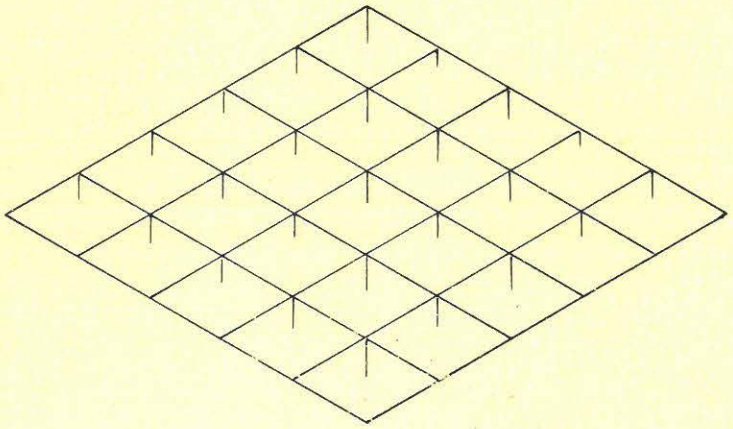
NATIONAL OCCUPATIONAL EXPOSURE SURVEY (NOES)

Contract No. 210–80–0057

Submitted to:
National Institute for Occupational Safety and Health
Robert A. Taft Laboratories
Cincinnati, Ohio 45226



Prepared by:
Robert Hanson, Principal Investigator
Diane Ward, Project Director
John Edmonds, Statistical Analyst
Joseph Escatell, Systems Analyst

Westat, Inc.
1650 Research Boulevard
Rockville, Maryland 20850
(301) 251–1500

November 1983

NATIONAL OCCUPATIONAL EXPOSURE SURVEY (NOES)

Final Report

Contract No. 210-80-0057

Submitted to:

National Institute for Occupational Safety and Health
Robert A. Taft Laboratories
Cincinnati, Ohio 45226

Prepared by:

Robert Hanson, Principal Investigator
Diane Ward, Project Director
John Edmonds, Statistical Analyst
Joseph Escatell, Systems Analyst

Westat, Inc.
1650 Research Boulevard
Rockville, Maryland 20850
(301) 251-1500

November 1983

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

List of Tables

TABLE OF CONTENTS (Continued)

List of Figures

TABLE OF CONTENTS (Continued)

List of Exhibits (Page E-1)

1. INTRODUCTION

1.1 Summary

The National Occupational Exposure Survey (NOES), is a response by the National Institute for Occupational Safety and Health (NIOSH) to the continuing need for information on nation-wide patterns of occupational exposure to health hazards. This report describes that portion of the NOES effort performed by Westat, Inc. under contract with the National Institute for Occupational Safety and Health.

NIOSH is charged with developing information on the types and extent of exposures to occupational health hazards. Data developed by NIOSH have important applications in the establishment of research priorities and in supporting the conduct of that research. To develop data of this type, NIOSH has carried out periodic on-site observational surveys of a sample of facilities representative of selected segments of American workplaces. The National Occupational Hazard Survey (NOHS I), conducted by NIOSH from 1972 to 1974, is an earlier example of this type of survey.

In summary, the objectives of the NOES were:

- To develop estimates of the number of workers in the target population exposed to chemical and physical and biological agents;
- To develop data that could describe the nature and extent of exposures to health hazards and the degree to which facilities have implemented programs to reduce occupational health problems; and

- To compile data such that analysis of industrial hazard exposure trends would be possible by comparison with similar data collected in NOHS I.

The target population was defined as the employees working in facilities or job sites located in the United States reporting more than seven employees with a primary activity or line of business on a list of target SIC's (a copy of this list appears as Exhibit 15). A facility was defined as an economic unit, generally at a single location, where business, service, or industrial activities were performed. Facilities were defined with a focus on operational rather than administrative or management functions.

