NATIONAL EMPLOYER HEALTH INSURANCE SURVEY SAMPLE DESIGN

David A. Marker and Edward C. Bryant, Westat
Chris Moriarity, NCHS
David Marker, Westat Inc., 1650 Research Boulevard, Rockville, Maryland 20850

1. INTRODUCTION

The goal of the National Employer Health Insurance Survey (NEHIS) is to gather information on the number of employers offering health insurance, their costs, the coverage and characteristics of their health plans, and also employee-related estimates. Estimates are desired at both the state and national levels, and nationally by size of establishment and by size of firm. NEHIS includes samples of both public and private sector employers. The public sector employers have been stratified by state and by level of government, while the private sector employers have been stratified by state and by establishment size within firm size. This paper summarizes the sampling frames, stratification, sample allocation to the states, and plan subsampling used for the 1994 NEHIS. For more information please refer to the NEHIS Survey Design Report (1994).

2. SAMPLING FRAMES

The initial task in planning the sample selection for NEHIS was to identify suitable sampling frames. Ideally, these frames would contain one record for every establishment in the country in existence as of one specific date. The date should be relatively current and the file should not contain records for establishments no longer in business. Given the dynamic nature of establishments it is impossible for such an ideal frame to exist. However, it is possible to identify frames that contain the vast majority of establishments that were in business recently.

This section describes the different sampling frames that are being used for the NEHIS. Discussions were held with the Bureau of the Census, Bureau of Labor Statistics, Internal Revenue Service, and Small Business Administration to attempt to identify the best possible frames for this purpose.

Section 2.1 describes the frames used for the Federal, state, and local governments; Section 2.2 the frame used for private sector establishments with employees; and Section 2.3 the frame for private sector self-employed with no employees (SENEs). For the public sector the unit of analysis is the governmental entity. However, in many cases multiple governmental entities purchase health insurance for their employees through a combined purchasing unit (PU). In such cases the unit on the sampling frame will be the single purchasing unit. If the PU is sampled, the respondent will be asked about the health coverage for every governmental entity in the state that participates in the PU.

For the private sector the unit of analysis is the establishment. OMB (1988) states that, "An establishment is usually defined as an economic unit, generally at a single physical location, where business is conducted or services or industrial operations are performed." For multi-establishment firms (MEFs) the decisions regarding health insurance can be made at a variety of different corporate levels. Since the appropriate decision-making level is unknown a priori for each firm, it is best to sample individual establishments. When more than one establishment of a single firm is included in the sample we will begin our contacts at the firm headquarters, but only collect data on health benefits for individual sampled establishments within the firm.

Overview of Sampling Frames

All Federal employees and employees receiving their health benefits through state government (throughout this paper, state includes the District of Columbia) will be included with certainty. These data will be collected by contacting the Federal Office of Personnel Management, and independent Federal agencies, and every state government to collect information on their union and non-union employee health benefits.

The frame for public sector local governments is a combination of the 1987 and 1992 Census of Governments (COG). The frame for private sector establishments with employees is the Duns Market Identifiers (DMI) frame compiled by Dun & Bradstreet as of late 1993. This establishment level file contains data on approximately 10.1 million establishments. Approximately 200,000 governmental entities on this frame are in standard industrial classification (SIC) codes that are exclusively governments and were eliminated from the frame. Any other governmental establishments (or self-employed with no employees) on this frame identified during the data collection phase will be considered out-of-scope, since estimates for these populations will be derived from other frames.

Complete and up-to-date sampling frames of self-employed persons with no employees do not exist. While the IRS does have lists, they are not available as sampling frames. We examined the possibility of using the DMI file for this purpose but were not satisfied that their coverage of the self-employed would provide a sound basis for the types of analyses demanded by NEHIS. Thus the last two quarters of the 1993 National Health Insurance Survey (NHIS) are used as a frame for this population. The

NHIS is a survey of the civilian non-institutionalized population of the United States, conducted by NCHS. That survey asks if the respondent is self-employed. Two quarters of the most recent data were available during the spring of 1994.

NEHIS attempts to collect information for the 50 states and the District of Columbia. No data are collected for Puerto Rico, the Virgin Islands, or other territories. Also, no data are collected for employees of Native American Indian tribal governments.

2.1 Public Sector Frame

The public sector of the employer survey includes the Federal government, quasi-governmental agencies within the Federal government, the state governments, and local governments.

