

USE OF CATI IN A BUSINESS SURVEY OF HEALTH INSURANCE PLANS

K.E. Boyle, F.J. Potter, S.A. Garfinkel, J.J. McNeill, RTI
P. McDonnell, HCFA
K.E. Boyle, RTI, P.O. Box 12194, RTP, NC 27709-2194

KEY WORDS: Enterprise Surveys

1. Introduction

Under contract to the Health Care Financing Administration (HCFA), the Research Triangle Institute (RTI) has designed and conducted a computer-assisted telephone interview (CATI) survey of businesses and other organizations to estimate the number and type of private health insurance (HI) plans in the United States. Private HI can be obtained for a person and his/her dependents through a variety of sources. These sources include:

1. a person's place of employment (through employer-sponsored or union-sponsored plans including Taft-Hartley Trusts);
2. membership associations or organizations (such as professional or business associations or religious organizations);
3. a student's post-secondary school;
4. direct contracts with insurance firms; and
5. direct contracts or arrangements with health maintenance organizations (HMOs).

The Survey of Health Insurance Plans (SHIP) is designed to provide coverage of the universe of private HI available to the civilian U.S. population by accounting for this variation in the location of the HI information (Garfinkel et al. 1990). The SHIP collects information on private HI by selecting a sample and gathering data from each of these sources.

The sample design incorporates two primary survey components: employer and non-employer surveys. The two survey components are differentiated by the primary source of private HI information not by whether or not the insurance is based on employment. Both the employer and non-employer surveys include employment-based insurance.

We describe briefly in this paper both survey components, but our major focus is one component of the employer surveys, i.e., the survey of private employers. We concentrate on the sample design for the private employer survey including the definition of survey units, the sampling frame and stratification variables, the distribution of the sampling frame by strata, sample selection, and sample size determination. The final section of our paper describes the response rates and other methodological results for the private employer survey.

1.1 Non-Employer Surveys

The non-employer surveys obtain information on the following:

1. union-sponsored group insurance;
2. association-sponsored group insurance from membership associations and organizations;

3. group insurance for students from post-secondary schools;
4. group insurance for civilian federal government employees from the non-HMO carriers in the Federal Employees Health Benefit Program (FEHBP) and independent federal agencies that provide non-FEHBP insurance plans;
5. health care provided by HMOs (including care provided to civilian/federal employees);
6. non-group HI provided by insurance companies; and
7. non-group HI provided by state risk pools.

We have used the Dun's Market Identifiers (DMI) data base as the source of the sampling frame for unions, membership associations and organizations, and post-secondary schools. (Details of this data base are provided in the following section.) However, for the other private insurance sources, we have used specialized data bases or composites of data bases; see Exhibit 1.

1.2 Employer Surveys

The employer surveys consist of two types of employers: a) state and local governments and b) private business enterprises. The private business enterprises include all commercial (for-profit) business enterprises and not-for-profit enterprises. Not-for-profit enterprises include unions, membership associations and organizations, religious organizations, and similar entities with employees.

The sampling frame for the private employer survey was the DMI data base from Dun's Marketing Services (DMS), a subsidiary of Dun & Bradstreet (D&B) Corporation. The DMI data base is developed from the D&B credit report data base. The DMI data base, reported to be the most comprehensive data base of business entities available commercially, contains information on approximately 6.7 million enterprises representing about 7.7 million establishments.

According to D&B, an enterprise is a corporate entity that has ownership or majority control of one or more establishments. An establishment represents any business with a unique, separate and distinct operation, including organizational units within a corporation, association, or organization. Each establishment in the DMI data base has a unique Data Universal Numbering System (DUNS) number, and establishments within an enterprise can be linked by the DUNS number to permit the identification of corporate families. Use of the DMI data base permits exploitation of the corporate structure linkage.

