

SAMPLING AND REPORTING UNIT ISSUES IN EMPLOYER-BASED CATI HEALTH CARE SURVEYS

Frank J. Potter, Kerrie E. Boyle, Melodie W. Rush, Research Triangle Institute
Frank J. Potter, RTI, Box 12194, RTP, NC 27709

Keywords: Establishment surveys, Business surveys

1. Introduction

As stated in the report "Quality in Establishment Surveys" (OMB 1988), data from these surveys "comprise an integral and important part of the nation's information base for policy-making and analysis."¹ Establishment surveys (ES), a special class of statistical surveys with similarities to population or housing surveys, have unique features which increase their complexity. The issues in ES have three primary sources: the complexity of the population of employers and employers themselves; the information being sought; and the inferential population for the survey. The population of employers and worksites is very dynamic and is continually evolving through the births of new employers, deaths and mergers of employers, and internal reorganizations. Employers are also very concerned about the confidentiality of information and fear unauthorized release of information to competitors. Recently, an increasing number of employers have imposed corporate policies against voluntary surveys, regardless of the survey's purpose or promises of confidentiality.

Within a large multi-location employer, data may be maintained at various units or levels. Cost-related data (for example on health insurance or health promotion activities) may be maintained either centrally at one location or in a decentralized fashion at subsidiary levels. On the other hand, employee characteristics (e.g., the number or percentage of employees by age, gender or race/ethnicity) are generally available at the individual worksite.

The information being sought is a second issue for ESs. Health-related costs and health care utilization, being generally detailed quantitative data, may require a different respondent in a multi-location employer than information on attitudes and opinions on the services or plans for changes to health care plans or programs. The information being sought is directly related to the issue of the inferential population for establishment surveys.

The inferential population for an ES may be worksites, entire corporations or firms, or individual health insurance or health promotion plans. For a portion of the population of establishments, a one-to-one

relationship may exist among these three entities. Vigilance is needed to ensure that the respondent (as well as the data collector) is discussing the data appropriate for the unit of inference. Confusion may exist on the unit of inference because of the complexity of employers and the inconsistent usage of terminology in establishment surveys.

Although some establishment surveys predate household surveys, the literature is less well developed than that for household surveys. The lack of information, in part, resulted in the 1993 International Conference on Establishment Surveys (ICES), which brought together many researchers involved in establishment surveys. In addition, the President's health care reform initiatives also led the Office of Technology (OTA) to sponsor a review of methods for an employer-based health benefits survey. Although OTA never issued a final report of this review, a draft report was widely distributed for review. More recently, the monograph, *Business Survey Methods*², was published with expanded versions of the ICES conference papers and other articles on ES topics. This monograph contains a wealth of information on definitions, issues, and current methods in ES. These recent professional activities indicate the level of concern over survey research methods specific to establishment surveys.

1.1 Establishment Surveys

Sampling and reporting unit issues in establishment surveys also reflect the complexity of employers, the information being sought in the survey, and the population to which inferences will be made. Establishment surveys is the general name given to all employer surveys in the U.S.³, but this name is somewhat misleading. Technically, an establishment is defined as "an economic unit which produces goods or services" at a single physical location and is primarily engaged in a single activity.⁴ Although rare, an establishment can have more than one physical location, and a physical location can have more than one establishment. An enterprise, on the other hand, is defined as the aggregation of all establishments (e.g., branches, offices, plants, and subsidiary companies) under common ownership and control.⁵ In a large national enterprise, both the corporate headquarters and the smallest regional office are establishments. An analysis of certain industries by RTI researchers illustrates a difference between establishments and enterprises with respect to the workforce; enterprises with 1,000 or more employees

account for 43 percent of the U.S. workforce for these industries, whereas establishments with 1,000 or more employees account for approximately 11 percent.⁶

RTI has conducted establishment surveys for the Federal Reserve System, the Environmental Protection Agency, the Health Care Financing Administration (HCFA), National Institute on Drug Abuse (NIDA), other agencies in federal and state governments, and commercial clients. In 1993, we conducted a national survey of 6,488 establishments to determine the prevalence and costs of employee assistance programs (EAPs) for NIDA and obtained a 90 percent response rate.⁷ We are currently implementing the 1995 cycle of this survey. In 1988, RTI conducted the National Survey of Small Business Finances (NSSBF) for the Federal Reserve System.⁸ In this survey, RTI collected data about the financial accounts, balance sheet, income statement, and other characteristics of the business from a national sample of 3,600 small business firms (enterprises). RTI also conducted national surveys on the characteristics and costs of employer-based health insurance for HCFA (Surveys of Health Insurance Plans, SHIP) and for a private health care consulting firm.

