

Published in *Journal of Marketing Research*, 14, 1977, 396-402.

## **Estimating Nonresponse Bias in Mail Surveys**

**J. Scott Armstrong**

The Wharton School, University of Pennsylvania

**Terry S. Overton**

Marketing Scientist, Merck, Sharp and Dohme

### **Abstract**

Valid predictions for the direction of nonresponse bias were obtained from subjective estimates and extrapolations in an analysis of mail survey data from published studies. For estimates of the magnitude of bias, the use of extrapolations led to substantial improvements over a strategy of not using extrapolations.

---

### **Introduction**

The mail survey has been criticized for nonresponse bias. If persons who respond differ substantially from those who do not, the results do not directly allow one to say how the entire sample would have responded – certainly an important step before the sample is generalized to the population.<sup>1</sup>

The most commonly recommended protection against nonresponse bias has been the reduction of nonresponse itself. Nonresponse can be kept under 30% in most situations if appropriate procedures are followed [25, 32, 44]. Another approach to the nonresponse problem is to sample nonrespondents [20]. For example, Reid [39] chose a 9% subsample from his nonrespondents and obtained responses from 95% of them.

Still another approach to the nonresponse problem, the one examined herein, is to estimate the effects of nonresponse [7, 21]. Many researchers have concluded that it is not possible to obtain valid estimates [10, 23, 29, 36]. Fillion [14] reanalyzed data from Ellis et al. [10] and concluded that, in fact, extrapolation did help. Furthermore, Erdos and Morgan [11] favor estimation where judgment warrants.

Estimates of nonresponse bias may be used for any of the following reasons.

1. *Reanalyzing previous surveys.* If the survey was carried out some time ago, the only way to deal with nonresponse bias is to estimate its effects. With the establishment of data archives [3, 24], the reanalysis of survey data is likely to increase in popularity.
2. *Saving money.* The effort to increase the rate of return becomes more difficult as the rate of return increases. If it were possible to estimate the nonresponse bias, it might be more economical to accept a lower rate of return. In other words, the estimation strategy might provide equivalent results at a lower cost.
3. *Saving time.* If respondents are expected to change substantially in the near future (as often happens in political surveys), obtaining a high rate of return may not be feasible because it requires too much time. In such cases, it would be desirable to estimate the nonresponse bias.

---

<sup>1</sup> Total sample refers to persons who presumably were contacted. Those obviously not contacted should be excluded (they would be primarily persons whose initial questionnaire was returned as undeliverable). Results from [38] indicate that the not-contacted group is more similar to respondents than to nonrespondents.

This article examines methods for estimating nonresponse bias. Predictions of the direction of nonresponse bias are evaluated, and estimates are made of the magnitude of this bias. An attempt was made to include all relevant previously published studies.

### **Methods For Estimating Nonresponse Bias**

The literature on nonresponse bias [e.g., 27, 46] describes three methods of estimation: comparisons with known values for the population, subjective estimates, and extrapolation.

#### **Comparison with Known Values for the Population**

Results from a given survey can be compared with “known” values for the population (e.g., age, income). However, as the known values come from a different source instrument, differences may occur as a result of response bias [17, 50] rather than nonresponse bias. Furthermore, even if the tested items are free from nonresponse bias, it is often difficult to conclude that the other items are also free from bias [10, 31]. The use of known values still can be helpful. For example, the *Literary Digest* survey failure in the 1936 Roosevelt-Landon election could have been averted by such a procedure [19].

#### **Subjective Estimates**

Several researchers [e.g., 5] have suggested that subjective estimates of nonresponse bias would be useful. It is not clear how one should obtain these subjective estimates of bias, although several approaches have been proposed. One approach is to determine socioeconomic differences between respondents and nonrespondents [26, 48]. For example, respondents generally are better educated than nonrespondents [6, 44, 49], and there may be differences in personality between respondents and nonrespondents [33, 48].