The Census of Governments (COG) listing of governments was used as the basic sample frame for local government agencies: counties, municipalities, townships, school districts, and special districts. The 1992 COG, having been recently completed, appeared to be the most complete and timely source, but its Directory had not been released at the time the sample was drawn and numbers of employees, by type of local government within state, also had not been released. The 1992 listing of local governments with names and addresses was made available to Westat by a special tabulation from the Bureau of the Census, however, and this provided the primary sample frame for local governments.

The 1987 COG states "A government is an organized entity which, in addition to having governmental character, has sufficient discretion in the management of its own affairs to distinguish it as separate from the administrative structure of any other governmental unit." The COG also contains dependent school districts even though they lack the degree of autonomy required of other governmental entities. Such districts are treated as governments for NEHIS. The other addition to the NEHIS list of governments resulted from telephone conversations with states to determine the composition of purchasing units. Some states stated that not all of their state colleges' and universities' employees were covered by the state health benefits PU. In such

cases, the uncovered colleges and universities were added to the sampling frame as separate governments. Numbers of employees in local governments were needed in order to stratify local governments by size, but these counts of employees were not available for the 1992 COG. They were available, however, for the 1987 COG. Since the Bureau of the Census used the same identification numbers for the 1987 and 1992 censuses, it was possible to match the two censuses and attach the 1987 number of employees to a large proportion of the 1992 list. The summary results of the matching are shown in Table 2-1. The column headed "Unknown" identifies the number of governments, by type, in the 1992 COG file that had no match in the 1987 file. The column headed "None" identifies the number of governments in the 1992 file that matched governments in the 1987 file that had zero employees in the 1987 census.

Out of 87,568 local governments in the 1992 COG file, 91.6 percent were found in the 1987 file ranging from counties matching 99.8 percent of the time to special districts with 83.1 percent. It was expected, of course, that the unknown number of 1992 employees would in some cases be substantially different from the numbers in 1987, but the size classes in 1987 were expected to be sufficiently close to the size classes in 1992 to serve adequately as a stratification variable for the 1992 frame. Preliminary results from the 77 local governments included in the NEHIS dress rehearsal indicate that, especially for special districts, changes from the 1987 frame size may reduce the utility of the stratification.

Purchasing Units

Health insurance for employees of local governments often is not provided by the local government. Common arrangements are for the state to cover the employees of aggregations of local governments as part of its overall state plan or for local governments to form groups with a common purchasing agent for health insurance. Both of these arrangements are referred to subsequently as "purchasing units" (PUs). Significant adjustment of the sample frame was necessary to accommodate these arrangements. Local governments that combined to

Table 2-1. Number of local governments in the 1992 COG file, by 1987 employee size classes

Government type	Employee size								
	Unknown	None	1-5	6-24	25-49	50-249	250-999	1,000+	Total
County	5	6	3	48	279	1,535	795	374	3,045
Municipality	175	1,175	11,961	14,318	2,931	3,952	1,143	316	35,971
Special District	5,546	12,789	8,779	3,363	969	996	211	84	32,737
School District	1,586	154	556	1,304	1,871	6,277	3,245	822	15,815
Total	7,312	14,124	21,299	19,033	6,050	12,760	5,394	1,596	87,568

have a common purchasing agent for health insurance were eliminated from the sampling frame, but replaced by the purchasing unit, an entity not originally on the government frame. The principle is clear enough, but effecting it required a number of decisions relating to particular issues that are described in the NEHIS Survey Design Report (1994).

To identify the large purchasing units in every state at least five organizations were contacted, but at least one central contact was attempted for State, County, Municipal, Special District, and School District employees in every state. This procedure was not intended to produce an exhaustive list of every PU in every state. Rather, it attempted to identify all large PUs whose inclusion in the sample would significantly impact the quality of survey estimates.

The final sampling frame was reduced from 87,568 to 77,691, a reduction of about ten thousand due to the purchasing unit aggregations. No attempt was made to identify all PUs – just the large ones. An unknown number of additional PUs may be identified during the field work.

2.2 Frame for Private Sector Establishments with Employees

A variety of alternative sampling frames were considered for the private sector. The main concerns with the DMI frame are its coverage of the farm sector, new establishments, and small establishments (including the self employed). While the DMI frame does not cover 100 percent of the farm sector, D&B's methods for frame construction (including reviewing public records, secretaries of state, bankruptcy courts, district courts, Departments of Motor Vehicles, etc.) appear likely to include most farms that have employees. The DMI frame's coverage of private employers and employees was compared with other government sources, and found to be sufficient for use as the sampling frame for NEHIS. Table 2-2 shows the national distribution of private sector establishments on the DMI frame by firm and establishment sizes after all establishments in SIC codes known to correspond to public sector establishments were eliminated.

