

**NATIONAL EMPLOYER HEALTH INSURANCE SURVEY (NEHIS)  
NATIONAL CENTER FOR HEALTH STATISTICS (NCHS)  
RESEARCH DATA CENTER (RDC) FILE DOCUMENTATION**

INTRODUCTION

The National Employer Health Insurance Survey (NEHIS) was developed to produce estimates on employer-sponsored health insurance data in the U.S.. The NEHIS was the first federal survey to represent all employers in the United States by State and obtain information on all plans offered to employees by their employers.

This documentation accompanies release of four files from the NEHIS: the *Establishment File*, the *Plan Analytic File*, the *Plan Variance File*, and the *Self-Employed with No Employees (SENE) File*.

<i>File Description</i>	<i>Unit of analysis</i>
<b>Establishment File</b> <sup>1</sup>	<b>Establishment or government</b>
<b>Plan Analytic File</b>	<b>Health insurance plan</b>
<b>Plan Variance File</b> <sup>1</sup>	<b>Health insurance plan</b>
<b>SENE File</b> <sup>1</sup>	<b>Person records</b>

The first three files are considered confidential because they are designed to produce State-level estimates. Because of confidentiality constraints, the National Center for Health Statistics (NCHS) is not able to publicly release NEHIS data along with State identifiers. By releasing the NEHIS Establishment and Plan Files through the Research Date Center (RDC), however, users can access these confidential (State) data in a controlled environment that allows RDC staff to identify and monitor potential confidentiality violations.

The NEHIS files at the RDC were created by NCHS staff upon completion of data quality control on files delivered by Westat, the NEHIS contractor. These files differ from the Westat-delivered files in three major ways: 1) they primarily contain constructed variables that are of analytic interest to data users and do not include the vast majority of the original survey variables used in variable construction; 2) 312 records with firm and establishment sizes of 1 were deleted because they were identified as Dun's Market Identifiers file out-of-scope cases; and 3) they have been stripped of fields that directly identify the sample unit, such as name and address of the establishment.

The RDC NEHIS file documentation consists of this Introduction, the *Codebooks* to the files, and

---

<sup>1</sup> This file includes survey design variables needed to analyze the data using SUDAAN software, which produces variance estimates that correct for complex survey design.

additional materials to assist users in better understanding the variables and how they were edited. The *Glossary of Variables* provides descriptions of the meaning and analytic use of constructed variables. *Sample SUDAAN Procedure Statements* are provided to assist users in computing sample variance estimates.

A brief description of the survey design and content follows. It should be noted that additional information about the survey may be obtained from previously published reports <sup>2, 3, 4</sup> and the NEHIS website (<http://www.cdc.gov/nchs/about/major/nehis/nehis.htm>).

### SAMPLE DESIGN

NEHIS was a stratified random sample survey of private establishments, governments, and self-employed individuals with no employees (SENE). Two of the survey's objectives had a major impact on the sample design employed for NEHIS. These were: 1) to provide national estimates and estimates for the 50 States and the District of Columbia of the number and percentage of private and public establishments that offer health insurance to employees; and 2) to characterize the number and type of health insurance plans that contract with these establishments and describe the costs and benefits they provide.

For the *private sector*, the sample unit was the establishment, defined as "an economic unit, generally at a single physical location, where business is conducted or services or industrial operations are performed." The major reason that establishments were sampled in NEHIS rather than firms is that establishments, by definition, are confined within State borders, thus facilitating estimates by State, while firms may cross State lines.

For the *public sector*, the sample unit was the government entity, i.e., Federal, State, county, municipality/township, school district, and special district. Exceptions were some government units that jointly purchased employees' health insurance through a purchasing unit, and here, the purchasing unit became the sample unit rather than the individual units that purchased insurance through the purchasing pool.

*Three sample frames* were used for the NEHIS:

- Private establishments were sampled from the October 1993 *Dun's Market Identifiers* (DMI) File, publicly available from Dun and Bradstreet.
- Local governments were sampled from the 1992 *Census of Governments* File maintained

---

<sup>2</sup> National Center for Health Statistics. *Employer-sponsored Health Insurance: State and National Estimates*. Hyattsville, Maryland. 1997.

