

DATA COLLECTION METHODS IN FEDERAL STATISTICAL SURVEYS
 Owen Thornberry, Jr., National Center for Health Statistics, DHHS
 William L. Nicholls, II, Bureau of the Census, DC
 Stanley Kulpinski, Bureau of Labor Statistics, DOL

I. Introduction

This Subcommittee of the Federal Committee on Statistical Methodology was asked to assess the Role of Telephone, Mail, and Personal Interview Surveys in Federal Statistics (1). The consensus of the members was that the Subcommittee could make its most important contribution by concentrating its efforts on the telephone interview in Federal statistical surveys. Federal surveys are defined as those conducted by a Federal agency or under contract for a Federal agency.

The focus on telephone interview surveys was made primarily for three reasons. The first is the need to reduce the costs of Federal statistical surveys and the potential for such reductions by replacement of personal interviews with telephone interviews where this does not reduce the precision of estimates or increase biases. The second is the less fully developed state of methodological knowledge about telephone interviewing (especially its newer forms such as random digit dialing and CATI) than about personal interview surveys and mailed questionnaires. The third is the opportunity to draw upon, summarize, and disseminate recently initiated developmental and evaluative work on telephone interview systems in progress at a number of Federal agencies.

A first task of the Subcommittee was to assess the frequency with which telephone, mail, and personal interview data collection methods are currently employed in Federal surveys. This paper is devoted largely to reporting the results of this first analysis and presents data on current collection methods by sponsoring Federal department or agency, type of respondent, collection agent, frequency of data collection, and several measures of respondent reporting burden and cost. The paper concludes with a brief description of further analyses and subsequent tasks planned by the Subcommittee.

II. Source of Data

The best source of existing data on Federal statistical surveys is the Office of Management and Budget (OMB) Reports Management System (RMS). The RMS is a computer data file constructed from the OMB Standard Form 83 or SF-83. Every collection of identical information from 10 or more respondents by a Federal agency or under contract for a Federal agency requires submission of an SF-83 for OMB approval. If OMB approval is obtained, information from the SF-83 is entered in the Reports Management System and remains active until the OMB assigned expiration date is reached. The records examined for this analysis were those active on August 22, 1981.

The RMS includes records not only for statistical surveys but also for administrative

forms requiring OMB approval. The reported purpose of the collection was used to identify statistical surveys within this larger set. After examination of a sample of RMS records, it was decided that two of the six categories of purpose--"application for benefits" and "regulatory or compliance"--were out-of-scope for this analysis. Only those records with an indicated purpose of "program evaluation," "general purpose statistics," "program planning or management," or "research" were used (1). The 2,137 records meeting this criterion (40 percent of the total file) are called "survey records" in the text and tables which follow.

Table 1 provides a distribution of these records by data collection method as coded on the SF-83 (2). Two-thirds give the self-administered questionnaire (3) as the only method of data collection. The personal interview and the telephone interview are reported as the sole method of data collection for 10 percent and 2 percent, respectively. About one-fifth of the records are coded as involving multiple data collection methods. These include the telephone in combination with the self-administered questionnaire (4 percent), the personal interview (2 percent), and both (3 percent). A total of about 11 percent involve the telephone interview - either as the only method or in combination with the other methods.

Table 1. Reported Data Collection Method of Active OMB Approved Survey Records, 8/22/81

N	%	Data Collection Method
1470	68.8	Self-administered only
46	2.2	Telephone interview only
201	9.4	Personal interview only
90	4.2	Personal & self-admin
88	4.1	Telephone & self-admin
36	1.7	Telephone & personal
62	2.9	Personal & self-admin & tel
144	6.7	All other*
2137	100.0	Total

*Includes 111 records with data collection method not given.

III. Limitations of Data

Some observations are appropriate to interpret the findings presented above and in the section which follows.

(1) The specific contents of the SF-83 have changed several times in the last several years. The file used contained entries from three different forms. Consequently there are missing data on some variables (Federal cost for example) and undoubtedly inconsistencies on others.