The “interest hypothesis” is another widely recommended basis for subjective estimates [2, 8, 9, 17]. It involves the assumption that people who are more interested in the subject of a questionnaire respond more readily [1, 30, 40, 41, 47], and that nonresponse bias occurs on items in which the subject’s answer is related to his interest in the questionnaire [4]. Finally, Rosenthal [42] concludes from a review of the literature that people are more likely to respond to a questionnaire if they would make a favorable impression upon anyone who reads the responses.

Despite the uncertainty about the use of subjective estimates, they are used. Furthermore, they have been shown to have some validity in [43], where the direction of bias was correctly predicted for each of 17 items.

#### **Extrapolation Methods**

Extrapolation methods are based on the assumption that subjects who respond less readily are more like nonrespondents [37]. “Less readily” has been defined as answering later, or as requiring more prodding to answer.

The most common type of extrapolation is carried over *successive waves* of a questionnaire. “Wave” refers to the response generated by a stimulus, e.g., a follow-up postcard. Persons who respond in later waves are assumed to have responded because of the increased stimulus and are expected to be similar to nonrespondents.

*Time trends* provide another basis for extrapolation [12]. Persons responding later are assumed to be more similar to nonrespondents. The method of time trends has an advantage over the use of waves in that the possibility of a bias being introduced by the stimulus itself can be eliminated. On the negative side, it is difficult to measure the time from the respondent’s awareness of the questionnaire until completion.

The method of *concurrent waves* involves sending the same questionnaire simultaneously to randomly selected subsamples. Wide variations are used in the inducements to ensure a wide range in rate of return among these subsamples. This procedure allows for an extrapolation across the various subsamples to estimate the response

for a 100% rate of return. The advantage of this procedure is that the extrapolation can be done at an early cutoff date because only one wave is required from each of the samples.

### **Estimating the Direction of Nonresponse Bias**

The prediction of the direction of nonresponse bias is useful for assessing uncertainty. Consider a questionnaire with an 80% rate of return where 1% of the respondents indicated an intention to purchase a new product. The possible limits from the complete sample could range from 0.8% if all nonrespondents had no intention to purchase to 20.8% if all nonrespondents intended to buy.<sup>2</sup> However, if one were able to predict the direction of nonresponse bias, these limits might be greatly reduced. In this example, if it could be stated that the nonrespondents would report lower intentions to purchase than respondents, the range of uncertainty would be 0.8% to about 1.0%, a substantial reduction.

### **Test Procedures**

To examine whether an estimate of nonresponse bias is valid, one must find other data about the nonrespondents to use as criteria. The criterion data can differ on two dimensions-completeness and method of data collection. "Completeness" refers to the percentage of nonrespondents about whom data are available. Thus, if the initial data covered 70% of the sample, what part of the remaining 30% is covered by the criterion data? The method of data collection is important because respondents may answer differently in a mail survey than they would in a personal interview if the questions concern sensitive items [16, 23, 49]. Mail survey results were used for the criterion in this study except where noted.

In the tests that follow, judges predicted whether the criterion response (from the second wave) was above, the same as, or below the response from the first wave. A two-tailed test (.05 level) of the differences of proportions from two samples of unequal sizes was used to divide the items into three categories, depending on whether the second-wave response was significantly above that for the first wave (U), not significantly different (N), or significantly lower (D).

Data were obtained from 16 previously published studies [1, 6, 8, 10, 15, 16, 23, 30, 34, 35, 39, 41, 43, studies A and E in 44, 45]. The sample sizes in each study were generally large. The first wave ranged in size from 60 to 7,900 with a median of about 1,000; the criterion waves ranged from 45 to 5,000 with a median of about 770. The part of the sample responding in the first wave ranged from 10 to 75% with a median of 42%. Finally, the nonrespondents covered by the mail criterion ranged from 13 to 92% with a median of 44%

### **Subjective Estimates**

Descriptions of the published studies were presented to three judges, professors at the Wharton School, who had prior experience with mail surveys but were not familiar with any of the studies in this sample. Each judge was asked to identify items that would be subject to nonresponse bias, and to state the direction of bias. He was instructed to use any basis he thought relevant in making these estimates.

