M., S.L. Bauman and P.J. Lavrakas. Nonresponse bias: Refusal conversions and callbacks in RDD telephone surveys. Paper presented at the Midwest Association for Public Opinion Research, Chicago, 1991.

Norusis, M.J. *SPSS/PC+ V2.0 Base Manual*. Chicago: SPSS, Inc., 1988.

Rosenthal, R. *Meta-Analytic Procedures for Social Research*. Newbury Park: Sage Publications, 1991.

Shaiko, R.G., D. Dwyre, M. O'Gorman, J.M. Stonecash and J. Vike. Pre-election political polling and the non-response bias issue. *International Journal of Public Opinion Research*, 1991, 3, 86-99.

Smith, T.W. The hidden 25 percent: An analysis of nonresponse on the 1980 General Social Survey. *Public Opinion Quarterly*, 1983, 47, 386-404.

Steeh, C.G. Trends in nonresponse rates, 1952-1979. *Public Opinion Quarterly*, 1981, 45, 40-57.

Stinchcombe, A.L., C. Jones, and P.B. Sheatsley. Nonresponse bias for attitude questions. *Public Opinion Quarterly*, 1981, 45, 477-96.

Traugott, M.W. The importance of persistence in respondent selection for preelection surveys. *Public Opinion Quarterly*, 1987, 51:48-57.

Wilcox, J.B. The interaction of refusal and not-at-home sources of nonresponse bias. *Journal of Marketing Research*, 14, 592-7.