For the survey of State and local governments, we used the 1987 Census of Governments (COG) data base compiled by the Bureau of the Census. This data base contains information on approximately 84,000 governmental

units including state, county, municipal and township governmental units; special governmental districts such as water and sewer authorities, fire districts and port authorities; and school districts. The COG is one of the most comprehensive listing of governmental units available.

2. Sample Design for the Private Employer-based Survey Component

The following sections describe the details of the sample design for the survey of private business enterprises.

2.1 Definitions of Survey Units

With the steady increase in the cost of health care and private HI over the past years, businesses and other organizations have become very innovative in designing HI plans and programs. Our preliminary investigations revealed that corporations and organizations vary substantially in how HI is provided to employees or members and how HI information is maintained. Unlike employment information (usually available from the establishment unit) or corporate financial data (usually available from corporate headquarters), HI information may be maintained anywhere in a corporate structure. A corporation may have all HI information located at corporate headquarters, at regional or subsidiary levels, at the establishment level, or at multiple levels.

The SHIP was designed to account for this variation in the location of HI information. We have defined the following units for our survey design.

1. The sampling unit (SU) is the organizational unit that is subject to initial sample selection; the sampling frame for a specific population is the list of all SUs.
2. The reporting unit (RU) is the organizational unit that is the source of the HI information for one or more sampling units. An SU may be the RU, or an SU may identify one or more RUs that are the source of the HI information. In addition, two or more SUs may identify the same RU.
3. The plan unit represents the HI plan. An RU may indicate that employees or members have one or more optional HI plans. Each HI plan is a plan unit.

In a sampling sense, the survey design allows for multiple stages of sampling, depending on how the HI information is maintained. We have attempted to minimize the amount of subsampling required; that is, sampling within SUs.

2.2 Sampling Frame and Stratification Variables

As indicated previously, the sampling frame for the private employer-based survey is the DMI data base which permits use of the Standard Industrial Classification (SIC), corporate structure, size (total number of employees), and geographic location for stratification. Another benefit of the DMI data base is that corporate linkages have been established by DMS. We have used these linkages in the sample design to define the sampling frame, or list, of enterprises; each enterprise is a sampling unit.

An important aspect of this survey design is the corporate structure information contained in the DMI data base. DMS categorizes entities either as a single DUNS number establishment or a multiple DUNS number entity. A single DUNS number establishment (called a single location organization by DMS) is, by default, a single DUNS number enterprise. A multiple DUNS number entity is an enterprise with more than one DUNS number, and thus, may contain subsidiaries, divisions, and headquarter entities as well as one or more branches. In a multiple DUNS number entity, one unit is defined as the utmost corporate headquarters (called an ultimate parent by DMS). All subsidiaries, divisions, headquarter entities and branches in a corporation are linked to this ultimate parent.

For sampling in this survey, we have defined a stratum of single DUNS number enterprises (or single location enterprises) and two strata of multiple DUNS number enterprises. The strata of multiple DUNS number enterprises include (1) a stratum of two-level enterprises (a headquarters DUNS unit and associated branch DUNS units), and (2) a stratum of multiple-level enterprises.

Our rationale for the definition of these three strata is as follows. We expected single location enterprises and two-level enterprises would be similar with respect to the location of HI data. That is, the most knowledgeable person for the HI data would be located at the headquarters of a two-level enterprise. On the other hand, in the multiple-level enterprises, the most knowledgeable person(s) for the HI data would be located somewhere within the corporate structure, not necessarily at the largest or highest unit (the ultimate parent or enterprise) or the smallest unit (the establishment).

For the multiple-level enterprises, we anticipated that the ultimate parent was more likely to have the data than the smallest (establishment) unit. Thus, we planned to sample ultimate parents. Nevertheless, it was feasible that many cases would send us to several RUs within the enterprise for the HI data. The enterprise-level sampling strategy required us to collect data for all or a sample of RUs within an enterprise. We have subsampled RUs within enterprises if the number of RUs exceed predetermined criteria and if the SU in question was expected to consume a disproportionate share of the survey's resources.