<sup>3</sup>Moss, AJ. *Plan and Operation of the National Employer Health Insurance Survey*. National Center for Health Statistics. Hyattsville, Maryland. 1999.

<sup>4</sup> Allen K, Park C. *Health Insurance Coverage for the Self-employed with No Employees*. National Center for Health Statistics. Hyattsville, Maryland. 1999.

by the U.S. Bureau of the Census and supplemented by the *Higher Education Directory*<sup>5</sup>. As with the private sector, the purchasing unit was used as the sampling unit for local governments who purchased through those collective units.

- The SENE sampling frame came from the *1993 National Health Interview Survey* Person Record Quarters 3 and 4 files (chosen to address the concern that the DMI file under-represented those individuals).

The Federal government and State governments were included with certainty. The District of Columbia and the State Governments of Maryland and Virginia, however, chose not to participate in the survey.

The final number of completed interviewed NEHIS sample units obtained from the three frames was: 34,604 private sector establishments (excluding SENE); 919 SENE cases; and 3,214 government (public sector) cases.

#### HEALTH INSURANCE PLAN SUBSAMPLING

In addition to a sample designed to provide State estimates of establishments and governments that offer health insurance, another unique feature of the NEHIS is a sample of plans representative of all health insurance plans offered to employees. The NEHIS design subsampled health insurance plans to reduce response burden. When five plans or fewer were offered, data were collected on all plans. However, for businesses offering six or more plans, five plans were subsampled using a methodology that obtained a representative mix of plans in the sample. A separate plan subsampling methodology was developed for firms with large numbers of establishments in the NEHIS sample, where no more than 13 plans were subsampled within the firm. No plan subsampling was performed for the Federal Government and State governments; these government units (with the exception of four States) reported on all of their plans. Subsampling of plans was implemented for about 11 percent of private sector establishments with insurance and for about 9 percent of local governments (and purchasing units) with insurance. A more detailed description of NEHIS survey design has already been published<sup>3</sup>. Alternatively this report- *Plan and Operation of the National Employer Health Insurance Survey* (#99-104) can be accessed through the following website:  
<http://www.cdc.gov/nchs/products/pubs/pubd/other/miscpub/miscpub.htm>.

#### DATA COLLECTION

The National Employer Health Insurance Survey (NEHIS) was conducted between April and December of 1994 (see *Plan and Operation of the National Employer Health Insurance Survey* #99-104). Interviews were conducted with respondents identified as the most knowledgeable about health benefits for the establishment. The overall final unit response rate achieved for NEHIS was about 72 percent, although there was substantial variation by type of sample case, as well as by State and firm size. Most interviews were conducted via telephone using a computer-assisted telephone interviewing (CATI) system. The CATI methodology was implemented for several reasons: complexity of the question sequence, expected use of multiple respondents, the

---

<sup>5</sup> 1994 *Higher Education Directory*, Higher Education Publications Inc. 1994.

large number of sampled cases, limited subject-matter expertise of interviewers, and rapid data turnaround. The survey contents included information about the availability and characteristics of employer-sponsored health insurance coverage, plan benefits, and costs. Information on plan characteristics was collected for both plans with enrollees and plans with zero enrollees (offered to employees but not used). The CATI questionnaire used to collect these data can be accessed via the NEHIS home page at: <http://www.cdc.gov/nchs/datawh/ftpserv/ftpdata/ftpdata.htm>.