2.3 Frame for Private Sector Self-Employed With No Employees (SENEs)

The National Health Interview Survey conducts in-person interviews with approximately 50,000 households each year. As part of that survey respondents are asked about their employment. Some respondents indicate that they are primarily self-employed. All self-employed respondents from the last six months of 1993 NHIS data collection (anticipated to be between 3,000 and 4,000 respondents) will be used as a frame for interviewing SENEs. (While some NHIS respondents (approximately 6%) do not have telephones, it is anticipated that a smaller percentage of the SENEs will not have telephones.) Any NHIS self-employed respondents who indicate that they have employees will be considered out-of-scope for this part of NEHIS. By including all NHIS SENEs in the NEHIS it is anticipated that at most 1,000 interviews will be completed with SENEs.

Unlike the NEHIS, the 1993 NHIS is not a state-stratified sample. Therefore, the data from the NHIS SENEs will not provide accurate estimates for the SENE component of individual states. Rather the SENE data will only be used when producing national estimates. State-level estimates from NEHIS will be restricted to establishments with employees.

3. SAMPLE ALLOCATION
3.1 Stratification

Public sector governments were stratified by type of government and number of employees. Private sector establishments were stratified by firm size and establishment size, where size was measured by number of employees. Before placing every government into a sampling stratum, the governments that were known to belong to a purchasing unit were replaced by a record for that PU. The sampling frame then consisted of a record for every government that did not belong to a PU, plus one record for every PU. Before placing every government into a sampling stratum, the governments that were known to belong to a purchasing unit were replaced by a record for that PU. The sampling frame then consisted of a record for every government that did not belong to a PU, plus one record for every PU.

Table 2-2. DMI sample frame by firm and establishment size

Firm size*	Establishment size								
	Unknown	1 no other	1-5	6-24	25-49	50-249	250-999	1000+	Total
1-49	1,197,959	1,105,384	4,884,932	1,539,707	210,621	0	0	0	8,938,603
50-999	26,094	22	105,387	147,541	61,946	190,392	16,991	0	548,373
1000+	29,504	29	83,353	147,134	51,041	77,550	18,825	7,478	414,914
Total	1,253,557	1,105,435	5,073,672	1,834,382	323,608	267,942	35,816	7,478	9,901,890

* Based on definitions of firms described in Section 3.1

The PUs were assigned to a size stratum based on the total number of employees in governments that are part of the PU. They were assigned to a type of government stratum based on the type of government of the plurality of the employees in that PU.

The DMI frame contains a series of variables that identify corporate family structures. For every establishment that is part of a corporate family DMI identifies its headquarters, the "ultimate" (highest level) headquarters of the firm, and how far down the corporate structure the establishment's headquarters is from the ultimate headquarters.

Since subsidiaries often have different fringe benefit programs, Westat identified as separate firms any first or second level subsidiaries (any headquarters who report directly to the ultimate headquarters or with only one intermediate step) along with the establishments who report through such headquarters. Each establishment was assigned to a firm stratum based on the total number of employees (excluding those who report to one of its first or second level subsidiaries) who report to the highest headquarters in that firm. Establishments that were not part of a MEF were assigned a firm stratum based on the number of employees at that one establishment.

The purpose of this firm creation procedure is to attempt to stratify establishments by the size of firm that is most likely to be controlling the health benefits decisions. Thus, for example, we have separated Electronic Data Systems (EDS) and Hughes Aircraft from General Motors. However, we have probably also separated some subsidiaries who do receive benefits through a higher level headquarters. Without frame information on health benefits, it is impossible to assign firm definitions that would match perfectly any particular analytic uses of the survey data.

3.2 Certainty Selections
All federal agencies will be included with certainty, by obtaining data from the agency responsible for their fringe benefits. All state employees receiving their health benefits through their state government will be included with certainty. In two states, the District of Columbia and Hawaii, all public sector employees are covered by the health benefits of the central government. In these "states" the entire public sector is therefore selected with certainty.

All local governments or PUs with more than 10,000 employees were included with certainty. Depending on the distribution within each state other public employers with more than 250; 1,000; or 5,000 employees may also have been selected with certainty.

In a few states there were only a few large private sector employers. In such cases they were also selected with certainty.

