*An imputation procedure is used to estimate the effects of nonresponse on issues of substantive interest in a social survey. Using this method, one can determine that nonresponse bias may have differential effects on variable means, depending on the combination of independent variables used in the ensuing substantive analysis of data from the survey.*

# ESTIMATION OF NONRESPONSE BIAS

DARNELL F. HAWKINS
*University of Michigan*

Within recent years many researchers have noted a significant rise in the nonresponse rate for various studies utilizing the survey research method.[1] In the past, such nonresponse has not been a topic of extensive sociological inquiry partly because there has been a tendency to view nonresponse as a methodological issue of primary concern to those whose principal interest is in the theory and practice of sampling. A declining response rate as well as an increased interest in all forms of measurement error has created, however, a new concern over problems inherent in survey nonresponse.

Nonresponse in survey research represents a failure to obtain measurements from some of the units being sampled (Kish, 1965: 532); however, the bulk of the nonresponse lies with observations which refuse to be interviewed or those observa-

AUTHOR'S NOTE: *I am grateful to Michael Flynn, James Jackson and David Segal, principal investigators of the 1973 Detroit Area Study, for the use of their data. E. M. Beck read earlier drafts and provided invaluable assistance. My thanks also to Mary Senter, John Siebs, Connie Gask, Paula Sa, and Diane Dunham for their suggestions and clerical assistance.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 3 No. 4, May 1975

tions which, even after repeated attempts, cannot be contacted. During the actual process of interviewing, various techniques are available to alter the nonresponse rate. Increasing call-backs and subsampling are the two most common actions, but after such supplementary precautions have been taken, the final nonresponse poses quite a serious problem—a problem which goes beyond the mere mechanics of response accumulation by adversely affecting the validity of the interpretations one makes using the data that have been collected.[2] Moreover, it is seldom the case in survey research that even sampling error is systematically computed and analyzed for important substantive values in the survey. Nonsampling error, such as nonresponse, receives even less serious scrutiny, and often appears as a footnote only to be forgotten in the ensuing substantive analysis.

Considering these practices, one must ask how the use of potentially biased data affects the kind of substantive analysis engaged in by sociologists. In other words, having analyzed fallible survey data and hence having made substantive inferences based on that data, we must ascertain to what degree the problem of nonresponse bias affects those substantive conclusions. The interest of this article is not on the formal properties (unbiasness, efficiency, consistency, sufficiency) of statistics derived from the data, but on the sociological inferences made from these summarizing statistics. To this end, this paper will use data from the 1973 Detroit Area Study (DAS) conducted by the University of Michigan to investigate the impact of nonresponse on the robustness of substantive conclusions based on fallible survey data.

The analysis consists of two major parts; one part is designed to measure the impact of supplementary precautions, such as call-backs, on the extent and nature of nonresponse bias. This procedure involves the analysis of changes in the demographic characteristics of respondents during interviewing. An estimate of substantive bias is made through the comparison of regression coefficients for three selected variables at various points in the interviewing process. The second part of the analysis uses an imputation procedure whereby some terminal nonrespondents (refusals) are given imputed values for the three selected variables; and a comparison is made between these imputed values for refusals and the actual values for respondents. The difference between these values is then used as an estimate of the bias due to nonresponse.

## NONRESPONSE: PAST STUDIES

Previous studies of nonresponse have been of two principal types. One group of studies has focused on attempts to describe more accurately the nature of the bias caused by nonresponse. Since it is unlikely that two surveys will have identical patterns of bias, these studies are case specific. It is hoped that through a series of such studies future researchers will be able to identify certain consistent "causes" and effects of nonresponse. Some analysts have sought to identify possible demographic correlates, such as sex, race, or age, to explain the varying patterns of response in surveys. Other researchers have looked at interviewer effects on nonresponse, the nature of the survey instrument itself, accessibility of respondents, and a wide range of other possible correlates.

The second major group of studies has taken a mainly statistical and mathematical approach to the problem of nonresponse bias. These studies are generally not case specific; but rather they are attempts to offer more generalized, statistical solutions to the problem. Attempts are made to offer ways of correcting for the bias of nonresponse by quantifying it, and then making needed adjustments in the survey statistics in order to cope with its effects. These solutions are offered as being applicable to diverse survey situations, and not merely to one case. Among the many investigators who have contributed to this tradition are Birnbaum and Sirken (1950), Hansen and Hurwitz (1946) and Hendricks (1949).

From the resulting assorted studies of nonresponse only a few significant demographic or other correlates of nonresponse have been identified, and these are often significant for explaining nonresponse in some surveys, but not in others. Few of the studies, however, are concerned with terminal nonresponse which is the real cause of bias. In the case of interview surveys, most studies of nonresponse involve a series of "wave" analyses of the characteristics of respondents and nonrespondents at various points in the interviewing process. All of the interim nonrespondents, except for strong refusals and those on the last call-back, eventually become respondents at the end of the survey. These interim nonrespondents are, for the most part, not-at-homes or mild refusals who are later found at home or persuaded to be interviewed. These studies of nonrespondents could more accurately be labeled studies of "reluctant respondents" as indeed some are (see Pomeroy, 1963; Robins, 1963).

Dunkelberg and Day (1973) have provided one of the most valuable studies of nonresponse in the wave analysis vein. In a study of the 1967 Survey of Consumer Finances conducted by the Survey Research Center at the University of Michigan, they have attempted to quantify the bias due to nonresponse. In their own words, they sought "to describe empirically the relationship between bias in the distribution of selected respondent characteristics in a personal interview survey and the number of calls used in the interviewing process" (1973 160).

The authors found that most of the categories of respondent characteristics converged on their population values after two or three call-backs. A few categories, however, were slow to converge and were characterized by rather large initial distributional errors. For each (demographic) variable of importance in the study, it was possible to calculate the number of calls required to reach a desired level of accuracy. As in most past studies of this kind, however, the level of accuracy used as a standard by the researchers is the final response group; hence terminal nonresponse is not analyzed. Another shortcoming of the wave analysis approach is its emphasis on distributional bias (i.e., bias in the distribution of demographic traits of respondents and nonrespondents) rather than on the impact of nonresponse on issues of substantive interest.

