TELEPHONE COLLECTION IN AN ESTABLISHMENT SURVEY OF JOB OPENINGS

Stuart Scott, Bureau of Labor Statistics
Mary Yurachek, Occupational Safety and Health Administration

To the Memory of Barbara Boyes

1. Introduction and Design

During 1979 and 1980 the Bureau of Labor Statistics, at Congress' request, conducted a pilot survey of job openings (Plunkert, 1981). In four states, establishments were asked the number and occupation of their current job openings, plus some related data, as of the last business day of the current quarter.[1] There are no plans to mount a full-scale survey, but it is hoped that some findings from the pilot tests can be applied to other surveys. Initially, data were collected by mail. For the last three quarters, telephone collection was used for small firms, i.e., firms with employment under 50. The aim of this paper is to describe our test of the quality of telephone collection in these firms. Our conclusions are that personal visit collection appeared to be better, but that the differences were small enough to make telephone collection viable.

In the very first quarter of the survey, phone reminders and "field follow-ups" were conducted to cut down nonresponse. For the field follow-up, data were collected by phone or personal visit. Telephone was successful in boosting the volume of response, and looked promising for obtaining more timely reponses. The time factor was important, since most small firms do not maintain pertinent records. Collecting job openings data by telephone appeared feasible because usually in small firms (1) the number of job openings is small and (2) personnel functions are concentrated in one location. As well as evaluating quality, the test was to observe costs in staff time. Information on staffing requirements and some dollar costs appear in the Appendix.

With limited resources available, a design was sought which would isolate collection method from other factors influencing the response variable, number of current job openings. Other factors considered were state, size (3 size classes within the 0-49 range), reference period or quarter, kind of business, geographical location, and interviewer. The experimental design which was adopted measured the effects of size and reference period and controlled to some degree for kind of business and geography. Handling of interviewers varied among the states, and will be discussed in the next section.

Figure 1 depicts the design, which features a set of three Latin squares for each state, a balanced set across the three size classes (cf. Winer, 1971). Each Latin square allocates collection method according to quarter (reference period) and panel. "P" denotes personal visit and "T" and "T'" telephone collection. Introducing panels into the design was intended to reduce the contribution of such factors as geography and kind of business to experimental error. Units participating in the pilot survey were grouped into triplets, by hand- matching as closely as possible on these factors. Then, triplets were selected by a probability mechanism and units within triplet randomly assigned to panels. Each panel within a size class consisted of one unit from each of 25 triplets selected for the experiment. To the extent that the matching worked, panel effects would be small.
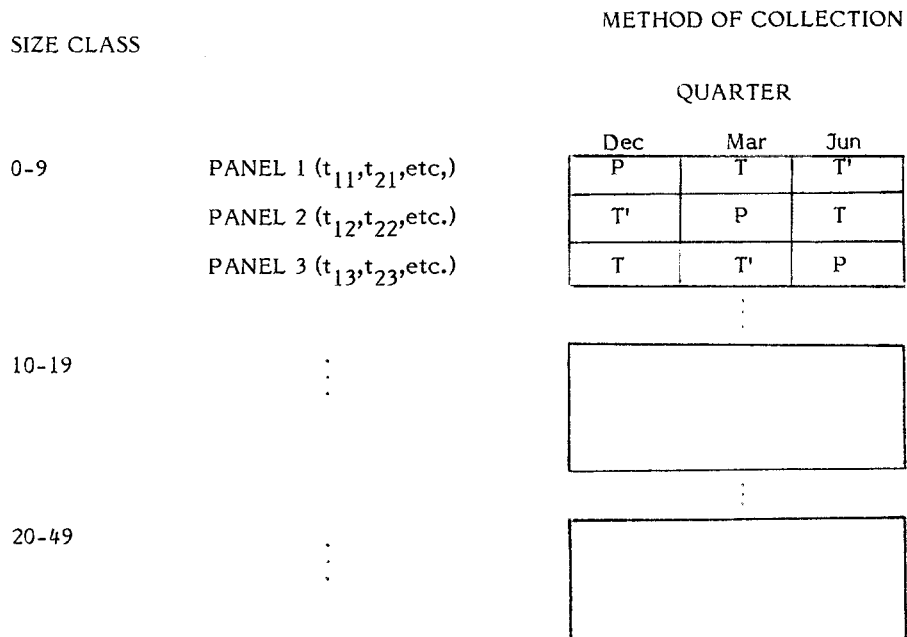
FIGURE 1. Latin Square Design