(2) As noted previously, an attempt was made to exclude those records in the RMS which did not represent Federal statistical data collection. However, the information available was insufficient for the identification of all such records and an unknown (but apparently small) number of entries with administrative or other non-statistical data collection purposes was included. The effect of this is to increase the number of records with "self-administered" as a data collection method.

(3) The choice of telephone interview as one of the explicit data collection methods which could be checked was not available on the form until September 1980. Prior to that time the category "other" had to be checked and the words "telephone interview" written in. The effect of these omissions, if any, is to underrepresent the use of the telephone method.

(4) How extensively the telephone is used for data collection varies greatly. For example, the telephone may represent the only method of data collection as in random digit dialed (RDD) surveys or it may be used only to obtain data from those not responding to a mail questionnaire. Further, the telephone may be used for purposes other than actual data collection, such as to contact a respondent to arrange a personal interview, to verify or supplement information collected by mail or personal interview, or to encourage the respondent to send in a mail questionnaire. It was not possible in the analysis based on the SF-83's to determine how frequently such auxiliary uses of the telephone are included in reported use of the telephone as a data collection method.

(5) Many surveys listed in the file involve several stages of data collection; e.g., longitudinal, pretest and survey, core questionnaire and supplements, etc. Sometimes one survey record appears in the file covering all stages and other times a separate record is available for each stage. Thus, each entry does not necessarily represent a distinct data collection effort.

(6) There is considerable variation in the designs of the surveys on which this analysis is based and consequently generalizations as to a "typical" Federal statistical survey cannot be made. For example, number of respondents can be more than a hundred thousand or as little as 10; estimated cost can be in the millions or only a few thousand; the number of responses per respondent varies from one to 52; and so on.

The reader should keep these qualifications in mind when reviewing the data which follow.

IV. Findings

Cross tabulations of data collection method by selected variables are presented in Tables 2-6. Table 2 provides a distribution by collection method for selected sponsoring Departments or agencies. The Departments with the largest numbers of survey records are

Health and Human Services (326), Commerce (316), Agriculture (257), Energy (132), Defense (111), Labor (106), and Housing and Urban Development (105).

While there is considerable variation among Departments, the self-administered questionnaire is the most common approach for each. The Departments with the largest numbers of personal interview surveys are DHHS (48) and Agriculture (31). None of the Departments make extensive use of the telephone relative to other methods. The telephone in combination with the mail or personal interview is used in 33 percent of the surveys of the Department of Agriculture.

Table 3 provides data on type of respondent. One half of the entries involve data collection from business or industry and about one-fifth each from individuals (or households) and State or local government. As single methods of data collection, the personal interview and the telephone interview are used disproportionately with the individual or household respondent. The telephone in combination with the mail is most likely to be used in surveys of business or industry or of farms, whereas in combination with the personal interview the most common use is with individuals or households.

Data on collection agent--either the Federal Government or a contractor--is given in Table 4. Three fourths of the records list a Federal agency as the data collection agent. Those involving the self-administered approach, either as a single method or in combination with the personal interview or telephone, were disproportionately conducted by a Federal agency. By contrast, almost two-thirds each of the telephone and the telephone-personal interview combination were contracted out to a non-federal agency. This may reflect the fact that government agencies currently have limited telephone data collection capabilities.

Information from other sources suggests that in combination with the mail the most common use of the telephone is either to encourage nonrespondents to return mail questionnaires or to provide the information over the telephone. This use generally requires neither as large a staff nor as sophisticated a system as does the telephone alone or in combination with the personal interview. Thus, the high proportion of telephone-mail surveys which are conducted by Federal agencies is not surprising.

Table 5 provides data on frequency of data collection. About one half the surveys were infrequent (either single or occasional) and one-half periodic (weekly-biennial). The telephone interview, the personal interview and combinations involving the personal interview were most common with one-time surveys whereas the mail and the mail-telephone combination were used disproportionately in periodic data collection efforts. These data suggest that the telephone approach has not been built into scheduled ongoing data collection for most Federal agencies other than to follow-up mail nonrespondents.