In order to deal with the problem of terminal nonresponse, many attempts have been made to estimate the bias of nonresponse, and then to incorporate that estimate into the survey statistics. However, the estimation of nonresponse bias is no simple matter. Unless expensive remeasurements or subsampling is used researchers generally have little knowledge of the effects and degree of such bias, especially in surveys of more restricted and rare populations. Other independent sources of information about nonresponse are seldom available. As a result, statistical remedies for nonresponse are often based on only partial information. Deming (1950), Cochran (1963), and Kish (1965) describe and analyze many of these various techniques and remedies. However, there are few tests of the practical value of such compensatory estimation procedures in the literature.

Fewer studies can be found in the literature which are designed to estimate the impact of terminal nonresponse bias on the substantive issues which almost every survey of sociological importance contains. For example, what is the effect of distributional bias in income or age variables on substantive issues in which these variables are used? How does nonresponse bias affect the results of such statistical procedures as correlation, regression, chi-square, or other tests of significance used in sociological research? These questions are especially crucial since most surveys in sociological research are used as instruments to test hypotheses and theories which require elaborate statistical manipulations and not merely the enumeration of gross demographic traits.

Three very important studies of nonresponse bias pose some of these questions and attempt answers. Suchman (1962) analyzed the effects of nonresponse bias on hypothesis testing. He concluded that where variables are independently related to the bias, then the use of the biased or unbiased data will show the same interrelationship between these variables. Thus, he concludes further that while distributional bias may be important, this bias does not always lead to significant biasing of substantive issues such as the direction of relationships between variables. He provides a well-documented and cogent argument. However, he is quick to point out that in using these knowingly biased samples, one cannot generalize to population cross sections. The emphasis is on the phenomenon being studied and not on its distribution in the general population (Suchman, 1962: 110). Yet such generalization to the general population is precisely the aim of most social scientists using the survey method.

It is also true that simple hypothesis testing in modern sociological methodology often takes a back seat to elaborate estimation procedures, such as multivariate analysis and various forms of causal modeling, which are more sensitive to nonresponse bias. The effects of nonresponse on these estimation procedures would be an important topic of investigation which would serve as an adjunct to the Suchman study.

In another study of substantive nonresponse bias, Lagay (1969-1970) uses two methods to estimate the effects of bias on survey values. The first method, a "check data" method, used independent sources of demographic information to check on the accuracy of survey demographic data. Check data on several variables of importance showed no significant differences between respondents and nonrespondents on these variables. The second method, the "dependent variable" method, was designed to measure the impact of bias on a dependent variable of substantive interest in the survey (a variable using the St. Paul family functioning scale). Both chi-square and rank-sum tests applied to the data indicated that the nonresponse and response groups differed significantly on this variable.

Schwirian and Blaine (1966-1967) explored the effects of nonresponse on mail questionnaire returns. Using the respondents to the second (final) mailing as "nonrespondents," the authors conclude: (1) that there is substantial bias due to nonresponse, (2) that the direction of bias is predictable, and (3) that of three selected pairs of variables, the association between one pair was significantly different for the two groups.

Unfortunately, the three studies discussed above stand almost alone as studies designed to measure the effects of nonresponse bias on substantive survey values and issues in a direct manner. This aspect of nonresponse deserves much more analysis and attention.

## WAVE ANALYSIS: THE PROBLEM OF INTERIM NONRESPONSE

The analysis of interim nonresponse alone fails to give a clear and accurate picture of the nature and extent of nonresponse bias in any survey, as noted above. However, when analyzed in conjunction with both interim substantive bias and terminal bias, it can provide valuable information. In such instances the analysis of interim distributional bias may help explain certain findings about the nature of interim substantive bias. This is especially important for comparison purposes with surveys that incorporate different call-back policies. Finally, the analysis of interim bias may reveal that there are trends in patterns of nonresponse during interviewing which result in the kind of final nonresponse group we observe.

Therefore, the first step in our analysis is to explore the changes in respondent characteristics through the process of interviewing conducted by the Detroit Area Study in 1973. This survey of adults was based on a multistage probability sample of housing units for the Detroit SMSA. Within households, respondent selection was based on the number of adults in the unit of age 18 or beyond. One adult was randomly selected as respondent from this listing of eligible adults. The final sample was composed of 845 households located on 109 blocks in the metropolitan area of which 576 became successful interviews, 177 were terminal refusals, 50 were not-at-homes, respondent absent or noninterview for other reasons, and 42 were non-sample. A total of 17 calls were made to some households in order to obtain the 576 completed interviews although the average (mean) number of calls was only 3.

Table 1 gives the cumulative distributions of respondent characteristics by the number of calls. For the purposes of this analysis, the relatively few cases requiring ten or more calls were combined into a single 10+ category. The changes observed in Table 1 can be taken as a rough quantification of the distributional bias of respondent characteristics at various points in the interviewing process.[3] Several trends can be noted by using the cumulative distribution after ten or more calls as the standard. For example, some distributional bias can be observed for the sex variable. The final sample contains 59% females, yet the sample after one call contains over 66% females, and after two calls, females are still slightly overrepresented. The same can be observed for various other respondent characteristics over the successive waves of calls. For example, white-collar workers make up 16% of all respondents after one call, but by the final series of calls they comprise nearly 32% of the completed sample.

One simple measure of the differences in the distribution of the selected respondent characteristics across the waves is reported in Table 1. This involves the calculation of the percentage error in the cumulative distributions between the first and last calls and between the sixth and last calls. These calculations show that some groups initially are greatly underrepresented, while others are overrepresented. Between the first and final calls, the white-collar group, the college-educated group, Jewish respondents, multiple-unit dwellers, the lowest income group, the aged, and housewives show the greatest percentage error. Low percentage error can be noted for low education groups, married persons, Protestants, and the middle-income groups. By the sixth call, many of these differences have diminished greatly and one observes an almost complete convergence on the final sample values. However, even between the sixth and final calls, blacks, the unemployed and retired, the over-65 age group, blue-collar respondents, and Jewish respondents still show an appreciable percentage of error. This suggests that more of these respondents were picked up on later calls and might have been missed in a similar survey of this population which used a more limited number of call-backs.