A scheme for assessing predictive accuracy is illustrated in Table 1. Data for this test were available from the studies mentioned in the preceding section, with the exceptions of [15, 16].

---

<sup>2</sup> The limits for complete sample response can be obtained from lower limit =  $RQ$  and upper limit =  $1 - R(1 - Q)$ , where  $R$  is the rate of return and  $Q$  is the proportion of respondents giving a specified response to an item. Tables for  $R$  and  $Q$  can be found in [12] and [18].

**Table 1**  
**Classification for Errors in Predicting Direction**

Estimated Direction of Bias	Actual Direction of Bias		
	Up (U)	None (N)	Down (D)
Up	C	S	I
None	S	C	S
Down	I	S	C

Key: C = correct, S = some error, I = incorrect.

An analysis of items on which the judges were either correct (the percentage of C's in terms of Table 1), somewhat incorrect (percentage of S's), or incorrect (percentage of I's) is summarized in Table 2. The judges did better than chance (that is, random choice of one of the three responses), yet they had no obvious superiority over the assumption that there was no bias. The latter method was correct on the 46% of the items which were not significantly biased.

**Table 2**  
**Accuracy of Experts in Predicting Direction for All Items**  
**(n = 136)**

Method	C	S	I
	Correct estimates (%)	Some error (%)	Incorrect estimates (%)
Judge 1	44	51	5
Judge 2	49	44	7
Judge 3	64	33	3
Chance	33	49	18
Assumption of no bias	46	54	0

How accurately could the experts predict for those items which were significantly biased (columns U and D of Table 1)? These results, summarized in Table 3, indicate that each of the three judges did better than chance; but in comparison with the assumption of no bias, the judges' higher percentage of correct predictions must be weighed against the higher percentage of incorrect predictions.

Disagreement among the judges was high. Interjudge reliability ranged from 56% identical predictions between judges 1 and 2 to 59% between judges 1 and 3. Although better than chance (33%), reliability was poor. Efforts to obtain more reliable estimates were expected to improve accuracy. Two approaches improving reliability were examined – the “interest hypothesis” and the “group consensus.”

*Interest Hypothesis.* It was believed that interjudge reliability might be improved by basing predictions solely on the interest hypothesis. (Follow-up interviews with the three judges had indicated that, although they used socioeconomic factors, they placed primary reliance on the interest hypothesis.) Consequently, six additional judges, also professors from the Wharton School, were selected and were given instructions to follow the interest hypothesis.

Contrary to expectations, the interest hypothesis provided little improvement in interjudge reliability. The average percentage of items classified identically was 62% in contrast to a 57% average for the experts with no formal instructions. Further, there was no gain in the accuracy of the predictions.

Although the interest hypothesis did not improve predictive ability, it did provide an inexpensive way to instruct novices how to estimate nonresponse bias. Three naive judges were selected, a housewife and two high school students. In terms of accuracy and interjudge reliability, their performance was no different from that of the other nine judges.

*Consensus.* A second attempt to improve reliability involved use of a group consensus. Group consensus is typically better than the average judge in a group and in some cases is superior to the best judge [51]. A consensus from the original three judges was selected on the following basis: an item was estimated to be biased if all three judges were in agreement as to the direction of bias for that item, or if two judges were in agreement and the third made no directional estimate; in all other cases, the item was estimated to be unbiased.

Results from the consensus estimates are presented in Table 3. Although the number of correct estimates was similar to that of the average judge, the percentage of incorrect estimates was cut by half. The reduction of incorrect estimates (from 9 to 4%) was achieved primarily at the expense of overlooking more biased items (from 25 to 32%), a tradeoff which may be favorable in some situations. Similar results were obtained when the analysis was repeated with the other three sets of judges.