This report describes the development of the sample design, the construction of the sample frame, and the selection of sample establishments interviewed in the field. The report also describes the method developed to project data from interviewed sample facilities to measurements at the national level utilizing a computer software package called VARNOES. In addition, the report presents key costs and other design parameters observed in the survey which will be useful for the design of future samples.

Staff for collection of data at sample establishments were supplied initially by the NIOSH. Later this responsibility was shared by Westat, beginning in June 1982, approximately halfway through the data collection effort. Editing and coding of the data was performed entirely by the NIOSH.

1.2 Tasks Covered in this Report

The following sections summarize the essential elements involved in each of four major tasks performed under this contract; the last two sections describe two additional tasks.

1.2.1 Sampling Strategy

The NOES sampling plan reflects the experience of the earlier NOHS I program and the assumed conditions that would apply for the new survey. A design comprising two stages of sampling was developed for the study. The first stage involved a sample of 98 first stage geographic sampling units (PSU's) and with the exception of samples of very large facilities, drawn irrespective of geographic location, the interviewed sample was confined to these PSU's. The decision was made to treat those very large facilities (2,500 or more employees) separately in order to maintain more nearly equal selection probabilities than would otherwise be the case. First stage geographic sampling units were defined by a unique system which involved construction of geographic clusters by the computer. The reasoning behind the two-stage sample design and the size of the sample is explained in detail in Section 2 of the report; the construction, stratification, and sampling of the first stage sampling units appears in Section 3; Section 4 describes the sampling procedures within first stage units.

1.2.2 Construction of the Sample

The study universe was comprised of facilities with eight or more employees and conducting business within a specific set of Standard Industrial Classifications (called the "target

SIC's"). The interviewed sample was designated in two steps: (1) a larger sample of 7,362 facilities was selected from the establishment file of Dun and Bradstreet (referred to as DMI) and contacted by telephone to identify those contacted units that were in the scope of the study; and (2) the in-scope facilities were visited and surveyed. A total of 4,504 facilities were designated for field interview, with all but 125 being interviewed. Substitutes were used for 111 of the 125, and the total number of completed interviews was 4,490. Further details of these steps are provided in Sections 4.1 through 4.5 and in Section 5.

1.2.3 Sample Frame

The Dun and Bradstreet Market File for 1980 (DMI) was the source of facilities interviewed for this study. The issues considered in the choice of list and the problems in supplementing the list from other sources are discussed in Section 4.6.

1.2.4 Estimation Procedures

The estimation procedures involved weights defined as the inverse of the probability of selecting the interviewed facility and modified by two stages of ratio estimation. The first stage employs universe counts of establishments in target SIC's in the DMI universe; the second stage uses counts of employees (counts of establishments for plants having over 1,000 employees) given by the County Business Pattern U.S. Summary for 1981, the most recent data available from this source. See Section 6 for details.

1.2.5 Computation of Estimates and Sampling Errors

A computer software package called VARNOES is supplied having two major parts; one part prepares projections to national levels for measures selected by the user involving data items appearing on questionnaires completed at sample facilities. The measures may involve totals, means, medians, or other functions of the characteristics reported at the facility. The second part of VARNOES produces sampling errors for the data items and measures identified in the first part. The estimation procedures are given in Section 6, the sampling error computation methods are discussed, with examples, in Section 7 and instructions for the VARNOES software package are given, with examples, in Section 9.

1.2.6 Future Survey Design Planning

Section 8 presents information useful for preparing a future design with objectives similar to the NOES. The costs and variances experienced for NOES that would be useful to approximate an optimum design with current conditions are compared with similar costs and variances used in developing the NOES. A portion of this section deals with proposals for the control and measurement of interviewer variability.

2. DERIVATION OF THE SAMPLE DESIGN

2.1 Original Design Parameters

2.1.1 Major Design Objectives

Textbooks on sampling describe the methods of optimizing the sample design for a survey. Typically, the methods involve establishing a cost function for the study, expressing the sampling variance as a function of the sample sizes and expected unit variances, and solving the equation which will produce the minimum variance for a fixed cost. This approach would be an oversimplification of the needs for the NOES because it assumes there is a single, unique statistic whose variance is to be minimized. There were several types of statistics for which data were needed for this program and quite different sample designs would be chosen depending on the statistic considered to be of greatest importance.

