ADJUSTING FOR NONRESPONSE IN A TELEPHONE SUBSCRIBER SURVEY

James H. Drew, GTE Laboratories and Robert M. Groves, University of Michigan
James H. Drew, GTE Laboratories, 40 Sylvan Rd. Waltham, MA 02254

Keywords: *Weighting Classes, Poststratification, Telephone Surveys, Administrative Data,*

Introduction

Each month, GTE administers a sizable number of telephone surveys of subscribers to its local telephone service, in order to gauge the extent of customer satisfaction with the service, to assess the impact of various service improvement programs, and to diagnose perceived problems which may lie outside the scope of the Corporation's internal measurements. From its billing records, each individual telephone company selects simple random samples of its residential and its business customers each month, from which a specified quota of individuals complete a fairly lengthy telephone-administered interview. Given the desire that the completed interviews accurately represent the perceptions of the general populations of GTE telephone subscribers, concerns have been raised over high levels of nonresponse, currently running at roughly 40-45% for residential customers and 25-60% for various segments of business customers.

Despite these high rates of nonresponse, there is hope that the nonresponse problem can be controlled through a combination of increasing interview attempts, loosening interview qualifications, and *post-hoc* data adjustments. The latter remedy is often applied by locating or defining some variable which "explains", or is plausibly associated with, an individual's nonresponse and whose data are available for the entire sample or the entire population from which the completed interviews are gathered. In the surveys we consider, candidate variables for this sort of use are available at the individual level through the telephone repair and service billing information GTE records as part of its administrative information.

These surveys are a major part of GTE's performance monitoring program, both in their extent and in the seriousness with which the results are interpreted. Consequent cost and scheduling concerns effectively eliminate the possibility of any casual deviation from the three-call rule by which the current survey interviews are administered. Thus, one must evaluate the consequences of unit nonresponse without recourse to any form of complete data. The interim strategy we have adopted during this study is to attempt several types of nonresponse adjustment with existing data and compare the adjusted estimates to gain an informal notion of the sensitivity of our survey estimates to the adjustment techniques.

We attempt three major types of nonresponse adjustment. For the first, we search for administrative data which are available for each unit in the sample, to be used to form weighting classes, as described by Oh and Scheuren (1983), for example. We note that,in this survey one might reasonably hope for variables which are associated with both the response mechanism and with survey responses, since telephone company data are likely to be related to telephone usage and hence to telephone survey availability, while that same data may well be also related to the telephone service evaluation that is the survey's subject. The second major type of adjustment, the administrative data are used to construct an explicit dichotomous variable model to predict response, which is then used as part of a selection bias model. The third adjustment strategy exploits the fact that these surveys are conducted by telephone with specific callback policies. Numbers of responses at each call are used to estimate response probabilities within (fairly flexibly defined) categories, from which adjustments are made.

In support of the comparison of the estimates based on these strategies, we devote the next section to a description of some of the characteristics of calling dispositions for our sample. In subsequent sections, the adjustment strategies are applied and compared.

Calling Histories

Seven hundred thirty nine households were selected as the target sample for the TEL-CEL General Services survey of February 1988. Each household was called by telephone up to three times, generally using a day-night-weekend calling sequence. Each call was classified according to its survey response result based on nine different categories. To condense the possible call patterns over the three-or-fewer calls, each household was uniquely categorized as one of the following:

COMPLETED - At one of the calls, the interview was completed.

REFUSAL - The final call disposition was a refusal.

OTHER CONTACT - Except for REFUSALs, contact was made at some call without the interview being completed. This category includes language problems and ineligible (e.g. child) contacts.

NOT AVAILABLE - At some call, the head of household was not available, although someone in the household was contacted.

NA,NA,NA - At each call, the telephone was not answered. Contact with an answering machine only is counted as a NA for this study.

OTHER NONCONTACT - Except for those with the NA,NA,NA pattern , no contact was made. This would include a combination of NAs and Busy signals.

NON-WORKING NUMBER - The final determination was that the called number was not in service.