SIZE CLASS

METHOD OF COLLECTION

QUARTER

	Dec	Mar	Jun
0-9 — PANEL 1 $(t_{11}, t_{21}, etc.)$	P	T	T'
PANEL 2 $(t_{12}, t_{22}, etc.)$	T'	P	T
PANEL 3 $(t_{13}, t_{23}, etc.)$	T	T'	P

10-19

20-49

Table 1. Percentage of Respondents

Collection method	12/79	3/80	6/80
Total	84.9	83.6	83.5
Personal visit	82.2	84.1	84.4
Telephone	86.2	83.3	83.0
Size Class	12/79	3/80	6/80
0-9	83.3	82.3	81.2
10-19	85.4	83.4	84.6
20-49	86.0	85.1	84.7

2. Survey estimates

Survey estimates, as well as ANOVA-type estimates, were made. The Job Openings pilot employed a probability sample, and the selection of triplets for the test involved well-defined subselection probabilities. Thus, standard Horvitz-Thompson estimates of total job openings for small firms could be made separately from telephone collection and personal visit collection. For the survey estimates, ability to detect differences due to collection method was dependent upon the effectiveness of the matching process at minimizing panel differences.

In terms of survey response, cooperation was just about the same for telephone and personal visit, as seen in Table 1. The figures used here are for purposes of comparison; they are not response rates in the pure sense.[2/] This was taken as supportive evidence for telephone collection, both because the rates were high and because they were not higher than those for personal visit. Higher rates for the telephone collection would have evoked suspicion, out of concern for respondent saying "no openings" when they meant "no interest in responding."

Cell estimates of job openings showed a great deal of variability, due essentially to the rarity of job openings. Thus, it was decided to switch from a variable measuring level of job openings to a variable indicating their occurrence, specifically, percent employment in units with one or more current job openings. This statistic measures percent of firms with current openings in an employment-weighted manner to facilitate summarization across size class.

Results were summarized to 24 cells: 4 States x 3 quarters x manufacturing--nonmanufacturing. Percentages for telephone and personal visit collection appear in Figure 2 (Massachusetts and Utah only) and Table 2. While there was not a consistent relationship between the methods, in 18 out of 24 cells, the personal visits gave higher values. A sign test of no difference between methods yielded a P value of 0.01. Averaging over all cells revealed that personal visits located approximately 5% more firms with openings than collection by telephone (cf. Table 3). The higher percentages of firms with openings for personal visit coincided with expectations, since earlier findings in the pilot survey revealed most errors to be in the direction of undercounting.

Table 2. Per cent Employment in Units with Openings

	Manufacturing		Nonmanufacturing	
	Personal Visit	Telephone	Personal Vist	Telephone
Florida				
12/79	34.5	24.5	33.7	14.6
3/80	23.5	22.6	24.7	18.1
6/80	32.1	11.7	12.1	10.8
Massachusetts				
12/79	15.0	23.9	15.1	14.1
3/80	25.1	13.6	12.5	18.1
6/80	16.6	20.5	13.3	11.0
Texas				
12/79	13.6	21.3	25.6	17.9
3/80	18.2	9.2	9.6	15.8
6/80	15.5	3.7	24.9	7.0
Utah				
12/79	29.4	24.2	13.3	20.4
3/80	14.8	13.7	18.1	10.4
6/80	36.2	14.5	22.5	7.8