Much of the analysis in the NOHS I referred to industry-by-industry breakdowns. For these kinds of analyses, the industries should have approximately equal reliability leading to a sample design with roughly equal sample sizes by industry. On the other hand, an efficient sample for analysis of statistics for all industries combined requires the sample size in each industry to be roughly proportionate to the size of the industry.

A second type of problem arose from the interest in data on the distributions of both plants and employees. An efficient sample design for statistics on employees would use higher sampling rates for larger facilities than for smaller ones; for statistics on plants, the number of plants, rather than their size, would be more important in the sample selection.

The sample design developed for NOES has been based on the decision to maximize the reliability of breakdown statistics concerned with numbers of employees. Breakdown by industry or data on the number of firms with specific characteristics were assigned lower priority in developing the sample design.

2.1.2 Cost Function

To determine the sample design that would produce the smallest sampling error for a given budget, we established a cost junction that expressed the total cost as the sum of costs over size strata, with the cost within a size stratum equal to the product of the number of sample establishments and the average cost of interviewing the establishments. Average costs were expressed as number of person-hours observed in the NOHS I study for the size group. As will be described later in this section, the number of PSU's did not enter the cost function. The cost function recognized only the unit costs and the total cost for those aspects of the survey that are directly affected by the sample size. The cost function we used was:

$$C = \sum_a n_a C_a \qquad (1)$$

where

C = The total cost directly related to the sample size,

C_a = The cost per sample establishment in the a^{th} establishment size class, and

n_a = The number of sample establishments selected in the a^{th} size class.

For the NOES the cost function recognized the following conditions.

1. Total Cost

The total of all costs of the survey included a number of more-or-less fixed charges that did not vary directly with moderate changes in the sample size; for example, writing specifications and computer programs for data processing, overhead costs, the cost of hiring and training surveyors, etc. The total of all directly related costs of the sample was taken as the total number of paid person hours to support a proposed number of surveyors working for an expected survey period of two years. This ignored the cost of travel and all fixed costs. It also meant that other costs expected to vary with the sample size were assumed to be small and not important in determining the sample size. For example, we assumed the total cost of telephone screening would not be importantly affected by variations in the allocations of the sample among size classes, or by minor changes in the number of sample cases.

We expected to have 21 surveyors available for the survey period. The surveyors were to be assigned to five teams, including a team leader for each team. We assumed each leader would produce about 80 percent of the production of the other team members (the other 20 percent of the leader's time to be taken up by supervisory responsibilities). Each team member was to work a 40 hour week for 48 weeks of the year; the remaining four weeks to be taken up by annual and sick leave and holiday time. For a 24 month data collection period, the hours contributed by the five team leaders would be:

(5 leaders)x(40 hours)x(48 weeks)x(2 years)x(.8) = 15,360 hours,

and from the remaining 16 team members:

(16 members)x(40)x(48)x(2) = 61,440 hours.

The total for all 21 surveyors = 76,800 person hours.

For a number of reasons, these assumptions turned out to be only a rough approximation of the survey conditions. For example, the period of data collection was about 32 months rather than the predicted 24. The surveyor force began initially with only 11 surveyors, rose to 15 after 5 months, then fluctuated between 10 and 22 for most of the remaining survey period. The size of the field staff meant that surveyor teams did not function as expected. Costs in terms of hours to survey plants for NOES were also found to differ from the NOHS I experience. The impact these changed parameters would have on an optimum survey design is discussed in Section 8.

2. Cost per Sample Establishment

The term C_a in expression (1) is the total number of person hours of surveyor time per establishment in the survey in the a^{th} size class. These unit costs vary according to the size of establishment and were taken from a tabulation of average surveyor hours per establishment by size class experienced in the NOHS I. We assumed the amount of time required to survey a sample firm would be similar to the experience in the NOHS I survey. The actual times observed for the two surveys appear in Table 8-2.