Coding by the interviewers was not consistently performed. Examination of the written call records indicated that a small number (10) of sampled units were coded as "Not Eligible" or "Designated Respondent Not Available" when a "Refusal" designation would have been more appropriate. Also, since the difference between a No Answer and contact with an answering machine is unrelated to a subject's availability or willingness to respond, answering machine contacts were recoded as "NA." These judgements are reflected in the table below.

Calling Patterns

Disposition Type	Number	Percent
Complete	395	53.5
Refusal	71	9.6
Other Contact	18	2.4
Not Available	37	5.0
NA,NA,NA	157	21.2
Other Noncontact	31	4.2
Non-Working Number	30	4.1

It is apparent that most nonresponse is due to sample units never being available to answer the telephone within the three call period. The next most frequent nonresponse disposition is a refusal.

The preponderance of nonavailability (NA,NA,NA) nonresponses in this particular survey parallels the findings reported by Sebold (1988) for a study based on the National Crime Survey. In that study, a substantial number of interviews were completed only after 25 or more calling attempts. A follow-up survey of these hard-to-contact respondents revealed that many noncontacts were away from their homes during the survey, with a near-majority (44%) being absent during the entire survey period. These findings are consistent with a recent GTE experiment in which a large number of unanswered telephones remained unanswered after nine callbacks.

In the next section, we consider the association of these nonresponse dispositions with repair request information gathered by a particular telephone company.

Repair Information and Response Rates

Consider the table below showing disposition type, as defined earlier, and three categories of repair report: no repair, repair call(s) within the Previous year, and all repair calls more than one year previous.

It is apparent from this table that repair recency is associated with sample response in two ways. First, completion rates are higher for those with repairs in the last year, while complete noncontact (NA,NA,NA) rates decrease with the interval between the most recent repair and the interview attempt. Also, refusal rates are higher for those with repair-to-interview attempt intervals exceeding one year. None of the other disposition codes show much difference among repair recency categories.

It is possible at this point to speculate that the complete noncontacts are to some extent actually non-working numbers which are not assigned out-of-service recordings by which the interviewer can make a definitive "Non-Working Number" classification. This however, is unlikely since sample is selected from current GTE billing records.

Call Disposition by Repair Recency

	Complete	Refusal	Other Contact	Not Available	NA,NA,NA	Other Noncontact	Non-Working Number
No Repair	211 50.00	36 8.53	7 1.66	23 5.45	106 25.12	17 4.03	22 5.21
Repair in Last Year	160 59.48	24 8.92	10 3.72	12 4.46	46 17.10	11 4.09	6 2.23
Repair, but not Within Last Year	24 50.00	11 22.92	1 2.08	2 4.17	5 10.42	3 6.25	2 4.17

Furthermore, while one might expect in this event that repair requests for the complete noncontacts are generally not as recent as those for respondents, this is not the case: for those noncontacts with repair requests, the mean time between the last request and the first interview attempt is 163.02 days (standard deviation = 140.68) while the analogous mean for respondents is virtually equal, at 164.62 days (standard deviation = 151.51).

One might also expect that the relative unavailability of a head of household to answer this survey indicates a more general absence from the household, which could possibly reflect his inability to notice or to report telephone problems in need of correction. This might be reflected in a tendency for nonrespondents to make fewer reports, or for nonrespondents to ignore more non-critical problems. One observes a significantly larger proportion of No Repairs among the complete noncontacts compared with the Completes, in support of the former suggestion. As a partial response to the latter question, we note that there is a substantial difference in the proportion of not-out-of-service calls between respondents and nonrespondents whose final call disposition was No Answer. In the respondent group, 26.8% reported at least one non-out-of-service problem, but of the No Answer group, only 13.5% (of 148) reported such problems. In contrast, the relative proportions of out-of-service problems reported was nearly equal, 9.9% for the Respondents to 8.1% for the No Answers. This is consistent with our suggestion above.

Thus, we can argue that complete noncontact is associated with repair reports categorized as being made earlier or later than one year before interviewing and that refusals are associated with the number of reports made at least one year prior to interviewing.