The purpose of this paper is to discuss the issues of the sampling unit and the reporting unit in the context of two national surveys conducted by RTI in 1993. The first survey used the **establishment** as the sampling unit and the second survey used the **enterprise**. The first survey collected information about the general employee benefits, the employee assistance program (if one existed), and the characteristics of and the workforce at the establishment. The second survey also collected information about the employee benefits, but focused more directly on the health insurance plans sponsored by the enterprise. In the second survey, we also obtained data on the characteristics of the workforce at the enterprise.

2. Description of Surveys

2.1 National Survey of Worksites and EAPs (NSWEAP)

The 1993 NSWEAP was designed to update the information collected in the 1985 National Survey of Worksite Health Promotion Activities, conducted by RTI, and the 1988 Survey of Employer Anti-drug Programs, conducted by BLS. The specific objectives of the 1993 NSWEAP were to:

- o estimate the prevalence of employee assistance programs (EAPs) in small-, medium-, and large-sized worksites;

- o determine the characteristics of existing EAPs in terms of type, location, and sponsorship;
- o determine if the prevalence and characteristics of EAPs vary by industry, worksite size, geographic region, and workforce characteristics;
- o estimate the annual costs of EAPs by type;
- o determine prevalence and characteristics of drug and alcohol testing activities, health promotion activities, and health benefits.

The target population of this research consists of all worksites with 50 or more employees of private business enterprises in the U.S. (excluding agricultural enterprises). The sampling frame was constructed using the Dun's Market Identifiers (DMI) data base from Dun's Marketing Services (DMS), a subsidiary of Dun & Bradstreet (D&B) Corporation. We assumed a one-to-one relationship between an establishment and a worksite for this survey.

The sampling strata were defined on the primary industry at the worksite (six categories) and the number of employees at the worksite (5 categories). Because of the expected difference between the D&B count of worksite employees and the reported count of employees, we included worksites with 40 or more employees in the sampling frame. However, only worksites that reported 50 or more employees were eligible for the analysis and reporting. The sampling frame included approximately 421,000 establishments with 40 or more employees. Geographic location (four Census Regions) was used as a secondary stratification factor within the sample selection procedure. The sample was allocated proportionally within each sampling strata across four geographic location strata and selected with equal probability within each stratum.

The final stratified sample contained 6,488 worksites and 5,828 (90%) worksites responded (see Table 1). These respondents included 3,204 eligible worksites and 2,624 ineligible worksites (i.e., government worksites, worksites with fewer than 50 employees and closed worksites). The response rate ranged from 80 to 96 percent across the 30 sampling strata. All survey data were collected using computer-assisted telephone interviewing (CATI).

2.2 National Survey of Employer-Sponsored Health Insurance

This national survey of enterprises was designed to collect information about the total number of

employees covered by health insurance and the costs for this health insurance by the general plan types (i.e., traditional, PPO, POS, and HMO). We also collected detailed data on the single largest plan of each type and collected data about other employee and retiree health benefits. The survey was a mail survey with limited telephone follow-up of nonrespondents using CATI.

We used a stratified simple random sample design of private employers and governments with some employers included into the sample with certainty. The certainty cases included a sample of respondents to a 1992 survey (conducted by the client), employers with 40,000 or more employees, state governments, the most populated cities and county governments, and the largest school districts. The sampling design was based on the information on the 1992 respondents and information from Dun and Bradstreet. For purposes of this paper, we will limit the discussion to the sample design for the private employers.

For the private employers, we used the DMI data base of enterprises and enterprises with less than 10 employees were excluded from the sampling frame. The sampling frame contained approximately one million employers. We stratified enterprises by size of the enterprise and oversampled a few selected industries. We controlled the selection of the enterprises in each size stratum to achieve an approximate proportional allocation across industry. From a sample of 1,897 enterprises, we obtained responses from 1,245 (66 percent) and health insurance and costs data from 880 enterprises (see Table 2).

3. Sampling and Reporting Unit Issues

3.1 Differentiation of Units

In the report "Quality in Establishment Surveys," the authors define the reporting unit as "the unit for which the data are to be collected."¹ We have found that in some establishment surveys, a more complex structure is needed. In this structure, units are categorized as either a sampling, reporting, or analysis unit.