**Table 3**  
**Accuracy of Experts in Predicting Direction for Biased Items Only**  
**(n = 74)**

Method	C Correct estimates (%)	S Some error (%)	I Incorrect estimates (%)
Judge 1	59	32	9
Judge 2	73	15	12
Judge 3	67	28	5
Average judge	66	25	9
Consensus of 3 judges	64	32	4
Chance	33	33	33
No bias	0	100	0

### Evidence on the Use of Extrapolations

The method of using concurrent waves for extrapolation was not tested, because no studies using this method could be found.

Limited data were available to test the time response method. Ford and Zeisel [16] did not find such extrapolations to be valid. Four additional studies provided evidence on the response time trend within the first wave. In total, however, there were only nine items – three items each from [1, 16], two from [44], and one from [45]. There were six correct and three incorrect predictions. Overall, then, there are not sufficient data to judge this method.

Extrapolations across two waves were examined by Zimmer [52]; improved predictions were found for five of seven items. The authors found eight additional studies with 63 biased items – one item each from [6, 30, 44], two items from [34], three from [1], seven from [39], and 24 each from [8] and [23]. Kivlin [28] was excluded because of an inadequate description. Data from the first two waves were used to estimate the direction of the bias and the third wave was used as the criterion. The criterion for [39] was based on a telephone follow-up and all the other studies used mail. The extrapolation method was superior to chance in all respects (row 1 of Table 4).

**Table 4**  
**Accuracy of Extrapolation in Predicting Direction for Biased Items Only**  
**(n = 763)**

Method	C Correct estimates (%)	S Some error (%)	I Incorrect estimates (%)
Extrapolation	89	0	11
Chance	33	33	33
Chance (with forced choice)	50	0	50
Subjective-extrapolation	60	38	2

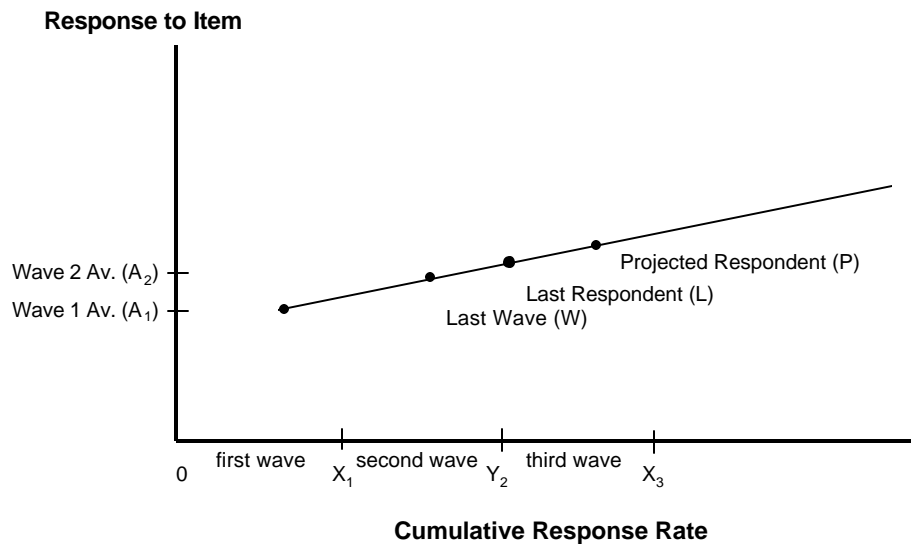
*A Combined Subjective-Extrapolation Method*

It seemed reasonable to expect a reduction in incorrect estimates if predictions were made only for items on which the subjective and the extrapolation predictions were in agreement. When the extrapolation and the consensus predictions were so combined, the results supported that hypothesis (see the last row in Table 4). For the 39 items on which a directional prediction was made by the combined method, there was only one incorrect estimate. A comparison of the first and last rows of Table 4 shows a reduction in incorrect estimates from 11% to less than 2%; this decrease was obtained at the expense of an increase in the percentage of items overlooked from 0 to 38%. Thus, the subjective-extrapolation method helps to reduce major errors. In this sense, it is a more conservative method.