Estimated median values for selected measures of sample size, respondent reporting burden, and cost are provided in Table 6.

These are approximate values and for each measure there is a wide range of values within each collection method.

Across all methods the median number of respondents and total responses (number of respondents X number responses per respondent) are around 550 and 1,400, respectively. The self-administered survey, the telephone survey, and the two in combination have the lowest median values on these measures. The telephone and the telephone-self-administered combination also have lower than average median values on the two respondent reporting burden measures--total hours reporting burden and minutes per response.

Estimated federal cost was available for only about one-third of the entries. Of these, the median values are \$70,000 total Federal cost and \$40 Federal cost per response. The least costly collection method on both measures is the telephone-self-administered combination, and the most costly is the telephone-personal combination. Relative to the single collection modes, the median total costs and per response costs for the telephone approach are about one-third higher than for the self-administered and around one-third lower than for the personal interview.

As noted, estimates of Federal cost were available for only one-third of the records. The relationships described above may be considerably different for the two-thirds of the records for which cost data were not available. In addition, the notes in Table 6 should be consulted relative to comparative interpretations of the data presented.

V. Summary

In summary, the purpose of this analysis of data from the OMB Reports Management System was to assess the extent of current uses of the telephone mode in federal statistical data collection. While there is some use of the telephone interview, that use is limited and is largely in combination with the other methods. The telephone is not used extensively as the primary mode of data collection by any Federal agency.

When employed as the sole collection method, it is most commonly used in surveys of individuals or households or of business or industry which are contracted out to private firms. These tend to be one-time surveys averaging less than one-half hour per response, with somewhat lower than average sample sizes and total hours reporting burden. Estimated Federal costs are somewhat higher than for mail surveys and lower than for those conducted by personal interview.

In combination with the mail self-administered questionnaire, the telephone is apparently used largely either to collect information from nonrespondents or to encourage the return of the questionnaire. Surveys of this type are disproportionately periodic surveys of business or industry or of farms which are conducted by the Federal agency rather than a private contractor. Median values for measures of respondent reporting burden and cost are lower than for the other collection methods.

The telephone in combination with the personal interview is most commonly used in one-time surveys of individuals or households which are conducted under contract. Relative to other methods, the personal interview-telephone surveys have disproportionately high median values on sample size, respondent reporting burden and federal costs.

Thus the Federal Government currently makes only limited use of the telephone interview. While some agencies have an in-house telephone data collection capability, none of them has a fully developed random digit dialing (RDD) or computer assisted telephone interview (CATI) capability. However, several agencies have made major commitments to the development of such systems (4).

VI. Future Work of Subcommittee

The interpretation of the presented tables has admittedly been restricted at many points by the limited survey design information required on the SF-83 forms. As a next step in the analysis, sponsoring Federal agencies have been asked to provide more detailed information about each survey utilizing the telephone as a primary or supplemental data collection method. These agencies also have been asked to report additional surveys which utilize telephone methods but were not so coded in the OMB Reports Management System. This enhanced data set will be utilized both for further statistical analysis and as a frame from which to select a range of surveys for detailed case studies. Special attention will be devoted to surveys recently substituting telephone interviews for personal interviews or mailed questionnaires. The focus of the case studies will be a more precise identification of methodological issues and a more meaningful analysis of cost. Federal agencies also have been asked to provide detailed descriptions of the developmental and evaluative work which they have undertaken or plan for the future.

The final report of the Subcommittee will include a summary of the research described above. In addition, the report will provide an overview of issues in telephone data collection and recommendations for the use of the telephone mode in future Federal statistical surveys.

Footnotes

(1) Since multiple purposes can be indicated, the excluded categories can appear in combination with the in-scope categories. Entries with both in-scope and out-of-scope purposes were retained and may include records of marginal relevance.