Figure 2. Per Cent Employment in Units with Openings, by Quarter


```
                    ┌─────────────────────────────────┐
                    │            LEGEND               │
                    │ ────────── PERSONAL VISIT       │
                    │ ---------- TELEPHONE COLLECTION │
                    └─────────────────────────────────┘
```

MANUFACTURING   NONMANUFACTURING

Massachusetts
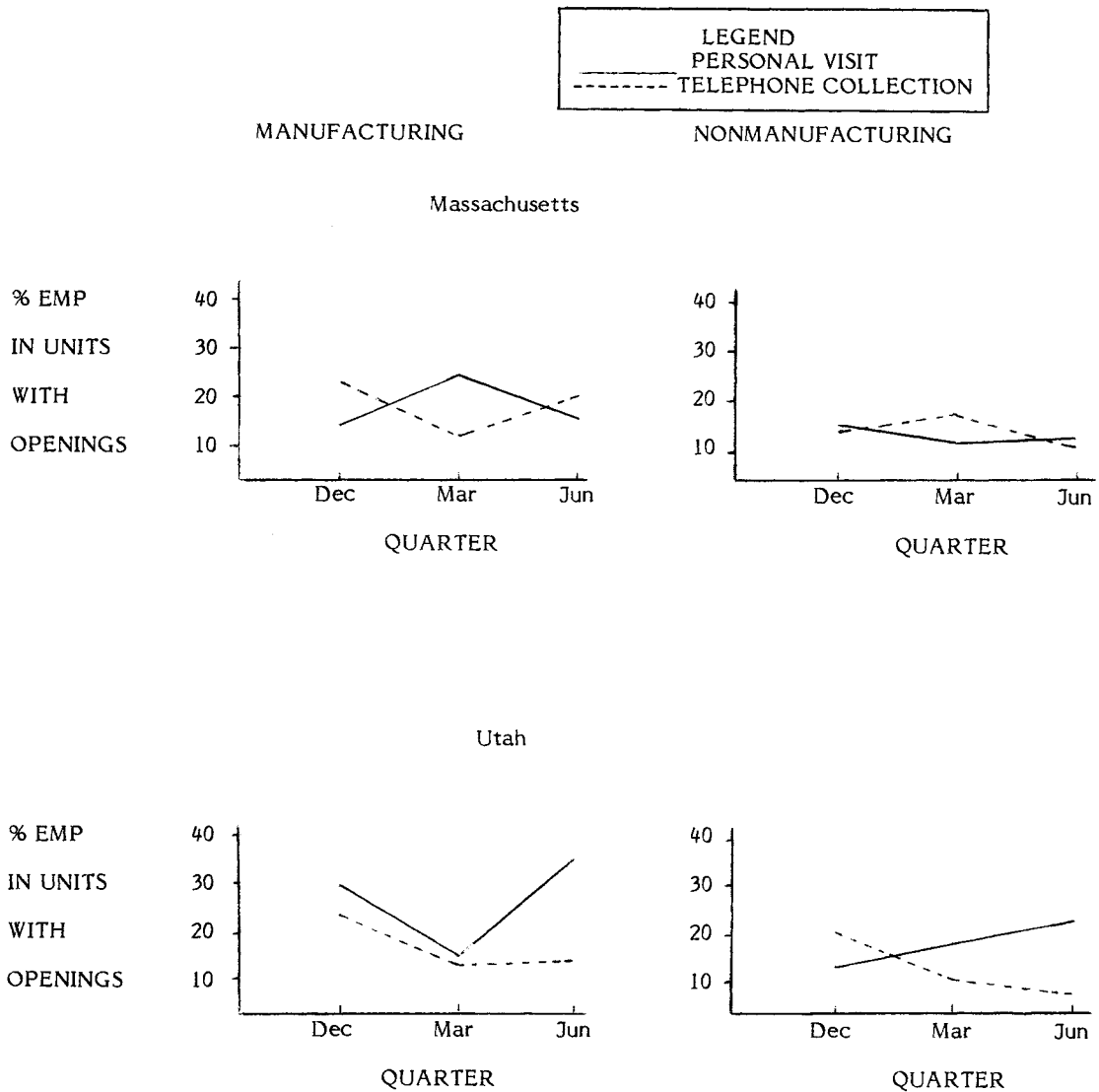
Utah

Table 3. Percent by which Personal Visit Exceeds Telephone
(Percent Employment in Units with Openings)