3. Cost per PSU

The cost function used did not have a term to reflect costs that vary by number of PSU's. There were several reasons for this. First, the cost of designating the sample of establishments, a major portion of which would involve the use of the

telephone, would be directly related to the number of sample establishments and would have little relationship to the number or location of the PSU's. Second, the cost of surveyor travel between PSU's was assumed to be relatively small and concentrated during weekends and should not influence the total cost based on person hours during the regular week. Third, there was a desire to restrict the time a team would be assigned a PSU below a 30-day maximum because of per diem reimbursement practices. As these conditions meant the number of PSU's would have little impact on the total survey cost, the reaction was to set the number of sample PSU's as large as administratively feasible to reduce the component of variance arising from the sampling of PSU's. We actually determined the number of PSU's by considering expected size of the PSU workloads and the required length of the team assignment without regard to other per PSU costs. This treatment of the PSU in the cost function also enabled a simpler variance function to be used.

2.1.3 Variance Function

1. Variance of Estimated Totals

The variance for the estimated total used in approximating the optimum design was:

$$\sigma^2(Y') = \sum_a N_a^2 (1/n_a - 1/N_a)\, s^2(\bar{Y}_a) \qquad (2)$$

where

Y' = a total, estimated from the survey

N_a = The total number of establishments in the universe of study in the a^{th} establishment size class (the universe of study is more completely discussed in Sections 2.2.1 and 4).

n_a = The number of establishments in the sample from the a^{th} size class, and

$s^2(\bar{Y}_a)$ = The estimated population variance of the mean of the Y characteristics among the establishments in the a^{th} size class.

2. Variance Among Establishments

The population variance for the a^{th} size class (that is, the variance among establishments within the class) is given by

$$s^2(\bar{Y}_a) = \sum_{i}^{Na} (Y_{ai} - \bar{Y}_a)^2/(N_a - 1) \qquad (3)$$

where

Y_{ai} = The number of employees having the characteristic Y in the i^{th} establishment in the a^{th} size class, and

$\bar{Y}_a$ = $\sum_{i} Y_{ai}/N_a$ is the average number of employees with the characteristic per establishment in the a^{th} class.

Values of $s^2(\bar{Y}_a)$ were not available when the sample design was developed. Although data from NOHS I could have been used to estimate the values of $s^2(\bar{Y}_a)$ for a selected set of characteristics, the time schedule prevented us from waiting for these variances to be prepared. We accordingly employed an approximation that makes use of the assumption that relvariances of characteristics tend to be about the same within most size classes. The relvariance in the a^{th} class is the ratio of the population variance in (3) to the square of the mean for the class. This approximation is

based on experience in other similar surveys and has generally held up pretty well. With this assumption:

$$S^2(\bar{Y}_a)/(\bar{Y}_a)^2 = \text{constant}, \qquad (5)$$

$$S^2(\bar{Y}_a) = \text{constant} \times (\bar{Y}_a)^2$$

expression (5) shows that if the value of the constant is known, an approximation to the variance for the a^{th} size class can be made if the mean for the class is known. However, the assumption of a constant relvariance is weakest in the largest and in the smallest size classes and we assumed the constant in expression (5) should be doubled for these classes (see column (5) of Table 1 in Exhibit 1). Estimated relvariances computed at the completion of the current study allow a review of this assumption, see Section 8.3.2.

3. Variance Among PSU's

The variance expression (2) did not show the contribution to the variance that arose because the sample was restricted to the sample PSU's. In Section 2.1.2, we discussed the omission of the term in the cost function reflecting the number of sample PSU's. With this cost function, the optimum sample design demanded as many sample PSU's as possible (because the cost function showed no cost penalty). However, the number of sample PSU's is important in assessing the variance of estimates from the NOES and the method of preparing estimated variances accounts for this. However, the between PSU variance contributions did not need to appear in the optimum calculations; we were guided only by the desire to minimize the variance between PSU's; i.e., to have as many sample PSU's as administratively feasible.