The Effect of Weighting Class Adjustment

In accordance with our findings of the previous section, five weighting classes were formed which were composed of the following classes of subscribers:

1) Those making no repair requests,
2) Those making only one request in the last year, and none previously,
3) Those making two or more requests in the last year, and none previously,
4) Those making one request, that request being at least one year prior to the interview period, and
5) Those making two or more requests, the most recent being at least one year prior to the interview period.

The variable based on these weighting classes was found to be significantly associated with subscriber response disposition (mostly through complete noncontact and refusals) and with the respondents' general quality rating in this survey.

To illustrate the effect of this adjustment, consider the proportion of subscribers rating their general telephone service using a particular set of response categories (such as "Excellent" or some other favorable response). The observed proportion among the respondents is 0.6401, with a standard deviation of 0.0243. Let $\underset{\sim}{f}$ be a column vector of the relative sizes of the five weighting classes in the sample, and let $\underset{\sim}{p}$ be a column vector of the proportions observed for the respondent in each weighting class. Then the adjusted estimator is

$$p^* = \underset{\sim}{f} \; {*}\underset{\sim}{p}',$$

and its approximate standard error is given as the square root of the scalar quantity

$$\mathrm{Var}(p^*) = \underset{\sim}{f}\,'{*}\, v(\underset{\sim}{p}){*}\underset{\sim}{f} + \underset{\sim}{p}'{*}v(\underset{\sim}{f}){*}\underset{\sim}{p},$$

where $v(\underset{\sim}{p})$ is the estimated covariance matrix of the observed proportions, and $v(\underset{\sim}{f})$ is the estimated covariance matrix of the sample quantity $\underset{\sim}{f}$.

For our data, we find that $p^* = 0.6424$ with standard deviation 0.0242, both quantities being nearly the same as those from the naive calculation which ignores the nonresponse. The insignificance of the adjustment is apparently a result of the two most changed weighting classes being associated with nearly identical quality ratings. Note also that the adjusted estimator has a slightly smaller standard deviation than the naive estimator, despite taking nonresponse effect into account. Apparently the weighting classes, being associated with the Quality rating, aid the precision of p^* by acting as a poststratification variable.

The Effect of Using a Selection Bias Model

A slightly different way of exploiting the administrative data developed above is to construct a model for selection bias, following the work of Heckman (1979). The essential idea is to use the administrative data, categorized as before, to construct a model for the response/nonresponse of the entire sample. Thus, if W_i is the indicator of the ith weighting class and π_i is the probability of response, one fits the model

$$\pi_i = \zeta^{-1}(\mu + W_i),$$

where ζ^{-1} is a specified function from $(-\infty, \infty)$ to $[0,1]$, usually a logit or a probit, and the predicted values $\hat{\pi}_i$ are used as variables in the equation

$$E(y_i) = v + a\hat{\pi}_i,$$

where y_i is the variable of interest. In our illustration, y_i is the logit of the proportion of the responses of interest in the survey population. The best candidate for weighting classes is found to be identical to that of the previous section. From this equation, predicted values of y_i are calculated for each poststratum defined by values of $\hat{\pi}_i$, and these predicted values (or appropriate transformations) are weighted by the known sample poststratum sizes to construct an adjusted variable estimate.

To estimate the population proportion, let y_i be the logit of this proportion and use the techniques just described to fit the model given above:

$$E(y_i) = 1.1221 - 1.1004\hat{\pi}_i,$$

from which one calculates the proportions

$$\frac{e^{y_i}}{1+e^{y_i}}$$

and weights them by the known sample sizes of each poststratum. Despite the significance of the regression given above, the adjusted proportion of responses is 0.6303, which is not significantly different from the unadjusted estimate.