3.3 Sample Allocation Plan
The sample allocation plan proceeded in six steps as follows:
1. Allocate a portion of the resources to the 100 percent coverage of federal and state employees and the sample coverage of self-employed with no employees from the NHIS.
2. Allocate the remaining sample units (in terms of completed responses) to the 51 states, taking into account the need to have state estimates and the need to protect the precision of national estimates.
3. Allocate the within-state samples between local government and private establishments.
4. Within private establishments, allocate the sample to the firm size by establishment size cells in terms of number of employees.
5. Within the local government sector, allocate the sample to the type of government by employment size cells.
6. Adjust the allocated sample sizes in terms of completed interviews to initial sample sizes, taking into account expected in-scope rates and response rates.

3.3.1 Allocation of Resources
Data for all Federal employees will be collected by contacting the Federal Office of Personnel Management and independent Federal agencies and all employees of every state government will be included with certainty, including the local governments that receive their health insurance through the state. Resources will remain for approximately 40,000 interviews to be distributed across the non-self employed private sector and local governments. The allocation of these resources is described in the following sections.

3.3.2 Allocation Across States
Estimates of employer health insurance plans and characteristics of employers are required for each state and the District of Columbia. One way to approach the design is to allocate the sample of approximately 40,000 completed interviews equally to the 51 states. An equal allocation would provide approximately 785 interviews for each state and would assure that the state estimates were approximately all equally precise. Equal allocation reduces the precision of national estimates because a large part of the total sample is allocated to small states that have a relatively small portion of total employees in the nation. A compromise was chosen that balances the need for state estimates against the need for precision in the national estimates.

Three models for allocation to the states were examined. They are referred to as allocations A, B, and C. Allocation A is equal for every state. Allocation B is in proportion to the total number of employees in each state. Allocation C is allocation

in proportion to the 0.3 power of the total number of employees in each state. To put this transformation in perspective, the 0.5 power is the square-root transformation and the 0.33 power is approximately the cube root, so the 0.3 power is a slightly more severe transformation than the cube root. The 0.3 power transformation seemed to balance the allocation between large and small states, permitting larger samples for the larger states, but permitting a sample for small states of at least 450 interviews.

This transformation balances the need for separate estimates by state against the need for obtaining precise estimates for the nation as a whole. Some relative standard errors of national estimates were estimated for the three methods. For these purposes, the strata were considered to be the states, so the comparison implies that the efficiency of the allocation to strata within the states is constant from state to state. Using this concept, every state under Allocation A would have a relative standard error of about 0.033. Under Allocation B the relative standard errors would vary from about 0.014 for California to 0.101 for Alaska, a factor of more than seven. Under the 0.3 exponential transformation, the state estimates would vary from about 0.024 for California to 0.043 for Alaska, a factor of less than two. Nationally, the relative standard error under Allocation C (the exponentially transformed allocation) would be about 18 percent less than with equal allocation (Allocation A), but would be about 21 percent higher than for the proportional allocation (B). Allocation C appeared to be a good compromise and therefore was chosen. The allocation is shown in the last column of Table 3-1.

3.3.3 Allocating the State Samples Between Private and Local Government

Initially, it was planned to allocate samples between the private sector and local government based on number of employees. Review of the proposed allocation led to consideration of a reduction in the number of local-government interviews and an increase in the private-sector interviews on the assumption that local governments are likely to be more homogeneous than are private employers with respect to the insurance plans made available to them and such characteristics as percent of employees covered. Therefore after reviewing alternative choices, it was decided to give local government employees 0.66 of the weight given employees of private employers in allocating the sample. The resulting allocation is shown in the middle columns of Table 3-1.

4. SUBSAMPLING OF HEALTH INSURANCE PLANS

Ideally, data would be collected from all sampled establishments regarding all of the health

Table 3-1. Expected number of completed interviews for the local government and private sector

State	Local government	Private sector	Total
AK	47	415	462
AL	65	748	813
AR	50	634	685
AZ	68	693	762
CA	118	1322	1440
CO	60	716	776
CT	40	738	779
DE	20	476	496
DC	0	531	531
FL	92	1049	1141
GA	75	897	973
HI	0	541	541
IA	59	656	715
ID	52	462	514
IL	75	1053	1128
IN	61	865	926
KS	66	634	699
KY	59	724	783
LA	71	728	799
MA	48	893	941
MD	53	785	838
ME	41	515	556
MI	77	961	1038
MN	65	803	868
MO	59	836	895
MS	67	612	679
MT	46	408	454
NC	75	919	994
ND	46	443	490
NE	58	565	622
NH	36	513	549
NJ	76	930	1006
NM	85	497	582
NV	37	572	609
NY	111	1175	1286
OH	75	1042	1118
OK	60	652	712
OR	58	667	725
PA	63	1063	1126
RI	21	504	525
SC	64	728	791
SD	42	460	502
TN	57	837	894
TX	111	1135	1247
UT	48	583	631
VA	70	863	933
VT	31	452	483
WA	69	786	855
WI	62	842	904
WV	50	545	596
WY	59	387	446
Totals	2,997	36,859	39,856

insurance plans that are available to their employees. Unfortunately, given the respondent burden, this is impractical in situations where the number of plans is large. In such cases it is necessary to subsample plans for which data will be collected.