Because of the potential for subsampling, the sample design for the multiple-level enterprises is a stratified multi-stage design where the number of stages varies for each SU. Although the same strategy is available for the single- and two-level enterprises, it has been rarely used. Thus, the sample design required for these strata can be considered as a relatively straightforward stratified random sample design.

A second reason for defining the three strata of enterprises was the availability and accuracy of additional data for stratification. All enterprises were further stratified by SIC category, number of employees (size), and geographic region. For the two- and multiple-level

enterprises, the units were stratified by SIC, geographic location of the ultimate parent and the size of the enterprise.

Stratification of multiple-level units by region may have limited usefulness for estimation purposes because the enterprise may span more than one region, and the highest level unit in an enterprise may not be located near the majority of its employees. The corporate headquarters (the ultimate parent unit) may be located in a financial center as opposed to near the majority of its business operations. Similarly, stratification by SIC may also have limited usefulness for multiple-level enterprises, i.e., the SIC of the highest level unit may not be an accurate indication of the industry of the majority of its employees.

For two-level enterprises, the SIC of the highest level unit may represent the industry of the majority of its employees. Furthermore, the size information is expected to reflect the total number of employees in the two-level enterprise. However, the location of the ultimate parent may not correspond to the location of the majority of the employees.

In summary, DMI data on corporate structure, SIC, size, and region have been used for classifying enterprises into strata. For the single-level enterprises (comprising the largest percentage of enterprise units), DMI data on SIC, size, and region accurately represent the classification of enterprises. For the two-level enterprises, the data on SIC and size are also expected to be accurately classify the enterprises, but the region in which the ultimate parent of the enterprise is located may not properly reflect the region for all employees. For the multiple-level enterprises, only the data on size is expected to provide a correct classification for the entire enterprise. The SIC code and region for the ultimate parent of the enterprise provides only an indication of the major industry and location of the enterprise and its employees.

The sampling strata have been defined as follows:

- Corporate Structure:
 - single location, two-level, and multiple-level enterprises
- SIC Classification (9 SIC divisions):
 - SIC Major Groups
 - 1. 01-09 Agriculture, Forestry and Fishing
 - 2. 10-14 Mining
 - 3. 15-17 Construction
 - 4. 20-39 Manufacturing
 - 5. 40-49 Utilities, Transportation and Communication
 - 6. 50-51 Wholesale Trade
 - 7. 52-59 Finance, Insurance, and Real Estate
 - 8. 60-67 Services
 - 9. 70-89 Other
- Size (number of employees):
 - 1. 0 or unknown
 - 2. 1-9
 - 3. 10-19

- 4. 20-49
- 5. 50-99
- 6. 100-199
- 7. 200-999
- 8. 1,000-4,999
- 9. 5,000-24,999
- 10. 25,000 or more

- Geographic Region (4 Bureau of the Census's Regions)
 - Northeast, Midwest, South, and West.

2.3 Distribution of the Sampling Frame by Strata

Of the 6.7 million enterprises on the DMI private employer sampling frame, 97.1, 2.2, and 0.7 percent were single location, two-level, and multiple-level enterprises, respectively. Furthermore, the marginal size distribution varied by corporate structure: as expected, the single location enterprises had relatively more smaller-sized enterprises than enterprises with more complex corporate structures. (The distribution of this frame by strata is not provided because of space limitations.)

2.4 Selection of SUs and RUs

An SU may be able to report for itself or indicate that other entities have the required HI data (i.e., are the RUs). A description of the selection of SUs in the private employer-based survey is followed by that for RUs.

2.4.1 Selection of SUs

The first-stage sample included all enterprises with 25,000 or more employees (303 enterprises) with certainty. With one other exception, the final sample was selected with probabilities proportional to size (PPS) within each combination of corporate structure and size category where size is the total number of employees. The exception to this rule occurred in the smaller size categories where we selected with equal probability when the size category was "0 or unknown" or "1-9" employees. We selected with PPS where the size measure was the number of employees, because the count of employees was expected to be highly related to the number of persons covered by HI and to health plan costs and benefits. For sample selection, we used a probability minimum replacement (PMR) sequential selection procedure (Chromy 1979).