3.1.1 Sampling Unit

The sampling unit (SU) is the unit to which the sample selection probabilities, or frequencies, are assigned, and the sampling frame is a listing of the sampling units. The four primary issues for SUs are: the completeness of the sampling frame; births, deaths and mergers; the sources of information (i.e., the reporting unit); and the authority to contact the reporting unit. The sampling frame is assumed to be a comprehensive listing of all possible sampling units. In almost every list-based survey, the list is not comprehensive of all potential

sampling units. ESs tend to be a classic example of incomplete sampling frames because business "births" and "deaths" can occur so quickly and the identification of "births" can require substantial sleuthing. The birth of establishments (as opposed to an enterprise) is very difficult to detect because the definition of an establishment is vague. An enterprise can form a new establishment essentially overnight and the identification of a new establishment within a complex enterprise may not be apparent to people external to the enterprise.

For an enterprise-level survey, the list of enterprises is also subject to a dynamic process. A new enterprise can be formed either by creating a new organization, by splitting of an existing enterprise, or merging of multiple existing enterprises. In the latter two situations, the sampling units can be linked to the new enterprise, and unbiased estimates can still be computed. The situations can be classified as a one-to-multiple and a multiple-to-one multiplicity, respectively. The effect of this dynamic process can be accommodated if adequate information is collected from the units selected in the sample.

The SU may be a source of only part of the desired data, and the respondent at the SU may identify additional sources at other locations (reporting units) for data. This new contact often requires a complete re-introduction of the survey. Prior authorization of RU contact will likely expedite the contact and increase cooperation from these "spawned" units. The initial respondents for the EAP survey (an establishment survey) frequently identified a person at a higher level in a corporation (e.g., a regional or corporate headquarters) as a source of the costs of EAP services. Telephone interviewers then made repeated attempts to secure the cost information.

3.1.2 Reporting Unit

We define the reporting unit as the unit from which data are obtained⁹. The reporting unit may be the same as the sampling unit, or the reporting unit may be completely external to the establishment or the enterprise. A single sampling unit (either an establishment or enterprise) may have multiple reporting units, or a reporting unit may be associated with more than one sampling unit. Both of these situations can impact the data collection and estimation procedures.

The primary issues related to the RUs are: maintaining the identity of the unit(s) for which data are being sought; identification of RUs external to the enterprise; minimization of respondent burden for RUs identified by multiple SUs; and collection of data specific to the unit of analysis. When an RU is contacted, the

telephone interviewer needs to clearly identify the unit for which the data are being sought. A corporate office may have little direct knowledge about a specific establishment. For example, external EAPs (EAP services contracted for externally by the employer) are servicing an increasing portion of worksites (relative to internal EAPs that are staffed by company employees). An EAP coordinator at a corporate office may manage different EAP external contractors because of different service areas of EAP contractors or different services required by worksite employees. The data collection instrument as well as the telephone interviewers need to ensure that the data pertain to the sampled establishment.

In the NSWEAP, some worksites referred us to the external contractor. In such cases, a single external source may provide data for multiple establishments. The RU/SU linkage is maintained to compute the correct multiplicity-adjusted sampling weights. Repeated contacts of an RU for data for multiple SUs need to be avoided (to the extent possible) and monitored to control the response burden. Our experience indicated that repeated contacts of RUs (especially in large enterprises) for data related to multiple establishments impacted not only the current data collection effort but also future surveys conducted by the same or other researchers.

3.1.3 Analysis Unit

The third unit is the unit for inference, the analysis unit. The primary issues are: differences in survey design when inferences are desired for establishments, for enterprises, or for some other unit (such as the HI plan or the EAP); and the need for data to fully characterize the desired analysis unit. In household surveys, the housing unit is selected and inferences are often made based on the persons in the household. In establishment surveys, the situation again becomes complex because of the relationships between the sampling unit, the reporting unit and the analysis unit. An establishment survey may have multiple analysis units, analogous to having both the household and the person within the household as analysis units. For example, an enterprise-level survey may have both the enterprise and health insurance plan as analysis units.

Data collected must fully characterize the analysis unit. In the EAP survey, we can make estimates on the prevalence of EAPs at worksites. However, we cannot make estimates from the EAP survey on the proportion of enterprises that have an EAP, since we have insufficient data to characterize the enterprises.