TABLE 1
CUMULATIVE DISTRIBUTIONS OF RESPONDENT CHARACTERISTICS BY NUMBER OF CALLS

| | Number of Calls | | | | | Percentage Error[a] | | |
|---|---|---|---|---|---|---|---|---|
| Characteristic | 1 | 2 | 3 | 6 | 10+ | Calls 1 & 10+ | Calls 6 & 10+ | N[b] |
| SEX | | | | | | | | |
| Males | 33.6 | 40.9 | 42.4 | 40.9 | 41.1 | −18 | −1[c] | 237 |
| Females | 66.4 | 59.1 | 57.6 | 59.1 | 58.8 | +13 | +1 | 339 |
| | N=113 | N=264 | N=361 | N=504 | N=576 | | | |
| RACE | | | | | | | | |
| White | 82.3 | 82.6 | 82.8 | 80.5 | 78.6 | + 5 | +2 | 451 |
| Black, other | 17.7 | 17.4 | 17.2 | 19.5 | 21.4 | −14 | −7 | 123 |
| | N=113 | N=264 | N=361 | N=502 | N=574 | | | |
| RELIGION | | | | | | | | |
| Protestant | 55.2 | 56.9 | 55.3 | 54.5 | 55.6 | − 1 | −2 | 289 |
| Catholic | 41.0 | 38.9 | 38.3 | 39.3 | 37.9 | + 8 | +4 | 197 |
| Jewish | 3.8 | 4.2 | 6.4 | 6.2 | 6.5 | −42 | −5 | 34 |
| | N=105 | N=239 | N=329 | N=455 | N=520 | | | |
| EDUCATION | | | | | | | | |
| 1-11 years | 29.2 | 30.0 | 30.1 | 29.3 | 29.2 | 0 | +1 | 167 |
| High school graduate | 33.6 | 34.2 | 30.9 | 31.5 | 32.0 | + 5 | −2 | 183 |
| Voc. Tech. & 1-3 years college | 29.2 | 23.2 | 24.5 | 25.6 | 25.2 | +16 | +2 | 144 |
| College graduate, or more | 8.0 | 12.5 | 14.5 | 13.6 | 13.5 | −41 | +1 | 77 |
| | N=113 | N=263 | N=359 | N=499 | N=571 | | | |

TABLE 1 (Continued)

| Characteristic | Number of Calls | | | | | Percentage Error[a] | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 6 | 10+ | Calls 1 & 10+ | Calls 6 & 10+ | N[b] |
| OCCUPATION | | | | | | | | |
| White Collar | 15.9 | 24.9 | 28.9 | 31.2 | 31.7 | –50 | –2 | 179 |
| Blue Collar | 29.2 | 26.8 | 25.6 | 25.8 | 27.1 | + 8 | –5 | 153 |
| Housewife & student | 44.3 | 37.5 | 35.1 | 33.1 | 31.9 | +39 | +4 | 180 |
| Unemployed & retired | 10.6 | 10.7 | 10.4 | 9.9 | 9.3 | +14 | +6 | 53 |
| | N=113 | N=261 | N=356 | N=493 | N=565 | | | |
| AGE | | | | | | | | |
| 18-24 years | 15.0 | 16.5 | 15.8 | 16.2 | 16.3 | – 8 | –1 | 92 |
| 25-44 years | 37.2 | 41.2 | 41.7 | 42.8 | 42.0 | –11 | +2 | 238 |
| 45-64 years | 35.4 | 32.7 | 32.1 | 31.3 | 32.7 | + 8 | –4 | 185 |
| over 65 years | 12.4 | 9.6 | 10.4 | 9.7 | 9.0 | +38 | +8 | 51 |
| | N=113 | N=260 | N=355 | N=495 | N=566 | | | |
| MARITAL STATUS | | | | | | | | |
| Married | 64.5 | 64.1 | 66.1 | 63.6 | 64.4 | 0 | –1 | 367 |
| Single | 12.7 | 14.5 | 13.7 | 13.4 | 13.5 | – 6 | –1 | 77 |
| Divorced, widowed or separated | 22.8 | 21.4 | 20.2 | 23.0 | 22.1 | + 3 | +4 | 126 |
| | N=110 | N=262 | N=357 | N=500 | N=570 | | | |
| FAMILY INCOME | | | | | | | | |
| $000-$5,999 | 23.7 | 21.7 | 19.3 | 18.9 | 18.6 | +27 | +2 | 96 |
| $6,000-$9,999 | 15.5 | 13.7 | 15.0 | 15.4 | 15.9 | – 2 | –3 | 82 |
| $10,000-$19,999 | 42.3 | 45.4 | 45.9 | 44.0 | 43.2 | – 2 | +2 | 223 |
| $20,000 or more | 18.6 | 19.2 | 19.9 | 21.8 | 22.3 | –17 | –2 | 115 |
| | N=97 | N=240 | N=327 | N=455 | N=516 | | | |