SENE data were collected by telephone using a paper questionnaire. The SENE questions were the same, or were modeled after, those items collected for establishments and health plans in the CATI questionnaire. Because some questions asked of establishments were not relevant for self-employed persons (e.g. number of employees), and there was only one possible respondent, CATI was not necessary for SENE. The SENE questionnaire is shown in Appendix III of the report - *Health Insurance Coverage for the Self-Employed With No Employees*.<sup>4</sup> This report can also be accessed through the website:

<http://www.cdc.gov/nchs/products/pubs/pubd/other/miscpub/miscpub.htm>

### NEHIS DATA FILES

All files are in ASCII format on CD-ROM. Apart from the SENE data, the NEHIS data base contains two distinct types of variables: establishment-based items, such as the number of employees at the establishment eligible for health benefits; and plan-based items, such as the amount of the premium paid by an employee for a specific plan offered at an establishment. The two types of analytic estimates that can be obtained also mirror how the NEHIS data files are configured. Separate files were constructed for the establishment data base and the plan data base.

- The *Establishment File* contains one record for each private sector sampled case (excluding SENE) and for each government case with a final result code equivalent to “completed or partially completed interview”.
- The *Plan Analytic File* contains one record for each health insurance plan that was offered at a business establishment or government with a completed or partially completed interview.
- The *Plan Variance File* is the same as the Plan Analytic File, except it includes dummy records and sample design variables needed to compute variances using the SUDAAN software.
- The *SENE File* contains one record for each self-employed individual (with no other employees) and 21 dummy records. All plan information for that individual, as well as sample design variables needed to compute variances using the SUDAAN software, is included on this record.

<i>File Description</i>	<i>Unit of analysis</i>	<i>Number of records</i>
<b>Establishment File</b>	<b>Establishment or government</b>	<b>37,818</b>
<b>Plan Analytic File</b>	<b>Health insurance plan</b>	<b>46,517</b>
<b>Plan Variance File</b>	<b>Health insurance plan</b>	<b>59,309 ( includes 12,792 dummy records)</b>
<b>SENE File</b>	<b>Person records</b>	<b>940 ( includes 21 dummy records)</b>

Note again that there are two versions of the Plan file: 1) an analytic version, and 2) a version designed for computing variances using SUDAAN software. The Plan Variance File includes sample design variables and dummy records needed to calculate standard errors with the SUDAAN software<sup>6</sup>. In the case of the Plan Variance File, nearly 13,000 dummy records were added so that design information was available for establishments that did not offer health insurance. Other than the SUDAAN variables and dummy records, the two Plan files are identical in content. The frequencies provided on the tape layout for these two files match the frequencies found on the Plan Analytic File (without dummy records). Two sample SUDAAN programs included with this documentation illustrates how the dummy records and SUDAAN sample design variables are used in the SUDAAN program.

The SENE File also includes dummy records<sup>7</sup> so that when computing variance estimates, complete sample design information is available for all strata and PSUs included in the National Health Interview Survey (the sampling frame for the SENE sample). Since only 21 dummy records were added to the SENE File, there is only one SENE File. The SENE File includes an indicator for the dummy records, SENEDATA. The frequencies provided on the SENE tape layout do not include dummy records ( i.e. SENEDATA=0). A sample SENE File SUDAAN program is also included with this documentation.

MERGING FILES

Although separate files were constructed for the establishment data base and the plan data base, the two files can be linked by establishment identification numbers. Each establishment is uniquely identified by the concatenated variables: CASEID||ESTBNUM. Each health plan is uniquely identified by: CASEID||ESTBNUM||PLANNUM. Thus, all plans within the same CASEID||ESTBNUM are offered by that establishment.

Linkage may be desirable if the file of interest is the Plan Analytic File, but a variable of interest is included only on the Establishment File. Linkage allows matching cases and copying variables

---

<sup>6</sup>Moriarity, Gousen, and Chapman. *Use of SUDAAN for the 1994 NEHIS Plan-Level Estimates*. Internal memo dated 6/19/97.