Quarter
December 1979	2.4
March 1980	3.1
June 1980	10.8

State
Florida	9.7
Massachusetts	-0.6
Texas	5.4
Utah	7.2

All Cells   5.4

Two possible explanations were found for these differences: Telephone collection appeared to deteriorate the last quarter. There, personal visit was higher in 7 out of 8 cases, and the percentage difference exceeded 10%. In two States, Florida and Utah, interviewers were assigned to collection method, and the best interviewers were assigned to personal visit. On the other hand, in Massachusetts, where neither factor applied, the two methods were virtually indistinguishable. There, the survey continued for two or three more quarters; June was not the end of the survey. Interviewers were systematically assigned across methods, so that each interviewer used both methods each quarter. Overall, examining the survey estimates by quarter (Table 3), differences appear significant only in the last quarter.

3. ANOVA estimates

Model   The model selected for the design described in Section 1 (cf. Figure 1) is

$$y_{ijkl} = \mu + \alpha_i + u(\alpha)_{j(i)} + \gamma_k + \delta_l + (\alpha\delta)_{il} + \epsilon_{ijkl},$$

where α_i denotes a size effect, $u(\alpha)_{j(i)}$ denotes panel effect (random effect nested within size), γ_k is reference period effect, δ_l is collection method effect, $(\alpha\delta)_{il}$ is the size-method interaction, and ϵ_{ijkl} is the random error component. Keep in mind that for each value of i only one-third of the jkl combinations actually occur. Each state is analyzed separately, due to the differences in both the economic structure and the conduct of the survey.

Higher-way interactions were omitted from the model as being unlikely to be significant. Also, no size-reference period interaction was anticipated. Size-treatment interaction was considered possible; for instance, the larger firms, 20-49, might have more problems with telephone collection than the others. While the design made it possible to isolate factor effects, significance testing depended on whether interactions were present among the Latin square factors. This led us to apply the Tukey test for additivity to the Latin squares (Scheffé, 1959). Since only 1 out of 12 yielded a large test statistic, omitting these interactions was considered acceptable. The only potential interaction appeared to be the reference period-collection method combination. Even if this interaction were present, it would spoil testing for the panel effect, but not nullify testing for the other main effects of the square. Most cells contained 16-25 observations. We chose to avoid the unequal cells situation by using cell means, rather than individual observations.

Limitations

One of the basic assumptions of ANOVA is equal variances for the observations. Departure from this assumption can be expected here due to unequal observations in the cells and some increase in standard deviation with size. On the other hand, the use of cell means lends some stability and the imbalance in observations per cell is relatively mild. The weakest aspect of applying ANOVA is the response variable itself. The low vacancy rate mentioned previously

Table 5.   ANOVA Statistics

Source	DF	Florida			Massachusetts			Texas			Utah		
		SS	F	Pr > F	SS	F	Pr > F	SS	F	Pr > F	SS	F	Pr > F
Size	2	.591	6.32	.02	1.166	22.41	.0002	.065	2.20	.16	.324	5.96	.02
Quarter	2	.419	4.48	.04	.155	2.98	.10	.074	2.50	.13	.028	.51	.61
Panel (Size)	6	.325	1.16	.40	.224	1.43	.29	.079	0.89	.53	.352	2.16	.14
Coll. Method	2	.066	0.71	.52	.023	0.45	.65	.020	0.70	.52	.093	1.71	.23
Size*Coll. Method	4	.096	0.51	.73	.185	1.77	.21	.101	1.70	.23	.091	.84	.53
Error	10	.468			.260			.148			.271		
Orthogonal Contrasts													
Telephone vs PV		1.41	.26			.63	.45		.02	.89		3.21	.104
Within Telephone		0.00	.96			.26	.62		1.37	.27		0.22	.65