(2) Because it was not possible to tabulate directly from the OMB file, selected data were abstracted from the appropriate entries, coded, and a data file created. The tabulations in this report were generated from that file. The Subcommittee is indebted to the Computer Systems and Programming Branch, Division of Health Interview Statistics, NCHS, for the file creation and data processing.

(3) The category "self-administered questionnaire" includes 1,415 records coded as "mail self-administered" and 55 records coded as "other self-administered."

(4) For a description of research and developmental work at the Bureau of the Census,

the National Center for Health Statistics, and the Department of Agriculture, see: James T. Massey, Kent H. Marquis, and Robert D. Tortora: "Methodological Issues Related to Telephone Surveys by Federal Agencies," in Proceedings of the Social Statistics Section, American Statistical Association, 1982.

Table 2: Percentage Distribution of Reported Data Collection Method of Active OMB Approved Survey Records by Sponsoring Federal Department or Agency, August 22, 1981

Department or Agency	Data Collection Method												Total	N	
	Self-Adm Only		Tel Int Only		Pers Int Only		Tel Int Self-Adm		Pers Int Self-Adm		Tel Int Pers Int Self-Adm				All Other
Health & Human Services	64%	3	15	5	2	2	2	3	5	5	2	3	5	100%	326
Commerce	78%	1	5	1	4	2	2	2	6	1	4	2	6	100%	316
Agriculture	40%	1	12	7	18	2	2	13	7	7	2	13	7	100%	257
Energy	73%	4	11	4	2	-	-	1	5	4	2	1	5	100%	132
Defense	76%	4	10	-	-	-	-	-	10	-	-	-	10	100%	111
Labor	45%	1	18	7	2	5	7	7	16	7	2	7	16	100%	106
Housing & Urban Development	74%	2	12	1	1	1	1	-	9	1	1	-	9	100%	105
Education	86%	-	2	6	2	1	1	1	1	6	2	1	1	100%	84
Treasury	95%	-	2	2	-	2	2	-	-	2	2	-	-	100%	63
Interior	84%	-	7	4	-	-	-	-	5	4	-	-	5	100%	57
Transportation	64%	7	7	4	-	-	-	-	18	4	-	-	18	100%	56
Justice	68%	5	7	-	5	2	2	-	14	-	5	-	14	100%	44
Veterans Administration	92%	-	2	4	-	2	2	-	-	4	-	-	-	100%	52
Environmental Protection A.	53%	13	20	-	-	7	7	-	7	-	-	-	7	100%	45
Railroad Retirement Board	77%	2	9	9	2	-	-	2	-	9	2	2	-	100%	47
Tennessee Valley Authority	43%	-	11	7	-	-	-	4	36	7	-	4	36	100%	28
All Others	78%	1	5	6	5	1	1	1	4	6	5	1	4	100%	308
Total	69%	2	9	4	4	2	2	3	7	4	4	3	7	100%	2137
N	(1470)	(46)	(201)	(90)	(88)	(36)	(62)	(144)							

Table 3: Percentage Distribution of Respondent Type of Active OMB Approved Survey Records by Reported Data Collection Method, August 22, 1981

Respondent Type	Data Collection Method								Total	N
	Self-Adm Only	Tel Int Only	Pers Int Only	Pers Int Self-Adm	Tel Int Self-Adm	Tel Int Pers Int	Tel Int Pers Int Self-Adm	All Other		
Individual/Household	14%	46%	50%	39%	12%	67%	10%	21%	20%	(425)
State/Local Government	21	11	9	2	7	6	5	32	18	(395)
Farms	1	2	10	6	23	6	31	3	4	(84)
Business/Industry	57	35	22	46	49	14	32	38	50	(1064)
Multiple	7	6	9	8	9	8	23	6	8	(169)
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	
N	1470	46	201	90	88	36	62	144	2137	

Table 4: Percentage Distribution of Collection Agent of Active OMB Approved Survey Records by Reported Data Collection Method, August 22, 1981