<sup>7</sup>Moriarity and Gousen. *Use of SUDAAN for 1994 SENE Estimates*. Internal memo dated 3/30/98.

of interest from the Establishment to the Plan files (one to many match). To match plan variables to the Establishment File, the plan variable of interest must be aggregated or summarized for the establishment level prior to merging the data. For example, the number of active employees enrolled in all major health plans (i.e., Health Maintenance Organization (HMO), Preferred Provider Organization (PPO), Conventional Indemnity plan, or Point of Service (POS) plan) as of the end of the plan year (EEENRP2) can be summed within an establishment on the Plan File and matched and compared with I\_EESCOV that contains employee enrollment on December 31, 1993 on the Establishment File (see discussion below). Several such variables are already included on both the Establishment and Plan Files. For example, establishment size (ESTSIZEN), firm size (FIRMSIZN), and industry code (SICCODE) from the Establishment File are included on the Plan files so that plan and establishment characteristics can be analyzed according to the same establishment, firm size, or industrial groups.

It may also be desirable to match variables from an external file such as the *Area Resource File* to the NEHIS files. Matching variables from such files is possible, but can only be performed by the RDC staff onsite so that they can ensure that the matching does not present confidentiality concerns.

It should be noted that some “apparently” similar estimates can be computed using either the Establishment File or the Plan files. These variables, however, have different names with different interpretations. For example, the Establishment File includes I\_EESCOV, the number of employees covered by a health plan on December 31, 1993, while the Plan files include EEENRP2, the number of employees covered by a health plan as of the end of the last plan year ending before April 1, 1994. Weighted estimates of these two variables, both of which include imputed values, will not be identical for the following reasons:

- I\_EESCOV and EEENRP2 represent covered employees at different points in time.
- Weighted plan and establishment estimates will differ because they benchmark to different population totals. The establishment weight includes adjustments to represent the population of establishments and governments selected. The plan weight includes additional adjustment factors to account for subsampling of plans and non-response.
- EEENRP2 is likely to include more double counting of employees than I\_EESCOV because it includes enrollment for single service and special plans (e.g., plans that covered only dental, vision, or prescription drugs, or plans that only covered long-term care services).
- EEENRP2 needs to be summed within the same establishment (CASEID||ESTBNUM) on the Plan files to get a count that is comparable to I\_EESCOV for the same establishment. This was done for the summary plan variable IPEESMAJ already included on the Establishment File. IPEESMAJ represents the total number of employees enrolled in a major plan during the 1993 plan year (i.e. sum of EEENRP2 for HMO, PPO, Conventional/Indemnity or POS plans within an establishment). Those two numbers, however, will only be the same if the 1993 plan year was the same as calendar year 1993 (about half of the plans).
- Non-sampling errors associated with response for the two variables differ.

Because of these differences, it is recommended that analysis of these particular variables be

confined to the files where they are located.

The SENE data base, on the other hand, cannot be merged with the Establishment or Plan files because of differences in sample design and unit of analysis. In addition, the SENE survey was designed to produce only national estimates, while the establishment and plan data were designed to produce both State and national estimates.

#### USE OF SURVEY WEIGHTS

The probability design of the NEHIS establishment and plan samples allows the data to be weighted to produce representative national and State-level estimates. ***Unweighted data should not be used for analysis of NEHIS files because the unweighted data yield biased point estimates.*** Use of the weights take into account the disproportionate sampling, clustering of observations, stratification, nonresponse and other adjustments used in NEHIS. The following website includes two publications with NEHIS self employed (#99-1024) and State and national estimates (#98-1017):

<http://www.cdc.gov/nchs/products/pubs/pubd/other/miscpub/miscpub.htm>.

Each NEHIS analytic file contains the appropriate survey weight for data analysis. On the Establishment File, this weight is called FESTB\_WT. By aggregating this weight, estimated establishment totals for national and State data can be obtained. For example, the sum of FESTB\_WT, cross-tabulated by INSURE2, gives the estimated number of establishments and governments that offered health insurance to their employees on December 31, 1993. The number of employees and other numeric characteristics can be estimated by multiplying that characteristic times the establishment weight and aggregating that sum. Thus, the estimated number of employees is the sum  $FESTB\_WT * ESTSIZEN$  across all cases on the Establishment File. For estimates of the private sector that match published estimates, the domain should be limited to SAMPTYPE=1 (private establishments).