2.2 Approximating an Optimum Design

2.2.1 Size of the Universe of Study

The cost and variance functions in expressions (1) and (2) include the terms N_a, the total number of establishments in the universe of study in the a^{th} establishment size class. To approximate the optimum, we used the number of establishments shown by tabulations of the County Business Pattern (CBP) summaries for 1979 [1]. The CBP data were taken as the most precise statement of the current number of establishments and employees in the target SIC's at that time even though access to the list of CBP establishments for sampling purposes was not possible.

The study universe was concerned with establishments reporting eight or more employees in a specific set of SIC's (see Exhibit 15); however, the size classes used in the CBP county summary records did not permit the same universe to be defined. To compute the optimum, we used approximations of the CBP counts for the correct size classes by using the proportions of the establishments that appear in the corresponding Dun and Bradstreet size classes (see Exhibit 2). In Section 4, we discuss the problems of defining this study universe in terms of the DMI and other files.

2.2.2 Sampling Rates by Size Class

The formulas that specify the optimum allocation of the sample to the size classes have been derived in Exhibit 1. This

exhibit shows that the optimum sample to be selected in the a^{th} size class is given by:

$$n_a = \frac{N_a s(\bar{Y}_a)}{\Sigma[N_a s(\bar{Y}_a)\sqrt{C_a}]} \times \frac{C}{\sqrt{C_a}} \qquad (6)$$

If the value of $s(\bar{Y}_a)$ that follows from expression (5) is entered in equation (6), the variance constant terms will cancel out of the numerator and denominator so that expression (6) involves only the relative sizes of the variance constants, the means for establishments in the size classes, the values of N_a, and unit costs. As indicated earlier, the CBP was used to produce estimates of N_a. The values of $\bar{Y}_a$ were directly available, although in some cases adjustments were necessary. (See Exhibit 1 for details). The values of C_a are shown in column (2) of Table 8-1 and were estimated from NOHS I.

Because the N_a's were based on the total numbers of CBP establishments, the values of n_a in expression (6) that define the optimum are given in terms of CBP establishments. However, the important result of the optimization computation is to produce the optimum sampling rates n_a/N_a for establishments in the a^{th} size class. These rates were then applied to the DMI file; the actual samples would be expected to differ somewhat from the sample numbers computed from the CBP universe.

A number of other factors also affected the sample sizes. When the computed sampling rates were applied to universe files having duplicate records for establishments or having records for establishments no longer in business, the usual result was a larger initial sample than expected; however, the telephone screening operation eliminated the out-of-business

sample cases and weight adjustments or deletions of sample cases were introduced to correct for multiple listings of the same establishment. In a like manner, when the incomplete file was sampled with these rates, the inadequate coverage was reflected by a corresponding shortage in the number of cases selected. The DMI file was expected to have both the undercoverage and multiple listing problems. The expected number of D & B selections resulting from the optimum sampling rates is given in Table 4-1. The actual telephone sample units processed and facilities surveyed in the field are discussed in Section 5.5.

3. SAMPLING OF PRIMARY SAMPLING UNITS

3.1 Definition of Primary Sampling Units

Sample establishments with expected numbers of employees less than 2,500 were designated within a sample of 98 primary sampling units (PSU's). The system for defining the individual PSU's and their selection probabilities was heavily influenced by the expected organization of the surveyor staff and the number of surveyors expected to be available for conducting interviews at the sample plants.

Original plans called for a staff of 21 trained surveyors available for interviewing in the field over a period of two years. It was assumed this staff would operate in five teams with each team working as a unit in visiting each PSU for the conduct of field interviews (the way these plans were fulfilled is discussed in Section 8). This organization plan and the expected number of surveyors had a direct bearing on the number of establishments and the number of sample PSU's assigned for the survey. The present section assumes the original plan in discussing the decisions made in designing the sample of PSU's.