**TABLE 1 (Continued)**

| Characteristic | Number of Calls | | | | | Percentage Error[a] | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 6 | 10+ | Calls 1 & 10+ | Calls 6 & 10+ | N[b] |
| CITY OF RESIDENCE | | | | | | | | |
| Detroit | 44.2 | 42.8 | 41.0 | 40.1 | 41.8 | + 6 | −4 | 241 |
| Suburbs | 55.8 | 57.2 | 59.0 | 59.9 | 58.2 | − 4 | +3 | 335 |
| | N=113 | N=264 | N=361 | N=504 | N=576 | | | |
| YEARS IN DETROIT AREA | | | | | | | | |
| 1-9 years | 9.1 | 11.5 | 11.2 | 11.0 | 10.7 | −15 | +3 | 61 |
| 10-40 years | 38.2 | 37.8 | 35.6 | 35.4 | 36.0 | + 6 | −2 | 205 |
| Over 40 years, or entire life | 52.7 | 50.7 | 53.2 | 53.6 | 53.3 | − 1 | +1 | 304 |
| | N=110 | N=262 | N=357 | N=500 | N=570 | | | |
| TYPE OF HOUSING UNIT | | | | | | | | |
| Detach single family | 81.4 | 82.3 | 81.5 | 79.5 | 77.7 | + 5 | +2 | 442 |
| Two-four family or row house | 9.7 | 9.2 | 10.4 | 11.5 | 12.5 | −22 | −8 | 71 |
| Apartment, trailer, other | 8.9 | 8.5 | 8.1 | 9.0 | 9.8 | − 9 | −8 | 56 |
| | N=113 | N=260 | N=357 | N=497 | N=569 | | | |

a. The percentage error is a measure of the extent to which the sample after 1 and 6 calls differs from the final sample, i.e.,

$$\frac{\%\text{Call 1 or 6} - \%\text{Call 10+}}{\%\text{Call 10+}}$$

b. The number of cases in this column refers to the total number of respondents after 10+ calls.
c. A "1" is also used to denote percentages which are greater than zero, but less than 1.

As to be expected, a sample consisting only of persons who had responded after one call (or even two calls) would be markedly different in some respects from the sample of respondents after ten or more calls. In most cases, at least three to six calls were required before a sample distribution approximating the final sample is achieved. Even after six calls, there are slight discrepancies in distributions which might cause problems of bias in the interpretation of substantive issues which are very sensitive to distributional changes in respondent-demographic characteristics.

Estimated regression coefficients are one form of substantively oriented survey values which are useful for the analysis of interim substantive bias. By observing the changes in these coefficients with the increasing number of call-backs, one can see how distributional bias is translated into one form of substantive bias. For these purposes three opinion-attitude questions often found in sociological literature were selected. (See Appendix A for these items as they appear in the survey.) One question asked the respondents about the strength of their political party ties (Political Affiliation); one sought to have respondents designate their social class status (Subjective Class); the other was designed to tap the respondent's attitude regarding the pace of government efforts to eliminate racial discrimination in employment (Race Discrimination).

The responses were used from these three survey items as dependent variables ($Y_i$), and sex ($X_1$), race ($X_2$), age ($X_3$), and family income ($X_4$) as independent variables, and a regression analysis was performed for various interim response waves which correspond to those used in the analysis of distributional bias in Table 1. Table 2 shows the changes in the estimated regression coefficients at various points in the interviewing process. One notes many changes in these coefficients as the number of calls made to households increases. Since there is also considerable flux in the standard error of these estimators, the extent and importance of these changes are not easily discernible. What would be ideal, therefore, is a statistical procedure which would allow one to calculate the number of calls required before one obtains "stable" coefficients, which are not greatly affected by additional calls. A procedure analogous to the one used by Dunkelberg and Day (1973) to estimate the optimum number of calls needed to minimize distributional bias is needed for determining how call-backs affect regression coefficients. However, such a procedure would be forced, by necessity, to use the final response group as a standard of comparison. Thus, while it would give us a more quantitative estimate of the impact of different call-back policies on the stability of the regression coefficients one obtains after many calls, it, like wave analysis, tells us nothing about terminal nonresponse and its possible impact on these coefficients.

Finally, it is obvious that the calculation of a "true" value for an attitude variable, such as the ones described above, is an important task. Attitude variables and many other nondemographic variables have no one value which can be verified, for example, by the census or other independent sources of information. At best, past studies may be consulted for some guidance; but one is never certain after such consultation whether change in values in later surveys reflects actual change in the population or is simply the result of measurement error. It is for these reasons that the present analysis consists of only a simple enumeration and charting of the changes in coefficients, while a more rigorous attempt at the quantification of bias is made only for terminal nonresponse.

Despite the use of less rigorous statistical methods, the analysis of interim substantive bias in Table 2 provides much useful information about the impact of call-backs on survey results, the nature of interim bias, and the effects of bias on a statistical procedure such as regression. First, one notes that some of the coefficients are affected more than others by additional calls. However, in all three regressions the race coefficient shows considerable change. This is especially important since it is a significant explanatory variable for all three dependent variables. Also, one notes that in all three regressions the sex coefficient is at times negative and at other times