**Estimating the Magnitude of Nonresponse Bias**

There are many ways one might extrapolate. Regression across waves was suggested by Filion [13, 14] for cases involving at least three response waves. In this study, in which extrapolation was based on only two waves with the third wave as a criterion, three simple methods were examined. The simplest, the “last wave” method, assumes the nonrespondents are like the average respondent in the second wave. The other two methods project the trend in responses across the first two waves: the “last respondent” method assumes the nonrespondents are like the projected last respondent in the second wave, and the “projected respondent” method assumes that the respondents are like the projected respondent at the midpoint of the nonresponse group. The methods are illustrated in the figure.

**Figure**  
**Response Trend Projections**



The prediction for this third wave was simple for the “last wave” method (termed W); the nonrespondents were assumed to respond as did those in the second wave. For the “last respondent,” a linear extrapolation was made by plotting the averages for the first and second waves, and drawing a line to the point representing the cumulative percentage of respondents at the end of the second wave. This can be calculated from:<sup>3</sup>

$$A_2 + (A_2 - A_1) \left( \frac{X_2 - X_1}{X_2} \right) = L$$

where A is the percentage response to an item within a given wave, X is the cumulative percentage of respondents at the end of a given wave, L is the theoretical last respondent, and the subscripts represent the wave.

For the “projected respondent” method, the extrapolation was carried out in the same way as for the “last respondent” method except that the line was extended to the midpoint of the criterion wave. With the same notation as before, where P is the projected respondent:

$$A_2 + (A_2 - A_1) \left( \frac{X_3 - X_1}{X_2} \right) = P$$

To keep the magnitude of the nonresponse problem in perspective, the criterion was based on the cumulative percentage response for each item at the end of the third wave. The predicted cumulative response, C<sub>3</sub>, can be calculated from:

$$\frac{C_2 X_2 + (X_3 - X_2) \hat{A}_3}{X_3} = C_3$$

where  $\hat{A}_3$  is the predicted response for the nonrespondents, W, L, or P. (Where extrapolation is used for the complete sample, X<sub>3</sub> would be 100%.)

Though the basic question was *whether* to extrapolate, the authors also asked *when* to extrapolate, hypothesizing that extrapolation is useful only if there are a priori grounds for expecting bias; in other cases there should be no extrapolation.<sup>4</sup> This approach is referred to as “selective extrapolation.”

The authors further hypothesized that the “last respondent” method would provide the most effective way to extrapolate. This method incorporates information about the trend in responses from early to later waves, yet it stays within the range of the historical data.

Data for testing the extrapolation methods were drawn from the 11 studies [1, 6, 8, 10, 15, 23, 30, 34, 39, studies A and E in 44] which had three waves and sufficient documentation to allow for subjective predictions of direction. There were a priori grounds for extrapolation on 53 of the 112 items from these studies.

The results showed that the error from extrapolation for these items was substantially less than that from no extrapolation (column 1 of Table 5). Improvement occurred on 43 of the 53 items, which is statistically significant ( $p < .001$ ) by using the sign test.

---

<sup>3</sup> The “last respondent in last wave” and “projected respondent” methods also carry the constraint that the prediction must exceed zero. This constraint was not needed for the studies reported here.

<sup>4</sup> The decision of when to extrapolate made use of the combined consensus extrapolation criterion discussed previously. If there was a consensus that bias should occur (that is, at least two judges agree and the third makes no prediction), and if the actual trend from wave 1 to wave 2 agreed with this consensus, the extrapolation was used.

**Table 5**  
**Accuracy of Extrapolation**

Method	MAPE*		
	A priori items (n = 53)	Non-a priori items (n = 59)	All items (n = 112)
	(1)	(2)	(3)
No extrapolation	4.8	2.7	3.7
Last wave	3.3	2.3	2.8
Last respondent	2.7	2.5	2.6
Projected respondent	2.5	2.7	2.6
Selective extrapolation	n.a.	n.a.	2.7

\* MAPE is the mean absolute percentage error, i.e., the average absolute error in predicting cumulative response, divided by the actual cumulative response.