94

results in a large coefficient of variation for the number of job openings. The distribution has a large spike at zero, and is skewed right. Again, this departure from normality for the underlying distribution is mitigated by the use of cell means for the analysis. Overall, the departures from the model appear to be mild enough to justify carrying out, with caution, basic F-tests via analysis of variance.

Results

Three of four states showed no significant differences due to collection method (Table 5). Utah approached significance with a P-value of 0.104 when the between personal visit and telephone component was separated out via orthogonal contrasts. Taking a look at means by collection method in Table 4, both Florida and Utah showed personal visit about 0.1 opening higher. In Massachusetts, telephone was higher, but by a smaller amount. Texas showed little difference. Calculations were performed by the ANOVA and GLM procedures of the 1979 version of SAS.

Table 4. Mean Job Openings
per Firm by Collection Method

	Average of T,T'	PV
Florida	.37	.47
Massachusetts	.37	.32
Texas	.20	.21
Utah	.27	.39

Was the test sensitive enough to detect any differences of real substance? Not really, in our view. A difference of about half the personal visit value was required to register as significant, given the error terms in the four states. This reiterated the volatility associated with estimating a rare characteristic.

As expected, mean number of job openings increased with size, with the size effect significant at the 5% level in three States. A general decline in level of openings over the three quarters, observed in the survey as a whole, was reflected in the reference period effects, significant at the 5% level in Florida and approaching significance in Massachusetts. No problems in telephone collection according to size emerged from this interaction term.

One of the principal findings of the ANOVA was the lack of a panel effect, except marginally in Utah. This gave support to using the quarter-by-quarter survey estimates of the previous section. It also offered a simplification for carrying out an analysis with categorical methods, since dropping panels would reduce the model to a complete design.

A natural counterpart to the survey estimates of per cent units with openings is categorical data analysis (Bishop, Fienberg, and Holland, 1975). Log-linear models were fitted to the data without the panel factor and with the response variable reduced to two categories (no openings, one or more openings). The resulting maximum likelihood estimates yielded an overall difference of 4.3% between telephone and personal visit collection for per cent of firms with openings compared to 5.4% from the survey estimates. Even at detailed levels, results looked very similar, and the categorical approach seemed to be more sensitive in testing for effects than ANOVA. The ANOVA approach has been retained, however, for it encompassed the full design and permitted examining panel differences. With additional time, efforts would be made to apply incomplete table methods to this Latin square design.

4. Conclusions

Our findings tend to support the feasibility of telephone collection for surveys involving small establishments. Overall, telephone collection appears to yield acceptable quality, since the differences from personal visit collection are not very large. In part, this favorable assessment depends on assigning some of the difference in two states to interviewer effects. In any case, the results are tempered by the large relative variability in the variable being measured, number of current job openings.

Survey managers found telephone collection to be timely and to produce high levels of response for a moderate expenditure of resources. The appearances that personal visit performed somewhat better and that interviewer performance influenced results suggest that telephone collection should be accompanied by periodic testing for quality.

FOOTNOTES

1/
Briefly, a current job opening is defined as an existing vacant job that is immediately available for filling and for which there is active recruiting extending outside the firm.

2/
The sample for the pilot survey was a probability sample of pairs of "like" units. The "shadow" unit of each pair was contacted in a subsequent quarter whenever the first unit dropped out due to nonresponse, refusal, or out of business status. The use of paired units in a controlled manner can be viewed as a nonresponse imputation technique designed to protect against nonresponse bias. Each figure used here is the ratio of the number of respondents, original or shadow, to the total number of in scope units solicited in the current quarter. To the extent that the pairing technique was not successful, this rate understates the potential for nonresponse bias. Any other set of response calculations considered were affected more strongly by response patterns in the nine months of the pilot prior to the Telephone Collection Test.