TABLE 2
CHANGES IN REGRESSION COEFFICIENTS BY CUMULATIVE WAVES OF RESPONSE

| Var. | Estimated Parameters | Number of Calls 1 | 2 | 3 | 6 | 10+ |
|---|---|---|---|---|---|---|
| Subjective Class | $b_0$ | 2.1642[a] (.34581) | 2.4021[a] (.19883) | 2.2218[a] (.16190) | 2.1264[a] (.14065) | 2.2101[a] (.13351) |
| | $b_1$ Sex | .00311 (.13405) | −.08188 (.07606) | −.04931 (.06441) | −.01969 (.05385) | −.02707 (.05130) |
| | $b_2$ Race | −.19269 (.17209) | −.41410[a] (.10327) | −.34102[a] (.08906) | −.31248[a] (.07170) | −.34421[a] (.06619) |
| | $b_3$ Age | .00034 (.00401) | .00098 (.00236) | .00195 (.00197) | .00167 (.00170) | .00130 (.00162) |
| | $b_4$ Income | .03651[a] (.01774) | .02263[a] (.01077) | .03150[a] (.00903) | .03899[a] (.00771) | .03343[a] (.00731) |
| | $R^2$ | .08(N=94) | .12(N=230)[a] | .12(N=317)[a] | .14(N=441)[a] | .13(N=500) |
| Race Discrimination | $b_0$ | 3.4593[a] (.52006) | 3.6593[a] (.33763) | 3.4598[a] (.26877) | 3.4014[a] (.24836) | 3.4690[a] (.23057) |
| | $b_1$ Sex | −.04326 (.19769) | −.01832 (.12740) | .05482 (.10574) | .08128 (.09421) | .06323 (.08767) |
| | $b_2$ Race | .60863[a] (.25505) | .69255[a] (.17328) | .76092[a] (.14641) | .72640[a] (.12584) | .77371[a] (.11325) |
| | $b_3$ Age | −.00312 (.00598) | −.00956[a] (.00400) | −.00698[a] (.00329) | −.00890[a] (.00301) | −.01107[a] (.00282) |
| | $b_4$ Income | −.03634 (.02712) | −.01680 (.01832) | −.01387 (.01504) | −.00590 (.01370) | −.00213 (.01268) |
| | $R^2$ | .10(N=95)[a] | .10(N=232)[a] | .11(N=317)[a] | .11(N=441)[a] | .13(N=499)[a] |
| Political Affiliation | $b_0$ | 5.5114[a] (1.0858) | 5.9079[a] (.67945) | 5.9281[a] (.56157) | 5.9336[a] (.49307) | 5.6850[a] (.45549) |
| | $b_1$ Sex | −.46034 (.41047) | .30929 (.25572) | .34728 (.22117) | .15784 (.18749) | .16176 (.17390) |
| | $b_2$ Race | 1.8068[a] (.54403) | 1.0923[a] (.34969) | 1.0331[a] (.30730) | .97448[a] (.25007) | 1.0745[a] (.22483) |
| | $b_3$ Age | −.00440 (.01258) | −.01091 (.00802) | −.01318 (.00682) | −.00937 (.00594) | −00743 (.00553) |
| | $b_4$ Income | −.03475 (.05571) | −.07737[a] (.03653) | −.08243[a] (.03111) | −.08873[a] (.02690) | −.07238[a] (.02479) |
| | $R^2$ | .14(N=94)[a] | .09(N=229)[a] | .10(N=315)[a] | .09(N=439)[a] | .09(N=499)[a] |

a. Coefficient or regression significant at .05 level.

positive at various stages of interviewing. This is to be expected, however, considering its level of significance.

It is important to note, however, that the stability of regression coefficients is affected also by nature of the experimental design chosen by the researcher. For example, Swindel (1974) and others have observed that if the independent variables in a regression are highly correlated, then the least-squares coefficients will be unstable; and one may even observe changes in the direction of the relationships for coefficients, such as that observed for the sex variable. Therefore, the correlation between the independent variables used in this analysis would have to be examined before definitive conclusions can be reached. A check of these correlations revealed very low intercorrelations among the independent variables. The highest correlation observed was a .36 correlation between age and family income after one call.

From Table 2, one also notes that by the sixth call most of the coefficients have somewhat stabilized and are not greatly affected by the subsequent round of calls. This state of relative stability after six calls was also noted for the distribution of respondent-demographic characteristics in Table 1. Similarly, one observes that the standard errors have also stabilized and consequently so have significance levels. Yet, one observes that between the sixth and final call the race coefficient in the Race Discrimination analysis changes from .72 to .77 without a considerable change in the standard error of the estimate. This change is considerably greater than that observed for any of the other coefficients, suggesting perhaps that this variable association is particularly sensitive to the characteristics of the last waves of reluctant respondents.

In the absence of knowledge about terminal nonrespondents, the conclusions that can be drawn from the data in Table 2 are limited. The real value of this kind of analysis lies in the effort to translate the rough demographic estimates of nonresponse bias shown in Table 1 into a more readily interpretable, substantive form. However, in general one may conclude from these data that: (1) there are substantial changes in some estimated regression parameters as the number of calls increases, and (2) the parameters for certain variables are affected more than those for other variables, i.e., nonresponse has a differential impact on survey values. A more conclusive and more complete picture of nonresponse can be reached by considering terminal as well as interim bias.

Almost every survey is also faced with the problem of terminal nonresponse (principally refusals and not-at-homes). The bias which results from this hard core of nonresponse cannot be calculated in a wave analysis, as can be done for interim nonresponse. Almost nothing is usually known about terminal nonrespondents in area probability samples, not even their exact demographic characteristics. Trends in the patterns of nonresponse may give some indication of the characteristics of missing respondents. However, there is no evidence of consistent, overall trends in the distributions of demographic-respondent traits in Table 1 which could be useful in drawing definite conclusions about the characteristics of the 177 refusals found in the 1973 Detroit Area Study. The final part of this analysis investigates the effects of the distributional bias of terminal nonresponse on the three substantive variables used in the analysis of interim bias.

## ESTIMATION OF TERMINAL NONRESPONSE BIAS

The 1973 Detroit Area Study interview cover sheets included space for information about persons who refused to be interviewed. Along with descriptions of the dwelling unit within which the refusal occurred, the age and sex of the refusing person were also obtained. The respondent selection procedure used for this survey necessitated a listing of household adult members prior to designating a respondent. This information was often obtained before the designated respondent had an opportunity to refuse. Most of the age data were obtained from these listings. Some of the age data were estimated by the interviewer when it was possible to see and talk to the respondent, but when he/she did not wish to divulge his/her age. The sex data were obtained from direct interviewer observations or from information given by another person in the household.

Of 177 terminal refusals, both age and sex data are available for 140. With this data—the racial composition of the blocks on which the refusal occurred (from 1970 census data) and income data from the survey itself—a profile of nonrespondents having four respondent-demographic characteristics was developed. These four characteristics are sex, age, race, and family income. The process used to develop this completed profile involved a series of imputations which are described below.

For each of the 140 refusals, the racial composition of the block where they resided was obtained from 1970 block census data for the Detroit metropolitan area. Of the 80 blocks on which refusals were located, 58 blocks were listed as having no nonwhite residents. An additional 7 blocks had nonwhite populations which ranged from 1% to 30%. Eleven blocks had nonwhite populations which ranged from 69% to 100%. The refusals located on the 58 blocks containing no nonwhites and those located on the 7 blocks with minimal nonwhite populations were given the imputed racial characteristic "white." The refusals living on the eleven blocks with high percentages of nonwhites (69% to 100%) were labeled "nonwhite." Thus, with this procedure, 65 of the blocks are considered to contain white refusals and 11 are considered to contain nonwhites.