DMS provided RTI with the frame of abstract records for all enterprises. These records were "abstract" records in the sense that no name or address information for these enterprises were provided. However, each enterprise abstract record included data needed to construct the strata (corporate structure, number of employees, SIC code, and region) as well as the DUNS number. RTI then sampled the appropriate number of cases from each stratum and sent a file of these sampled abstract records to DMS. (The specific sample sizes are discussed below and were estimated based on a cost and variance optimal allocation of sampling units.) DMS matched these records by DUNS number to records on their complete file and returned the augmented

records to RTI. These data provided the necessary input to RTI's CATI system.

The study sample was also randomly partitioned to develop subsamples or waves. Each sample wave was a valid probability sample and, therefore, a sample of the waves is also a valid sample. The sampled units were then contacted via RTI's CATI system to provide the required HI cost and benefit data.

2.4.2 Identification and Selection of RUs

The RUs for an enterprise (i.e., SU) are the sources of information for its HI data. The number and characteristics of the RUs will vary for different enterprises. An RU may be a subsidiary or an establishment of the enterprise or may be an insurer, an insurance agent, or a third party administrator. An enterprise, especially a large enterprise, often sponsors multiple health plans:

- one or more traditional plans,
- one or more single service plans, and
- one or more HMOs.

Our objective was to collect data on each HI plan that is offered, except for HMOs. We asked for only aggregate HMO data for the enterprises. In this sampling and data collection structure, the HI plan is the analysis unit.

To control data collection costs and prevent a few SUs from consuming a disproportionate share of resources, the number of RUs per SU contacted via CATI was limited. A sampling of RUs was used if the number of RUs exceeded a predetermined number.

During the conduct of a previous HI survey, we found only a few enterprises directed the interviewers to other RUs for data on traditional and single service health plans. Additionally, we expected that most enterprises will have, on average, 3 or fewer reporting units. However, because some enterprises will have more than 3 reporting units, the sample size of RUs sampled within each enterprise has been set via a sliding scale.

The RUs within an enterprise were selected with probability proportional to the number of employees; therefore, the unconditional probability of selecting the RU is proportional to the number of employees accounted for by the reporting unit. The benefit of using PPS sampling of enterprises and RUs is that, if the number of persons covered by HI is directly proportional or equal to the number of employees, then the variance is reduced and the sampling precision increased. Substantial precision improvement can be achieved even if the number of persons covered and the number of employees are highly positively correlated.

2.5 Sample Size Determination

For sample size estimation, cost and variance models were developed that incorporated stratification by corporate structure and employer size and the sampling of enterprises, reporting units, and health insurance plans. Our cost-variance optimization attempted to minimize the cost function subject to the constraints of the relative variance for specific

estimates and used an algorithm based on the Kuhn-Tucker theory (Chromy 1987).

The distribution of the sample of 3,956 private enterprises by corporate structure and size that was suggested by the cost-variance optimization appears in Exhibit 2. The response rates by corporate structure and size category follows.

3. Methodological Findings

We define an enterprise as a respondent if we determined whether or not the enterprise had eligible HI plan data. Given this definition, the overall response rate for the private business enterprise sample was 91.5 percent. The response rate was highest for single location enterprises (95.1 percent). The response rate then decreased as the complexity of the corporate structure increased: 92.2 percent for two-level enterprises and 86.5 percent for multiple-level enterprises. There was no pattern in the response rate by size category across corporate structure.