Collection Agent	Data Collection Method								Total	N
	Self-Adm Only	Tel Int Only	Pers Int Only	Pers Int Self-Adm	Tel Int Self Adm	Tel Int Pers Int	Tel Int Pers Int Self-Adm	All Other		
Federal Government	81%	33%	53%	67%	86%	39%	73%	73%	75%	(1610)
Private Contractor	19	67	47	33	14	61	27	27	25	(527)
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	
N	1470	46	201	90	88	36	62	144	2137	

Table 5: Percentage Distribution of Frequency of Data Collection of Active OMB Approved Survey Records by Reported Data Collection Method, August 22, 1981

Frequency	Data Collection Method								Total	N
	Self-Adm Only	Tel Int Only	Pers Int Only	Pers Int Self-Adm	Tel Int Self-Adm	Tel Int Pers Int	Tel Int Pers Int Self-Adm	All Other		
Single or occasional	38%	66%	63%	65%	34%	72%	45%	46%	44%	(938)
Weekly-semiannual	27	17	9	12	25	8	23	26	24	(514)
Annual or biennial	31	6	18	16	26	11	13	22	27	(582)
Other	2	11	10	7	15	8	19	6	5	(103)
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	
N	1470	46	201	90	88	36	62	144	2137	

Table 6: Estimated Median Values for Selected Measures of Sample Size, Respondent Reporting Burden and Cost of Active OMB Approved Survey Records by Reported Data Collection Method, August 22, 1981

Selected Measures	Data Collection Method								Total
	Self-Adm Only	Tel Int Only	Pers Int Only	Pers Int Self-Adm	Tel Int Self-Adm	Tel Int Pers Int	Tel Int Pers Int Self-Adm	All Other	
Number of respondents	400	1000	1500	3500	800	4300	2600	800	550
Number of responses	1100	1300	2000	4700	1600	4700	7600	2000	1400
Hours reporting burden	1200	500	1000	3300	600	1800	1800	1800	1200
Minutes per response	45	20	30	20	15	30	15	35	30
Total number records	1470	46	201	90	88	36	62	144	2137
Federal cost (\$000) *	\$ 55	\$ 90	\$ 140	\$ 125	\$ 30	\$ 245	\$ 175	\$ 50	\$ 70
Federal cost per response *	\$ 40	\$ 60	\$ 110	\$ 40	\$ 15	\$ 130	\$ 20	\$ 60	\$ 40
Number with cost estimate	378	25	82	43	61	28	49	19	685

*Caution should be exercised in any comparative interpretations of the estimates of Federal cost. For example:

- (1) These figures were derived from estimates provided by the agencies requesting OMB clearance. There is no reason to assume consistency among agencies in the procedures followed nor in the components of cost included in deriving these estimates. Further, data were available only for one third of the entries.
- (2) These figures do not represent pure estimates of cost by method. Data collection method is only one component of cost in any survey design. The universe sampled or other aspects of the design may dictate the collection method(s). For example, surveys of rare populations may require screening by telephone prior to a personal interview; i.e., TI/PI approach. In such surveys, the cost for the TI/PI combination would be expected to exceed average PI costs. That is, the higher median cost estimate for TI/PI as contrasted to PI does not indicate that a particular survey would cost more with the former rather than the latter collection method. Where a dual-mode approach is possible (TI/PI), it may be considerably less expensive than PI alone.
- (3) While all figures are probably underestimates of actual costs, this error is much greater for those conducted in-house (by the federal agency) than under contract with a private organization. Cost estimates for the latter group are based largely on negotiated contract costs. Estimates for the in-house surveys are not comparable because they generally do not include all related components of actual cost. Thus, relative differences between TI, PI, or TI/PI (disproportionately contract) and SA, PI/SA, or TI/SA (disproportionately in-house) are overstated. For example, while with identical survey designs the telephone approach (TI) may be more costly than data collection by mail (SA), the magnitude of the difference as shown above is greatly exaggerated.
- (4) Periodic surveys are generally less costly than one-time surveys. The SA and TI/SA as contrasted to other methods are disproportionately periodic surveys. Thus, the magnitude of differences between these and the other methods is overstated.