Four of the 80 blocks had almost equal percentages of white and nonwhite populations (49%, 55%, 60% nonwhite). The probability of selecting a respondent of either racial category, therefore, was about equal. For these 4 blocks, additional sets of data from the census were used—the percentage of owner- and renter-occupied households with blocks as heads, and the number of each such units on each block. The use of these measures resulted in refusals living on three of the four remaining blocks being given the nonwhite label and one the white label.[4] The final result of this procedure was a refusal sample of 15 nonwhites and 125 whites.

The income data for each refusal were imputed from the incomes of respondents on their respective blocks. When 2 or more respondents could be found on the block on which a refusal occurred, this refusal was given an imputed income equal to the mean block income of the respondents. This procedure was followed for 122 of the 140 refusals. The remaining 18 refusals were white females aged 65 or more. After a routine check of the mean income of all racial and age groupings, it was found that the imputed incomes for the refusals and those for the actual respondents were reasonably comparable for all racial and age groups except for older, white females. Theirs was greatly overestimated. Since there is evidence from both the responding sample and from past sociological studies that older people especially older women, tend to have relatively low incomes, an adjustment was made for these 18 aged, women refusals. They were given the mean income of all white females, 65 years old or more in the responding sample instead of their mean block incomes.[5]

Table 3 shows the differences between the 140 refusals and the 576 respondents on the four characteristics selected—age, sex, race, and family income, plus an additional variable "Type of Housing Unit." There is no substantial difference between the sex distributions in either group, although slightly more men tend to refuse than do women. However, there are more middle-aged and older persons and fewer nonwhites among the refusals than among the respondents. The mean income imputed for the refusals is also somewhat lower for the refusals than for the respondents. Considering the trend toward a decreasing percentage of "Single Family Unit" respondents noted in Table 1, one finds that a surprisingly large number of refusals are from this category of housing. In an effort to investigate how this distributional bias in respondent characteristics may affect areas of substantive interest in the survey itself, the following test for substantive bias was used.

The same three variables used in the estimation of interim bias are used in the estimation of terminal bias. The coefficients of the final response group (10+) on these three attitude

TABLE 3
COMPARISON OF THE DISTRIBUTION OF THE CHARACTERISTICS OF RESPONDENTS AND REFUSALS

| Characteristic | Respondents | Refusals | Percentage Difference [a] |
|---|---|---|---|
| | | Percentage | |
| SEX | | | |
| Male | 41.1 | 43.6 | −6 |
| Female | 58.9 | 56.4 | +4 |
| | N=576 | N=140 | |
| RACE | | | |
| White | 78.6 | 89.3 | −14 |
| Black | 21.4 | 10.7 | +50 |
| | N=574 | N=140 | |
| AGE | | | |
| 18-24 years | 15.0 | 4.3 | +71 |
| 25-44 years | 37.2 | 29.0 | +22 |
| 45-64 years | 35.4 | 46.4 | −31 |
| 65 years or more | 12.4 | 20.3 | −63 |
| Mean Age (X) | 41.8 years | 50.4 years | |
| | N=566 | N=140 | |
| FAMILY INCOME | | | |
| $000-$5,999 | 23.7 | | |
| $6,000-$9,999 | 15.5 | (See Appendix B) | |
| $10,000-$19,999 | 42.3 | | |
| $20,000 or more | 18.5 | | |
| Mean Income (X) | 10.65 ± $9,500 (See Appendix A) | 10.0 ± $9,000 (See Appendix A) | |
| | N=516 | N=140 | |
| TYPE OF HOUSING UNIT | | | |
| Detached single family | 77.7 | 81.1 | −4 |
| Two-four family or row house | 12.5 | 10.8 | +14 |
| Apartment, trailer, other | 9.8 | 8.1 | +17 |
| | N=569 | N=191[b] | |

a. Percentage difference is calculated as $\frac{\%\text{Respondents} - \%\text{Refusals}}{\%\text{Respondents}}$.

b. Includes all 177 refusals, plus 11 terminal not-at-homes and 3 terminal RAs.

variables are used to estimate (impute) a value for the refusals for these same variables in the following manner:

$$Y = b_0 + b_1X_5 + b_2X_6 + b_3X_7 + b_4X_8$$

where $b_0$, $b_1$, $b_2$, $b_3$, and $b_4$ are the coefficients estimated from the final response group (N = 499 or 500), and where $X_5$ is the sex of the refusal, $X_6$ is the race, $X_7$ is the age, and $X_8$ is the refusal's family income. Then in order to examine the extent of terminal bias, one computes the estimated means of the three variables for both the respondent and refusal samples, and a difference of means test to test for significant differences in such means.

Table 4 shows the results of the imputation procedure as outlined above. Test statistics indicate significant differences in the variances of the dependent variable values for respondents and refusals. This is true for all three variables. Considering the nature of the imputation procedure, one finds this predictable. In addition, however, the means of the two samples on the Race Discrimination item is significantly different at the .01 level. Of course, a more accurate estimate of the bias of terminal nonresponse can be made by comparing the respondent sample means to the means of the respondents plus refusals samples. These differences reflect the bias which resulted from a relatively high rate of refusals in the Detroit Area Study.