We also examined the phenomena of the aging of the DMI frame. From a practical standpoint, several months elapsed between "freezing" of the frame for sample selection and releasing the sample to RTI's CATI system. Exhibit 4 presents the distribution of closed business enterprises by corporate structure and size. Of the 3,956 enterprises in the sample, we determined that 137 enterprises closed. Not unexpectedly, the majority of closings occurred among small-sized, single location enterprises, supporting the belief that small enterprises are a risky business.

4. Summary

RTI has designed and conducted a CATI survey of private business enterprises as part of a larger effort to estimate the number and type of HI plans in the United States. For this survey component, we selected a sample of approximately 4,000 enterprises (i.e., SUs) from our sampling frame, i.e., the DMI data base, stratifying on corporate structure, number of employees, and SIC category. We used RTI's CATI system to locate the organizational units within each enterprise that was the source of HI information (i.e., RUs) and to collect HI information from these RUs. To control costs, we sampled RUs within an SU when the number of RUs was large. Our survey design and protocol achieved an overall response rate for the private business sample of 91.5 percent.

References

- Chromy J.R. (1979) "Sequential Sampling Methods," American Statistical Association Proceedings of the Section on Survey Research Methods, pp 401-406.
- Chromy J.R. (1987) "Design Optimization with Multiple Objectives," American Statistical Association Proceedings of the Section on Survey Research Methods, pp 194-199.
- Garfinkel S.A., McNeill J.J., and Potter F.J. (1990) "Survey of Health Insurance Plans: Project Plan Report," Prepared for HCFA under Contract No. 500-88-0065.

Exhibit 1
Description of Frames for Non-Employer Surveys

Survey Component	Frame
Unions	Dun's Market Identifiers (DMI)
Associations	DMI
Post-Secondary Schools	DMI
FEHBP Carriers	List of 22 carriers from the Office of Program Management
HMOs	Frame developed from Group Health Association of America, InterStudy, and the Office of Prepaid Health Care files
Insurers:	
Commercial	List of 1,341 companies from A.M. Best Co.
Blue Cross/Blue Shield (BCBS)	List of 73 insurers from the BCBS Association
State risk pools	List of 14 State risk pools
Third Party Administrators (TPAs)	List of 357 TPAs in the Society for Professional Benefits Administrators (SPBA)

Exhibit 2
Distribution of the Private Business Enterprise Sample
by Corporate Structure and Number of Employees

No. of Employees	Corporate Structure			
	Total	SL ¹	TL ²	ML ³
Total	3,956	1,936	906	1,114
0 or Unk.	280	180	60	40
1-	455	350	65	40
10-	470	350	70	50
20-	900	530	240	130
100-	400	200	110	90
200-	425	190	115	120
1,000-	410	110	120	180
5,000-	313	26	106	181
25,000+	303	0	20	283

1 SL: Single location enterprises
2 TL: Two-level enterprises
3 ML: Multiple-level enterprises

Exhibit 3
Response Rate (Percent) Among the Private Business Enterprise Sample
by Corporate Structure and Number of Employees

No. of Employees	Corporate Structure			
	Total	SL	TL	ML
Total	91.5	95.1	92.2	86.3
0 or Unk.	92.0	95.0	90.0	82.8
1-	95.1	96.7	91.5	88.2
10-	96.0	96.2	98.5	91.1
20-	92.0	94.3	91.6	83.6
100-	91.4	95.3	88.7	86.0
200-	91.2	93.9	88.2	89.7
1,000-	90.5	94.2	92.4	87.0
5,000-	88.6	91.3	97.1	83.2
25,000+	87.7	--	100.0	86.9

Exhibit 4
Distribution of Closed Business Enterprises in the Sample by
Corporate Structure and Number of Employees

No. of Employees	Corporate Structure			
	Total	SL	TL	ML
Total	137	105	17	16
0 or Unk.	39	33	4	2
1-	46	38	4	4
10-	23	18	3	2
20-	13	7	4	2
100-	9	5	1	3
200-	6	3	1	2
1,000-	0	0	0	0
5,000-	0	0	0	0
25,000+	1	--	0	1