Similarly, the SENE File weight, NRWT1, can be aggregated to estimate the number of self-employed individuals with no employees in the United States. The SENE File can produce only national estimates due to small sample size. For example, of 919 responding SENE cases, information on health plan coverage was collected for only 630 cases- the remainder did not have health insurance coverage. Since the SENE sample used the National Health Interview Survey (NHIS) as the sampling frame, additional information from that survey's Health Insurance Supplement is also included on each SENE record, as well as SENE-related demographic and socio-economic variables.

In contrast to the Establishment and SENE Files, the Plan files have both a national weight and a State weight. NCHSPLWT produces unbiased national estimates, but estimates in certain States are problematic due to excessive weight factors assigned to plans because of subsampling. The State weight, STATEWT, includes an additional post-stratification adjustment to improve State-level estimates.

Finally, it should be noted that estimates derived using the Plan files can be estimated by either per-plan or per-enrolled employee. For example, the monthly premium for single coverage can

be estimated for a “typical plan” or as the average amount for enrolled employees. This may be illustrated using the combined monthly premium for single coverage (ISNGPREM), the plan sample weight (NCHSPLWT), and the number of employees with single coverage (IEESINER).

- If one is interested in the typical (per-plan) premium of single coverage offered by employers, it is estimated by the average of  $ISNGPREM * NCHSPLWT$ .
- The average amount paid for employees with single coverage (per-employee) is estimated by the average of  $ISNGPREM * NCHSPLWT * IEESINER$ .

Users should recognize that “per-plan” estimates weight plans with larger enrollments more heavily than when the plan weight is multiplied by the number of employees (per-employee estimates).

Users should also be aware that the number of cases included in these two computations will differ because premiums (or premium equivalents) for single coverage were collected for plans offered even when no employees were enrolled in the plan. Thus, the number of cases with a premium in ISNGPREM is 37,140 cases, but only 30,890 cases involving a major plan (HMO, PPO, Conventional/Indemnity, or POS) had employees enrolled with single coverage. Premium information was collected only for these four types of plans <sup>8</sup>.

#### LIMITATIONS OF DATA

Users of the NEHIS data should be aware of several limitations/differences associated with them. First, even though the survey of establishments has a large number of observations, response among establishments in firms with over 1,000 employees was relatively low- 55 percent. Caution should be exercised in interpreting results based on estimates with low response rates. Item response rates also vary by survey question. Unit response rates for selected domains as well as item response rates of key NEHIS survey variables have already been published <sup>3</sup>.

Second, estimates from NEHIS may differ somewhat from employer estimates from other nationally representative government surveys such as the ES-202 program, the *Employee Benefits Survey (EBS)* from the Bureau of Labor Statistics, or estimates from the *County Business Patterns* report published by the Bureau of the Census. Possible reasons for discrepancies include: differences in definitions of establishments, differences in definition of employees included in the sample, differences in coverage of the NEHIS establishment universe with those used by other employer surveys, and reporting error. For example, EBS and NEHIS collected information for different groups of employees. The EBS collected health insurance premiums only for full-time employees, whereas NEHIS collected premiums for all employees.

Third, NEHIS estimates of State government employees differ from published estimates from the *1992 Census of Government (COG)* and the *1993 Survey of Government Employment (SGE)*. An internal evaluation of estimates from these three sources found that NEHIS estimates of State government employees diverged from COG and SGE estimates, ranging from 37 percent higher to 27 percent lower in eight “problem” States. Users should be aware of these differences when

---

<sup>8</sup> Premium information was also collected for 34 Medicare Supplemental (Medigap) plans. Premiums for these plans are excluded in ISNGPREM and IFAMPREM; however, they are included in the unimputed premium fields.

analyzing public sector employees, particularly in these eight States- Kentucky, Minnesota, Mississippi, Nevada, New Mexico, Oregon, Washington, and West Virginia.