These findings suggest that, as in the case of interim bias, the distributional bias of respondent characteristics caused by terminal nonresponse may have differential effects on the substantive issues within the survey itself, though in its instance the differences (bias) are slight. Some areas or topics of interest may be affected more than others: the effect of refusals on the Race Discrimination item is somewhat greater than its influence on the other items. Distributional and concomitant substantive bias can thus be seen as having varying effects on the accuracy of probability statements depending on the variables used in calculating these statements. While there are differences between the sample means of refusals and respondents for all three variables considered, only one of these is significantly

TABLE 4
RESULTS OF IMPUTATION PROCEDURE

| Variable | (1) Respondents N=509 | (2) Refusals N=140 | Respondents + Refusals N=649 | t-test between columns 1 & 2 |
|---|---|---|---|---|
| Political party | 6.0751 | 6.0670 | 6.0734 | 0.151[a] |
| Subjective class | 2.1985 | 2.1702 | 2.1924 | 1.456[a] |
| Race discrimination | 4.0124 | 3.8511 | 3.9776 | 4.849 |

a. Not significant at the .01 level.

different—Race Discrimination. One must now look at the differences in the distributions of respondent-nonrespondent characteristics from Table 3 to see if they are sufficient to explain the substantive bias one observes in Table 4.

The linear regression of the Race Discrimination variable for the 576 (499 valid cases) respondents reveals that of the four explanatory variables used in the analysis, race and age are the most significant:

| | Partial Correlations |
|---|---|
| Race | .29308 |
| Age | −.17368 |

Nonwhite and younger respondents would tend to raise the mean value of the variable. The sample of refusals is "older" than is the sample of respondents; it is also "whiter." Therefore, the decline in the mean score is predictable given the characteristics of the persons in the refusing sample. The older and whiter sample has a mean score on the variable which indicates that more people are satisfied with the present rate of speed of government attempts to eliminate racial discrimination in employment.[6]

The slight decline in the mean for the subjective class variable can probably be explained by the lower mean income of the sample of refusals. A similar decline in the political party variable is due to the lower income and perhaps also to the older age of the sample of refusals (see Appendix B).

## SUMMARY AND CONCLUSIONS

For sociologists whose main interest is not sampling theory but substantive sociological theory, the method of estimating nonresponse bias outlined in this paper may be of more value than many of the methods found in the literature on nonresponse. The accumulation of facts and figures about the characteristics of nonrespondents such as their age, sex, race, and education, while of much value, gives sociologists little information concerning the impact of such characteristics on the kind of substantive work done by these researchers.

I have attempted in this paper to reinterpret nonresponse bias in the kinds of terms understood best by sociologists who are forever indebted to survey research as a valuable research tool, but who have neither the desire nor inclination to explore the intricacies of sampling theory. Yet, sampling theory and its implications for the accuracy and reliability of sociological research cannot be ignored by any social scientist. And the problem of nonresponse deserves more than a casual reference via a footnote. Much more research should be done on nonresponse, especially studies which make the concept of nonresponse bias more understandable to the social scientists who frequently use survey research. The procedure outlined in this paper must be duplicated to see if the findings are merely a fluke due to a faulty imputation procedure or whether this is a reliable way of estimating nonresponse bias.

No doubt any researcher who works with the problem of race discrimination opinions would control for race in any interpretations that were made about the public's opinion on this issue. However, the fact that difference in the sample means was also caused by the age of the nonrespondents suggests that on such an issue age must also be considered. For political scientists and pollsters who frequently are concerned with the percentages of persons having a given opinion on an issue such as race discrimination, nonresponse bias may greatly affect the validity and accuracy of their interpretations.

It must also be noted that the substantial change noted in the coefficient for race on the Race Discrimination item as seen in Table 3 cannot be explained by changes in the percentage of blacks in the sample between the various calls (see Table 1). It may well be that surveys are systematically missing a more middle-aged and older segment of the population in general population samples whose opinions on such issues as race and other social issues are different from those of younger persons who are inclined to accept interviews. These nonrespondents could also be vastly different from the sample of older persons who do become respondents. Studies by Benson et al. (1951) and Schuman and Gruenberg (1970) have commented on the conservative attitudes of white survey nonrespondents. These questions deserve a great deal of investigation. One practical step for future researchers would be to increase persuasion efforts for middle-aged respondents, as well as to consider the use of such techniques as subsampling and oversampling this group when possible. All such efforts will help to understand and control the kind of substantive bias which, though seemingly slight, is evident in this analysis of nonresponse.

## NOTES

1. For example, see the American Statistical Association Conference (1974). Participants in the conference reported that response rates of 60% to 65% are average completion rates for current surveys. In the Detroit Area Study from which the data for the current analysis is taken, response rates have dwindled from 85% or more in the 1950s to current rates of near 70% or less.

2. In sampling theory the terms precision and accuracy are used to refer to the effects of bias and variable errors. It can be shown that the total survey error is a function of variable errors due to sampling and the bias due to measurement errors.

$$\text{Total Error} = \sqrt{\text{VE}^2 + \text{Bias}^2}$$

High rates of nonresponse will affect the latter term, and consequently increase the total survey error (Kish, 1965: 510). These terms are also roughly synonymous with the terms reliability and validity used in psychology and referred to by Campbell and Stanley (1963).

3. Eight variables chosen for this analysis are standard socioeconomic types (age, sex, race, occupation, family income, education, marital status, and religion). "City of residence" was added to investigate possible Detroit-suburban response differences. "Years in the Detroit Area" was added to check for the effects of geographical mobility and migration patterns. The "housing type" variable was added to investigate respondent accessibility effects.

4. The 2 additional census descriptions already referred to were also used to check on the accuracy of the racial designations for the 76 other blocks.

5. The imputed mean income for these refusals using the mean block income was approximately $11,000. With mean income for their age group, it was lowered to $6,000.

6. Because of the fact that there were 37 refusals with missing data that were not used in the nonresponse analysis, a final check of the racial characteristics of these refusals was made. Using the race imputation procedure outlined above, one finds that 30 of these would have been labeled "white" and 7 "black." If these 37 are then added to the 140 refusals with no missing data, the percentage of blacks in the sample of refusals would increase from 10.7 to 14.1. This small increase would not have seriously affected the results obtained in this analysis.

## REFERENCES

American Statistical Association Conference (1974) "Report on the ASA Conference on Surveys of Human Populations". Amer. Statistician 28, 1 (February) 30-34.