For an establishment survey, such as NSWEAP, similar problems are encountered with the reporting unit being different from the sampling and analysis unit.² For

this establishment survey, cost data were often available for multiple establishments from a single reporting unit. There are several techniques available to accommodate this situation. Depending on the analyses, either multiplicity adjustments to the sampling weight, proration of the cost data using employee counts, or the computation of per employee costs were used to compute the estimates.

4. Effect on Estimation

The goals of the survey have a major impact on the units encountered in an ES. For the estimation of totals, such as the total cost of health insurance, the use of the enterprise as the sampling unit may facilitate the collection of cost data because of the location of the HI data. In the survey of employer-sponsored HI, the goal was for ratio estimates of “per employee” costs for the health care costs associated with the majority of employees in the enterprise. We collected sufficient data to characterize the enterprise with respect to the majority of the employees, but insufficient data for the computation of total HI costs. For prevalence estimates (for example the proportion of worksites with an EAP), the analysis unit will need to be fully characterized.

State, regional, and industry-specific inference problems exist in the computation of quantitative estimates (such as costs, employees covered, etc.) because an RU may provide information that crosses these estimation domains. Data can be prorated to specific geographic regions or industries, but the proration process (usually a ratio estimation process) tends to assign average estimates to these domains.

5. Discussion

The data being sought and the desired inferential population generally define the primary sampling unit (the establishment or the enterprise) for the survey. Zarkin et al. (in press) explored the issue of the choice of the sampling unit for employer-based health insurance surveys. For some data (e.g., number of full-time and part-time employees, wages, and state-level estimates), the establishment is the preferred sampling unit, while for other data (e.g., enrollment in health plans, risk and administrative structure, and variables affecting health insurance decisions), the enterprise (i.e., the corporate headquarters) is the preferred sampling unit. For a survey seeking all of these items, a multi-stage design using both enterprises and establishments as sampling units is appropriate. However, the source of some data may be at neither the enterprise nor the establishment, either internal or external to the enterprise.

References:

1. Federal Committee on Statistical Methodology. (1988) Quality in Establishment Surveys. Statistical Policy Working Paper 15. Washington D.C.: Office of Management and Budget.
2. Cox BG, Binder DA, Chinnappa BN, Christianson A, Colledge MJ, Kott PS (eds.). (1995) Business Survey Methods. New York: J Wiley & Sons, Inc.
3. In Canada, Australia, and many other countries, these are called generically business surveys.
4. Bureau of Labor Statistics (1988) "BLS Handbook of Methods" Bulletin 2285, Washington D.C.: U.S. Government Printing Office.
5. Federal Committee on Statistical Methodology. (1988) Quality in Establishment Surveys. Statistical Policy Working Paper 15. Washington D.C.: Office of Management and Budget.
6. Zarkin GA, Garfinkel SA, Potter FJ, McNeill JJ, (in Press) "Employment-Based Health Insurance: Implications of the Sampling Unit for Policy Analysis" INQUIRY.
7. Boyle KE, Potter FJ, Rush MW, Snodgrass J, (1994) "Survey Methodology and Performance Characteristics for the 1993 National Survey of Worksites and Employee Assistance Programs in the US." Presented at the 1994 American Public Health Association Conference.
8. Cox BG, Elliehausen GE, Wolken JD. (1989) "Surveying Small Businesses About Their Finances." Proceedings of the Section on Survey Research Methods, pp 553-557.
9. In "Frames and Business Registers: an Overview," Colledge (Colledge, 1995 in Cox, et al. 1995) uses the term "collection unit" to refer to the unit from which the data are collected.

Table 1. Sample Distribution and Response Status for NSWEAP

		Worksite Employees (From DMS)				
		Total	40 - 49	50 - 99	100 - 999	1,000 Plus
Sample	Count	6,488	707	1,700	3,283	798
Responding Worksites	Count	5,828	657	1,589	2,871	711
	Percent	90	93	93	87	89
Eligible Responding Worksites	Count	3,204	84	630	1,960	530
	Percent	55	13	40	68	75

Table 2. Sample Distribution and Response Status for an Enterprise Survey

		Enterprise Employees (From DMS)			
		Total	10 - 499	500 - 999	1,000 Plus
Sample	Count	1,897	301	241	1,355
Responding Enterprises	Count	1,245	232	164	849
	Percent	66	77	68	63
Eligible Responding Enterprises	Count	880	115	133	632
	Percent	71	50	81	74