The Effect of Using Survey Wave Data

Because much of the nonresponse in this survey is due to the sample unit's unavailability to answer the survey, as indicated by the preponderance of NA,NA,NA call dispositions, one can exploit the consequent approximate independence of individual calling results. One simple form this might take is to estimate response probabilities for each survey call for some categorization of the data, so ρ_k is the probability of a response at a given call for a unit in the kth category, and then model the proportion f_k of the kth category as proportional to

$$\frac{r_k}{1 - (1-\rho_k)^R},$$

where r_k is the number of responses in the kth category and R is the total number of calls. Note that the categories can be constructed from any combination of variables available for the respondents, and can indeed be based on the variable of interest; category information from the full sample need not be used. Weighting class information can be incorporated by estimating f_{wk}, where f_{wk} is the proportion of the kth category in the wth weighting class, $k=1,2,\ldots,K$; $w=1,2,\ldots,W$. Then an overall proportion for the kth category is estimated as

$$\sum_{w=1}^{W} n_w f_{wk}$$

where n_w is the sample size of the wth weighting class. More complicated models can of course be used to estimate f_{wk}, $w=1,2,\ldots,W$, $k=1,2,\ldots,K$, generally based on hypothesized differences in response rates across calls, and on possible categorical compositions of those in the sample who cannot furnish interviews even with extensive calling. See Drew and Fuller (1980) for details.

To estimate the proportion considered in the preceding sections, let the Quality variable itself define the categorization, and use the previously defined weighting classes. Current evidence suggests no difference in response probabilities across weighting classes, so constant ρ_k values were used for each category. The adjusted percentages for the five weighting classes are

$$\{0.6480, 0.6437, 0.5135, 0.5798, 1.000\}$$

which give a final estimate of 0.6275 for the population proportion, which we note is not significantly different from the naive estimate which ignores the nonresponse.

In view of the simplicity of this callback model and the tenuousness of its assumptions (such as the composition of the unreachable households), other data and slightly more complicated models were also used. In one set of models data from the 9-call experiment of April 1989 were used to estimate response probabilities ρ_k based again on the Quality variable, and these in turn were used to construct estimates of the percentage for Quality, say $\hat{z}_9$. The exact form of these estimates depends on the callback model used, the existence of nine calls allowing relatively great flexibility in those forms. One line of modeling postulated individual response probabilities ρ_{rk} for $r < R^*$, and category dependent $\rho_{rk} = \rho_k$ for $r \geq R^*$. Another line of modeling assumed that the unreachable households had categorical compositions like only those reached on sufficiently high call numbers. For percentages $\bar{z}_3$ based on three calls from the April experiment, and the three call percentages $\bar{x}_3$

from the original experiment, an adjusted estimator is given by

$$\bar{x}_3 \frac{\hat{z}_9}{\bar{z}_3}.$$

These adjusted estimates, however, are again little different from the naive estimates ignoring the nonresponse.

Summary

We have explored the issue of nonresponse in an industrial telephone opinion survey for which some limited administrative data are available for nonresponse adjustment. Telephone repair information about the number and recency of repairs in an eighteen month period was found to be significantly associated with both noncontact and refusal nonresponse, and with telephone service ratings. However, the weighting class adjustment based on this information yielded a quality rating estimate which was not statistically different from the unadjusted estimate.

The effect of this survey's nonresponse was further assessed by two other kinds of nonresponse adjustment: Heckman's selection bias model, and various callback models. None of these general techniques produced estimates which were significantly different from the unadjusted estimate. This notion must, of course, be verified by further experimentation in which greater efforts are made to interview noncontacted households and to convert refusing households.

References

Drew, J. and W. Fuller (1980). Modeling nonresponse in surveys with callbacks, *Proceedings of the Section on Survey Research Methods of the American Statistical Association*.

Heckman, J. (1979). Sample selection bias as a specification error, *Econometrica*, 47, 1, 153-161.

Oh, H. and F. Scheuren (1983). Weighting adjustment for unit nonresponse, in *Incomplete Data in Sample Surveys* (Vol 2), ed. W. Madow, I. Olkin and D. Rubin, New York: Academic Press.

Sebold, Janice (1988). Survey period length, unanswered numbers, and nonresponse in telephone surveys, in *Telephone Survey Methodology*, ed R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls II, and J. Waksberg, New York:John Wiley and Sons.