The projected respondent method yielded the most accurate predictions for a priori items. The improvement over the last respondent method was of little practical significance (2.5 versus 2.7%), but it was statistically significant as improvements were found on 39 of 53 items ( $p < .001$ ).

Because there were relationships among the items within a study, the analysis of MAPE was repeated with each study used as the unit instead of each item. Here, the last respondent method yielded the lowest error (2.2), followed by the last wave (2.5), the projected respondent (2.7), and no extrapolation (3.5). The same type of results were obtained by use of mean absolute error instead of mean absolute percentage.

Results for items with no a priori expectation of bias are presented in column 2 of Table 5. The last wave extrapolation led to a reduction in the cumulative response error compared with no extrapolation, but the reduction was not statistically significant as it occurred on only 30 of 57 items (there were two ties). These data also were analyzed by treating each study as a unit; the rankings of the methods were the same. The same conclusions were obtained when mean absolute error was used.

“Selective extrapolation,” in which the last respondent method was used for a priori items and no extrapolation was used for other items, was compared with the “no extrapolation” and “total extrapolation” methods. The results (column 3 of Table 5) indicated that any of the extrapolation methods were more accurate than no extrapolation. But the different methods of extrapolation did not lead to important differences among themselves. Apparently, one need only find a “reasonable” approach to extrapolation, a selection that can be made on the basis of cost and simplicity.

The study made predictions for third wave responses. In practice, one would be predicting to 100% of the sample. To examine whether these conclusions are valid as one approaches 100%, results from the six studies with the highest cumulative rate of return after the third wave (average return = 92%) were compared with those from the five studies with the lowest cumulative rate of return after the third wave (average return = 60%). The fact that no differences were found between these groups in the percentage of items on which the prediction of magnitude was improved suggests that the results can be generalized to 100%.

An additional study [38] was obtained after the analyses had been completed. Because data were available for the complete sample, this study was useful in examining an extrapolation to 100%. The criterion was based on an earlier mail survey, which protected against differences due to response bias. Four questions were available, and each met the criteria for the interest hypothesis. The MAPE for no extrapolation was 10.3, which was inferior to the MAPE's of 1.7 for the last wave, 1.6 for the last respondent, and 6.6 for the projected respondent.

### Summary

Judges made valid predictions for the direction of nonresponse bias for items in mail surveys. These estimates were most accurate for items which were significantly biased (66% correct, 9% incorrect). Use of a



consensus led to further improvements (64% correct, 4% incorrect). The direction of bias also was predicted by extrapolation; when extrapolations from the first two waves were used to predict bias in the third wave, they were correct for 89% and incorrect for only 11% of the significantly biased items. A combined subjective-extrapolation method was correct on 60% of these items and incorrect on only 2%. These results show that it is possible to obtain valid predictions of the direction of bias. Such predictions are useful for reducing the confidence intervals for mail survey results.

Predictions of the magnitude of bias also were examined, by using results from the first two waves of a survey to predict the third wave. Extrapolation led to a reduction of error by almost half of that found with no extrapolation. The results were not very sensitive to the use of different methods of extrapolation.

The authors recommend that the theoretical last respondent be used as a prediction for the nonrespondent in cases where there are a priori grounds; in other cases, there should be no extrapolation. But a simple extrapolation across all items also produced favorable results. These results in favor of extrapolation contrast sharply with the conclusions found in the literature.