BENSON, S., W. P. BOOMAN and K. E. CLARK (1951) "A study of interview refusals." J. of Applied Psychology 35 (April): 116-119.

BIRNBAUM, Z. W. and M. G. SIRKEN (1950) "Bias due to non-availability in sample surveys." J. of Amer. Stat. Ass. 45 (March): 98-111.

CAMPBELL, D. T. and J. C. STANLEY (1963) Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

COCHRAN, W. G. (1963) Sampling Techniques. New York: John Wiley.

DEMING, W. E. (1950) Some Theory of Sampling. New York: John Wiley.

DUNKELBERG, W. C. and G. S. DAY (1973) "Nonresponse bias and callbacks in sample surveys." J. of Marketing Research 10 (May): 160-168.

HANSEN, M. H. and W. H. HURWITZ (1946) "The problem of non-response in sample surveys." J. of the Amer. Stat. Ass. 41: 517-529.

HENDRICKS, W. A. (1949) "Adjustment for non-response bias caused by non-response in mailed surveys." Agricultural Economics Research 1 (April): 52-56.

KISH, L. (1965) Survey Sampling. New York: John Wiley.

LAGAY, B. W. (1969-70) "Assessing bias: a comparison of two methods." Public Opinion Q. 33 (Winter): 615-618.

POMEROY, W. B. (1963) "The reluctant respondent." Public Opinion Q. 27 (Summer): 287-293.

ROBINS, L. N. (1963) "The reluctant respondent." Public Opinion Q. 27 (Summer): 276-286.

SCHUMAN, H. and B. GRUENBERG (1970) The Impact of City on Racial Attitudes. Ann Arbor, Mich.: Institute for Social Research.

SCHWIRIAN, K. P. and H. R. BLAINE (1966-1967) "Questionnaire return bias in the study of blue collar workers." Public Opinion Q. 30 (Winter): 656-663.

SUCHMAN, E. A. (1962) "An analysis of 'bias' in survey research." Public Opinion Q. 26 (Spring): 102-111.

SWINDEL, B. F. (1974) "Instability of regression coefficients illustrated." Amer. Statistician 28 (May): 63-65.

U.S. Department of Commerce. Bureau of the Census (1970) Block Statistics Detroit, Michigan Urbanized Area. 1970 Census of Housing HC (3)-120.

# APPENDIX A

## VARIABLES USED IN THE REGRESSION ANALYSIS

### Political Party (strength of affiliation)

This item follows a question which obtains the respondent's party affiliation.

Do you consider yourself a strong Republican or a not very strong Republican?

Do you consider yourself a strong Democrat or a not very strong Democrat?

Do you generally lean more toward the Republican or Democratic Party? (asked of Independents)

1. Strong Republican
2. Not strong (Republican)
3. Lean Republican
4. Neither party
5. Lean Democrat
6. Not strong (Democrat)
7. Strong Democrat

### Subjective Class

If you were asked to use one of these four names for your social class, which would you say you belong in:

| | Recoded as: |
|---|---|
| 1. Working Class | 1. Lower Class |
| 2. Middle Class | 2. Working Class |
| 3. Lower Class | 3. Middle Class |
| 4. Upper Class | 4. Upper Class |

### Race Discrimination

Do you feel that the government is moving much too fast, too fast, much too slow, too slow, or just about right in its efforts to eliminate racial discrimination in employment?

1. Much too fast
2. Too fast
3. Just about right
4. Too slow
5. Much too slow

### Respondent's Age

Respondent's exact age in years was used. These ranged from 18 (lower limit set in study) to 84 among the final respondents.

## APPENDIX A (Continued)

### Respondent's Sex (Dummy Variable)

1. Males
2. Females

### Respondent's Race (Dummy Variable)

1. Whites
2. Nonwhites (included only four persons who are not black)

### Respondent's Family Income

A categorical variable coded as follows:

1. $000-999
2. $1,000-1,999
3. $2,000-2,999
4. $3,000-3,999
5. $4,000-4,999
6. $5,000-5,999
7. $6,000-6,999
8. $7,000-7,999
9. $8,000-8,999
10. $9,000-9,999
11. $10,000-11,999
12. $12,000-14,999
13. $15,000-19,999
14. $20,000-24,999
15. $25,000 or more

## APPENDIX B

**Because of the fact that the incomes for the refusals are the mean block incomes of the respondents, many are not in the same units as those of the respondents (see Appendix A). The following is a listing of the exact units in which the incomes of the refusals appeared.**

| Value | Number Persons | Value | Number of Persons |
|---|---|---|---|
| 3.00 | 1 | 10.50 | 5 |
| 3.50 | 1 | 10.57 | 3 |
| 4.33 | 1 | 10.60 | 3 |
| 5.00 | 1 | 10.67 | 1 |
| 6.00 | 2 | 10.80 | 1 |
| 6.60 | 3 | 11.00 | 1 |
| 6.67 | 2 | 11.20 | 2 |
| 6.94 | 18[a] | 11.25 | 2 |
| 7.00 | 1 | 11.33 | 1 |
| 7.33 | 2 | 11.40 | 4 |
| 7.40 | 1 | 11.50 | 1 |
| 7.50 | 3 | 11.60 | 1 |
| 8.33 | 5 | 11.67 | 2 |
| 8.40 | 1 | 11.75 | 3 |
| 8.67 | 4 | 11.80 | 1 |
| 8.71 | 1 | 12.00 | 4 |
| 8.83 | 1 | 12.20 | 5 |
| 9.00 | 4 | 12.50 | 6 |
| 9.25 | 1 | 13.00 | 3 |
| 9.33 | 1 | 13.25 | 2 |
| 9.50 | 3 | 13.33 | 1 |
| 9.71 | 1 | 13.50 | 3 |
| 9.80 | 1 | 13.80 | 2 |
| 10.00 | 4 | 14.00 | 6 |
| 10.20 | 3 | 14.50 | 2 |
| 10.25 | 3 | 14.85 | 1 |
| 10.40 | 4 | 15.00 | 1 |

a. White females 65 years or more.