Fourth, it is recommended that for public sector estimates, users analyze only combined estimates of State and local governments (GOVTYPE=2-4). In a few States, the distinction between State and local government units is blurred, particularly in the case of dependent school districts. In “dependent school districts, ” public school systems may be operated by the county government, but all public school system employees are State employees.

Fifth, the NEHIS treatment of health insurance plans with major medical “Wraparound coverage” may differ from that of other surveys. A “Wraparound Plan” is a major medical plan that supplements a basic hospital/surgical medical plan<sup>9</sup>. The Wraparound plan covers all charges other than those provided by the basic hospital plan. From the employer’s point of view, the two plans supplement each other to provide employees comprehensive health insurance coverage. In the NEHIS, 262 such plans were identified with the assistance of a health benefits consultant. The presence of these plans required special edits. Employee enrollment in the two plans was often identical or double counted. This caused the sum of employees enrolled across all plans offered within a business with wraparound plans to exceed the total number of employees. In addition, premiums and employee contributions are underestimated if these two plans are treated separately.

Special edits for basic plans and their wraparound coverage plans were performed to adjust for these problems. Edits performed for these 262 plans included:

1. *Creating a recode.* SUMWRAP identifies a plan pair as either a “basic” plan or a “wraparound” plan. SUMWRAP is blank otherwise. Another field, WRAP, identifies the plan ID number (CASEID||ESTBNUM||PLANNUM) of the basic plan on 132 wraparound coverage plans.
2. *Unduplicating employee enrollment in these plans.* If the same employees were covered by the two plans, the enrollment was retained on the “basic” plan record, while enrollment for the wraparound record was set to “inapplicable”. If enrollment for the two plans double counted employees but the count was not identical, the difference in enrollments was assigned to the wraparound plan, while the “larger” enrollment number was assigned to the basic plan. This edit assumes employees had a choice in selecting wraparound coverage; thus, the “larger” enrollment count reflects employees with only basic coverage.
3. *Combining premiums for each level of coverage (single and family).* This was done when enrollment was identical in both plans. The combined premium is included on the “basic” plan record. The premium was then set to “inapplicable” on the wraparound record. When enrollment was double counted but was not identical, the combined premiums were recorded only on the wraparound record.
4. *Editing the base plan fields for covered services to “yes” (coded ‘1’).* When it was reported as such for the wraparound plan, edited base code reflects the combined plan’s more comprehensive coverage.

---

<sup>9</sup> Employee Benefit Plans: A Glossary of Terms International Foundation of Employee Benefit Plans. 1997.

Data fields reflecting coinsurance/copayments for these two plans were not changed on either record because both fields are of interest analytically. For instance, in Basic+wraparound supplemental coverage plans, deductibles from the basic plan apply first, while deductibles from the wraparound plan go into effect afterwards.

In conclusion, most of the variables contained on these NEHIS files have been checked for consistency and validity and when necessary, subjected to data consistency and logical edits. These edits are noted in the Glossary pertinent to each file. However, because of the complexity of the data arising from multiple respondents for a given establishment, the fact that respondents were offered a choice of responses (e.g. per year/ per month, aggregate), and differential item non-response rates, there are plan records with inconsistencies among some variables. These “problematic” variables, for the most part, involve components of annual health insurance costs. Also, BENEPAID (annual benefits paid out by the plan) had a particularly low item response rate of 28 percent.

**NATIONAL EMPLOYER HEALTH INSURANCE SURVEY (NEHIS)  
RDC FILE DOCUMENTATION**

**Table of Contents**

1. Indexes and Codebooks

Establishment File Variable Index .....	i
Plan Analytic and Variance Files Variable Index .....	ii-iv
SENE File Variable Index .....	v-vi
Establishment File variables and codes .....	1 to 40
Plan Analytic and Variance Files variables and codes .....	1 to 82
SENE File variables and codes .....	1 to 52

2. Glossary of variables

Establishment File .....	G-1 to G-13
Plan Analytic and Variance Files .....	G-14 to G-31
SENE File .....	G-32 to G-34

3. Sample SUDAAN procedure statements .....

	S-1 to S-6
--	------------