### References

1. Baur, E. Jackson, "Response bias in a mail survey," *Public Opinion Quarterly*, 11 (Winter 1947-48), 594-600.
2. Benson, Lawrence E., "Mail surveys can be valuable," *Public Opinion Quarterly*, 10 (1946), 234-241.
3. Bisco, Ralph L., "Social science data archives: progress and prospects," *Social Sciences Information*, 6 (February 1967), 39-74.
4. Blair, William S., "How subject matter can bias a mail survey," *Media/Scope*, (February 1964), 70-72.
5. Brown, Rex V., "Just how credible are your market estimates?" *Journal of Marketing*, 33 (July 1969), 46-50.
6. Clausen, John and Robert N. Ford, "Controlling bias in mail questionnaires," *Journal of the American Statistical Association*, 42 (1947), 497-511.
7. Daniel, Wayne W., "Nonresponse in sociological surveys: A review of some methods for handling the problem," *Sociological Methods and Research*, 3 (February 1975), 291-307.
8. Donald, Marjorie N., "Implications of nonresponse for the interpretation of mail questionnaire data," *Public Opinion Quarterly*, 24 (1960), 99-114.
9. Edgerton, Harold A., Stuart H. Britt and Ralph D. Norman, "Objective differences among various types of respondents to a mailed questionnaire," *American Sociological Review*, 12 (Summer 1947), 435-444.
10. Ellis, Robert A., Calvin M. Endo, and J. Michael Armer, "The use of potential nonrespondents for studying nonresponse bias," *Pacific Sociological Review*, 13 (Spring 1970), 103-109.
11. Erdos, Paul L. and Arthur J. Morgan, *Professional Mail Surveys*. New York: McGraw-Hill Book Co., 1970.
12. Ferber, Robert, "The problem of bias in mail returns – A solution," *Public Opinion Quarterly*, 12 (Winter 1948-1949), 669-676.
13. Filion, F. L., "Estimating bias due to nonresponse in mail surveys," *Public Opinion Quarterly*, 40 (Winter 1975-76), 482-492.
14. Filion, F. L., "Exploring and correcting for nonresponse bias using follow-ups of nonrespondents," *Pacific Sociological Review*, 19 (July 1976), 401-408.

15. Finkner, A. L., "Methods of sampling for estimating peach production in North Carolina," *North Carolina Agricultural Experiment Station Technical Bulletin*, 91, 1950. As cited in William G. Cochran, *Sampling Techniques*. New York: John Wiley and Sons. Inc., 1963. p. 374.
16. Ford, Robert N. and Hans Zeisel, "Bias in mail surveys cannot be controlled by one mailing," *Public Opinion Quarterly*, Fall 1949, 495-501.
17. Franzen, Raymond and Paul Lazarsfeld, "Mail questionnaire as a research problem," *Journal of Psychology*, 20 (1945), 293-320.
18. Fuller, Carol H., "Weighting to adjust for survey nonresponse," *Public Opinion Quarterly*, 38 (1974), 239-246.
19. Gallup, George H., *The Sophisticated Poll Watcher's Guide*. Princeton, NJ: Princeton Opinion Press, 1972.
20. Hansen, Morris H. and W. N. Hurwitz, "The problem of non-response in sample surveys," *Journal of the American Statistical Association*, 41 (December 1946), 517-529.
21. Hendricks, W. A., "Adjustment for bias by non-response in mail surveys," *Agricultural Economics Research*, 1 (1949), 52 ff.
22. Hochstim, Joseph R., "A critical comparison of three strategies of collecting data from households," *Journal of the American Statistical Association*, 62 (1967), 976-989.
23. Hochstim, Joseph R. and Demetrios A. Athanasopoulos, "Personal follow-up in a mail survey: Its contribution and its cost," *Public Opinion Quarterly*, 34 (Spring 1970), 69-81.
24. Hyman, Herbert, *Secondary Analysis of Sample Surveys*. New York: John Wiley and Sons, Inc., 1972.
25. Kanuk, Leslie and Conrad Berenson, "Mail surveys and response rates: A literature review," *Journal of Marketing Research*, 12 (November 1975), 440-453.
26. Kirchner, Wayne K. and Nancy B. Mousley, "A note on job performance: differences between respondent and nonrespondent salesmen to an attitude survey," *Journal of Applied Psychology*, 47 (Summer 1963), 223-244.
27. Kish, Leslie, *Survey Sampling*. New York: John Wiley and Sons, Inc., 1965.
28. Kivlin, Joseph E., "Contributions to the study of mail-back bias," *Rural Sociology*, 30 (Fall 1965), 332-336.
29. Lansing, John B. and James N. Morgan, *Economic Survey Methods*. Ann Arbor: Survey Research Center, University of Michigan. 1971.
30. Larson, R. F. and W. R. Carton, Jr., "Can the mail-back bias contribute to a study's validity?" *American Sociological Review*, 24 (April 1959), 243-245.
31. Lehman, Edward C., Jr., "Tests of significance and partial returns to mail questionnaires," *Rural Sociology*, 28 (September 1963). 284-289.
32. Linsky, Arnold S., "Stimulating responses to mailed questionnaires: A review," *Public Opinion Quarterly*, 30 (Spring 1975), 82-101.
33. Lubin, B., E. E. Levitt, and M. S. Zukerman, "Some personality differences between responders and non-responders to a survey questionnaire," *Journal of Consulting Psychology*, 26 (1962), 192.
34. Mayer, Charles S. and Robert W. Pratt, Jr., "A note on nonresponse in a mail survey," *Public Opinion Quarterly*, 30 (Winter 1966), 637-646.