4.1 Single Establishment Firms (SEFs)

For single establishment firms (SEFs) that offer five or fewer plans data will be collected on all plans. If more than five plans are offered data will be collected on only five. In such situations a maximum of one single-service plan (e.g., dental or eye care) will be selected, along with four (or five, if there are no single-service plans) other plans. The SEF procedure for subsampling plans will be based on probabilities proportional to enrollment.

Enrollment data for each plan will be collected for the actual establishment that is being sampled. If enrollment data by plan are not available the plans will be selected with equal probabilities.

4.2 Multi-Establishment Firms (MEFs)

The situation is more complex with sampled MEFs. MEFs are defined here by a common health benefits grouping. MEFs may offer single-service plans, company-wide plans (it is only required that such plans cover all sampled establishments, whether or not nonsampled establishments are covered is irrelevant), and local plans which are available to employees of some but not all sampled establishments. It is not feasible to collect enrollment information on every plan at every sampled establishment. Therefore the following rules have been developed for subsampling plans of MEF.

MEFs with five or less plans will have all plans selected and data collected on all sampled establishments.

For MEFs with more than five plans, if all plans are company-wide the SEF rules for plans will be followed. Again, data will be collected on all sampled establishments.

For MEFs with more than five plans, including some local plans, the procedures are as follows. For the MEFs that were sampled in five or fewer states subsample one establishment per state with equal probability. Sample one company-wide single-service plan and two other company-wide plans, in addition to *two* local plans per subsampled establishment. This allows for a maximum of 13 plans per MEF (although the average burden per sampled establishment is under three plans).

For the MEFs sampled in 6-10 states, the procedures are the same except for selecting only *one* local plan per subsampled establishment. This again allows for a maximum of 13 plans per MEF (although the average burden per sampled establishment is now under two plans).

The remaining MEFs sampled in more than ten states are likely to require unique, off-line, procedures for many sections. For these MEFs the establishments are sorted by Census Division, then by state, and by size within state. A systematic equal probability subsample of ten establishments (while assuring that no two of the subsampled establishments are in the same state) will be selected. The same procedure for sampling plans as for MEFs sampled in 6-10 states will be used. This continues the maximum of 13 plans per MEF (while reducing the average burden per sampled establishment to under one plan).

Selection of a single-service plan will be done with equal probability. Also, for MEFs in 6-10 states, where only one local plan is to be selected per subsampled establishment, the local plan is selected with equal probability. When two company-wide or two local plans are being selected determine whether any one plan has at least 50 percent of the enrollment in company-wide (or local) plans company-wide. If so that plan is taken with certainty and the one other plan is selected with equal probability. If no plan exceeds the 50 percent threshold, two plans are selected with equal probability.

The above procedure will be downwardly biased for estimating "the percentage of *establishments* with a plan offering option X," as would any procedure involving subsampling of plans. This bias only affects establishments with more than five plans. However, it will be unbiased for estimating *plan* characteristics.

4.3 Public Sector

There will be no subsampling of plans for state or Federal governments. Information will be collected on every plan that is offered, along with data on the total number of employees, etc. in the state government (and Federal employment by state).

Local governments that are part of state government purchasing units will have information collected on all of their plans as part of the state government interview. For local governments that are part of other (non-state government) PUs and for non-PU local governments, we will subsample health care plans in an analogous fashion to SEFs. For a sampled non-PU local government we will use the SEF rules described in Section 4.1, except that any plan with an enrollment equal to less than 5 percent of total enrollment in major health plans at that government will have its measure of size, for subsampling purposes, set equal to 5 percent of total enrollment. For all local governments that are part of a PU plans will be subsampled as for non-PU local governments, except that the measure of size for each plan is based on total enrollment for the plan across the entire PU.

5. Reference

NEHIS Survey Design Report submitted by Westat, Inc. to the National Center for Health Statistics, May 5, 1994.

OMB Statistical Policy Working Paper 15, Quality In Establishment Surveys, July 1988.