Appendix. Survey Procedures and Staffing Costs

The Job Openings Pilot Survey was conducted by four states under contract with the Bureau of Labor Statistics. The survey staffs were located in research/analysis offices within state employment security departments, where other Federal-state cooperative surveys are handled. Some information

was retained, mostly on their own initiative, with respect to staff utilization and cost.

Mostly, states either already had available or made arrangements for a set of phones with direct long-distance dialing capabilities at special rates (WATS lines or the local equivalent). One state offered callbacks, and kept open an 800 line for that purpose. While some delays were experienced in the middle of the day waiting for open lines, mostly the direct dial capability operated well. Response was strongest in the core hours of the day, say 9:30 a.m. - noon and 1-3 p.m., and in the core days of the week. Fridays and days adjacent to holiday weekends were especially weak. Both to utilize the better parts of the day and to avoid interviewer fatigue, three states (all except Utah) tended to limit interviewers to about half a day for calling. In these states, professional analysts were used for both personal visits and telephoning. Massachusetts' assignment of interviewer's to triplets, (cf. Section 2), so that the same individual conducted both telephone and personal visit collection, proved workable. With the exception of a very few cells, 90% or more of total responses over six weeks were reported as completed to BLS-Washington within 3½ weeks.

In Florida, typically an interviewer made about 20 calls in half a day, and obtained about 12 responses. Utah interviewers called all day long, except for Monday mornings and Friday afternoons. They reported 20-25 responses per day, which sounded comparable to Florida on an hourly basis. Massachusetts, which used telephone more extensively than the other states, reported 20 responses for about 25 cells over about 5 hours in a typical day, a somewhat higher ratio of responses to calls. Florida and Utah estimated telephone costs at $5 per response. Florida derived this from $3 per call, $2.50 for labor and $.50 for telephone charges. Management overhead was not included in these figures.

States averaged 4-5 personal visit collections per day, and 6-7 in major urban areas. Careful planning was required in each state, especially Texas, for scheduling visits to outlying areas. In scattered cases, Texas utilized local analysts. The cost figures were $30 and $40 per response for Utah and Florida, respectively. A substantially higher figure for Florida was understandable, since the major urban areas involved travel from Tallahassee. Florida's cost estimates included labor, per diem, and transportation costs.

Respondents were favorable to telephone collection. States put considerable effort into solicitation and into prenotification letters accompanying each collection. Over the phone, data requirements were easily explained; definitions were clarified; and, interviewers had the opportunity to suggest checking records.

Problems and limitations, not all endemic to telephone collection, were reported by each of the states. Telephoning seemed too easy at times. For instance, a respondent might immediately respond "no openings." When asked about new hires, on the other hand, there would be a pause, then some data based on payroll records. With no records in a majority of cases, snap answers sometimes contributed to an undercount of job openings. The same phenomenon occurred at times due to losing the sense that the survey was important, due to fatigue with responding to same survey or to realization that the survey was ending. A couple of states made the disquieting observation that small firms not in the telephone collection test, which received phone calls all three quarters, tended to exhibit lower levels of job openings. These factors served to emphasize the requirement for strong survey procedures to make telephone collection effective.

REFERENCES

Bishop, Yvonne, Fienberg, Stephen, and Holland, Paul (1975), Discrete Multivariate Analysis, MIT Press.

Plunkert, Lois, (1981), "Job Openings Pilot Program," Bureau of Labor Statistics report.

Scheffé, Henry (1959), Analysis of Variance, Wiley, Sec. 5.1.

Winer, B.J. (1971), Statistical Principles in Experimental Design, 2d ed., McGraw-Hill, Ch. 9.