35. Newman, Sheldon W., "Differences between early and late respondents to a mailed survey," *Journal of Advertising Research*, 2 (June 1962), 37-39.
36. Ognibene, P., "Correcting non-response bias in mail questionnaires," *Journal of Marketing Research*, 8 (May 1971), 233-235.
37. Pace, C. Robert, "Factors Influencing questionnaire returns from former university students," *Journal of Applied Psychology*, 23 (June 1939), 388-397.
38. Pavalko, Ronald M. and Kenneth G. Lutterman, "Characteristics of willing and reluctant respondents," *Pacific Sociological Review*, 16 (1973), 463-476.
39. Reid, Seerley, "Respondents and non-respondents to mail questionnaires," *Educational Research Bulletin*, 21 (April 1942), 87-96.
40. Reuss, C. F., "Differences between persons responding and not responding to a mailed questionnaire," *American Sociological Review*, 8 (August 1943), 433-438.
41. Rollins, Malcom, "The practical use of repeated questionnaire waves," *Journal of Applied Psychology*, 24 (December 1940), 770-772.
42. Rosenthal, Robert, "The volunteer subject," *Human Relations*, 18 (1965), 389-406.
43. Schwirian, Kent P. and Harry R. Blaine, "Questionnaire-return bias in the study of blue-collar workers," *Public Opinion Quarterly*, 30 (Winter 1966), 656-663.
44. Scott, Christopher, "Research on mail surveys," *Journal of the Royal Statistical Society, Series A*, 124 (1961), 143-191.
45. Stanton, Frank, "Notes on the validity of mail questionnaire returns," *Journal of Applied Psychology*, 23, (February 1939), 95-104.
46. Stephan, Frederick F. and Philip J. McCarthy, *Sampling Opinions*. New York: John Wiley and Sons, Inc., 1958.
47. Suchman, Edward A. and Boyd McCandless, "Who answers questionnaires?" *Journal of Applied Psychology*, 24 (December 1940), 758-769.
48. Vincent, Clark E., "Socioeconomic status and familial variables in mail questionnaire responses," *American Journal of Sociology*, 69 (May 1964), 647-653.
49. Wallace, David, "A case for-and-against mail questionnaires," *Public Opinion Quarterly*, 18 (1954), 40-52.
50. Wiseman, Frederick, "Methodological bias in public opinion surveys," *Public Opinion Quarterly*, 36 (Spring 1972), 105-108.
51. Zajone, Robert B., *Social Psychology: An Experimental Approach*. Belmont, CA: Brooks/Cole Publishing, 1967.
52. Zimmer, H., "Validity of extrapolating non-response bias from mail questionnaire follow-ups," *Journal of Applied Psychology*, 40 (1956), 117-121.