3.1.1 Distinctive Features of NOES PSU's

The counties in the 50 States and the District of Columbia were combined into 604 PSU's for this survey. The PSU's have the following features.

1. PSU's as Combinations of Counties

The PSU's have been made up of contiguous counties; in Louisiana, parishes, and in Alaska, census divisions, take the

place of counties. Independent cities, where they appear, have been combined with their neighboring counties.

Using the county as the basic building block for PSU's was expected to enable the telephone interviewer, the surveyor, and the respondent to use a generally familiar boundary for the limits of the PSU's. Published data for the County Business Patterns and records for other files of potential use in sampling also record the county location of the establishments as a part of the address information.

2. Metropolitan PSU's

PSU's in metropolitan areas were made up of the counties that composed SMSA's at the time of the 1980 census; however, in a few cases, additional non-metropolitan counties were added to the smaller SMSA's to provide sufficient interviewing workloads.

3. Non-metropolitan (Non-metro) PSU's

PSU's in non-metro areas were made up of groups of counties that had common boundaries. They were prepared by ordering the counties in a contiguous sequence and defining PSU's as individual counties or groups of consecutive counties that provided sufficient interviewer workloads and were large enough so that a self-weighting sample could be selected.

4. State Boundaries and PSU's

Although the intention was to construct PSU's from counties within the same state, multi-state PSU's did occur because some SMSA's appeared in more than one state, and because the assigned sequence of non-metro counties occasionally caused a PSU to include parts from more than one state.

5. Surveyor Workloads

Each PSU was constructed to contain a sufficient number of establishments to keep a four-person surveyor team busy for a period of two to four weeks. All establishments reporting 2,500 or more employees were excluded in determining the expected interviewing workloads or the PSU's. Sample establishments with 2,500 or more employees were selected without regard to their PSU location so that some appeared in non-sample PSU's.

3.1.2 Establish the Size of the PSU's

The need to define PSU's large enough to provide a sufficient workload for an assignment period imposed conditions on the PSU sizes. In this section we show how the conditions were stated in terms of the PSU measures of size and expected person hours of survey time once the approximate number of strata was established.

The design incorporated a self-weighting sample within establishment size classes but with sampling rates differing according to the size of the establishment. A self-weighting sample of establishments for the size class having the smallest sampling fraction was defined using the probabilities that satisfied the following condition (see also Exhibit 1):

$$\frac{1}{k} = \frac{M_{hj}}{M_h} \left(\frac{M_h}{\overline{M}_{hj}} \times \frac{1}{k} \right) \qquad (7)$$

where

M_{hj} = $\sum_a N_{hja} f_a$ is the measure of size of the j^{th} PSU in the h^{th} stratum,

N_{hja} = The number of establishments in the a^{th} employee size class (according to CBP) in the hj^{th} PSU,

f_a = The oversampling ratio for establishments in the a^{th} class, and

M_h = $\sum_j M_{hj}$ is the measure of size of all PSU's in the h^{th} stratum.

The first term on the right of expression (7) is the probability of selecting the sample PSU from its stratum and the remaining term on the right states the probability to be used to select sample establishments, in this case from the establishment size class having the smallest sampling ratio.

For the a^{th} size class, the probabilities in the following expression define the sampling system:

$$\frac{f_a}{k} = \left(\frac{M_{hj}}{M_h} \right) \times \left(\frac{M_h}{M_{hj}} \frac{f_a}{k} \right) \qquad (8)$$

The terms on the right of expression (8) have the same meaning as in (7); the PSU selection probability is the same but the within PSU selection probability in (8) reflects the larger overall sampling fraction f_a/k that applies to the a^{th} class. The within PSU selection probability given in (8) is the basis

of one of two conditions considered in defining PSU's:

- Condition 1 -- The within PSU selection probability should not exceed 1 (so that a self-weighting sample is selected); that is,

$$\frac{M_h \quad f_a}{M_{hj} \quad k} \leq 1$$

From which it follows that the PSU measure of size would have to satisfy:

$$M_{hj} \geq M_h \, f_a \, / \, k \qquad (9)$$

Expression (9) was evaluated by using the values of f_a, k and the total measure of size $M = \sum_h M_h$ computed from the information given in Table 1 of Exhibit 1. We did not use the values of f_a for the employee size classes of 2,500 or more in this computation because firms of this size were not sampled within PSU's. In fact, for a given PSU, a particular f_a was of concern only if the PSU universe contained an establishment in the a^{th} class. Writing $\tilde{f}_{ahj}$ as the value of f_a for the largest class with an establishment in the hj^{th} PSU allowed expression (9) to be replaced by the condition setting a lower bound on the measure of size for the hj^{th} PSU:

$$M_{hj} \geq M_h \tilde{f}_{ahj} \, / \, k \qquad (10)$$

- Condition 2 -- At least two team weeks of effort should be required to survey the sample expected from the PSU.

This condition was expressed algebraically as

$$2(139.7) \leq \frac{M_h}{M_{hj}} \sum_{a=1}^{8} \frac{f_a}{k} N_{ahj} C_a$$

where C_a was the per firm surveyor hours for the a^{th} class. The right side of this expression shows the total surveyor hours in the hj^{th} PSU as the sum of the products of the number of sample firms in the classes and the per firm survey hours needed.

The total number of surveyor hours for a four-person team in a two-week period was expected to be 2(139.7) hours (see Section 3.2.1). From this expression it followed that an upper limit on the PSU measure of size would be given by

$$M_{hj} \leq \frac{M_h}{2(139.7)k} \sum_{a=1}^{8} fa \, N_{ahj} C_a. \qquad (11)$$

Although the conditions could be stated explicitly, it was not always practical to adhere to them rigidly. For example, the measure of size for some PSU's could be made large enough to satisfy Condition 1 only by defining PSU's covering excessive area, and some PSU's had to be defined with measures that did not meet this condition. As a result, special weighting was required for some of the firms in size classes 3 through 8. This problem had to be dealt with in the weighting specifications (see Section III-B of Exhibit 7).

3.1.3 Procedure for Constructing the PSU's

We made extensive use of the computer in defining the PSU's. The five features enumerated in Section 3.1.1 above, were adapted to computer operations by the steps summarized here. The detailed specifications for these operations are given in Exhibits 3 and 4.

1. List the Counties in Contiguous Sequence

This operation produced a sequentially numbered list of all counties, parishes, census divisions, and independent cities within the United States in an order corresponding to their adjacent locations on the ground. The list was prepared by manually assigning sequence numbers to the counties on a series of maps. The ordering tried to minimize "cross-overs" from one side of a significant geographical feature to the other and from a state to its neighbors. Particular care was taken to minimize cross-overs from one Census Region to another.

2. Define SMSA's as PSU's

All counties within each SMSA were treated as a single PSU. In a few instances, one or more adjacent non-metro counties were added to smaller SMSA's to improve the conformity with the two conditions.

Although each of the very large SMSA's was treated as a single PSU in the sampling operation, the interviewing was occasionally apportioned to more than one team for survey at different times during the field interviewing phase. For example, it was convenient to survey one-half of the Chicago SMSA by all

of the interviewers available at the start of the survey and complete the remaining portion of the PSU sample later as a separate assignment. The order of surveying PSU's was decided by the NIOSH staff.

3. Combine Non-Metro Counties into PSU's

After ordering the counties by sequence number, the computer defined PSU's in non-metro areas by combining them until the two conditions of Section 3.1.2 were satisfied. A visual inspection of the computer results looked for evidence of awkward geographical combinations that would make them inappropriate assignment areas. We resequenced the counties in a few states and established a revised set of PSU's. A few PSU's of very large area were generated by the computer in some of the Western States. Later, when one of these large PSU's (in Alaska) was identified as a sample PSU, we selected a subsample of the PSU area to permit manageable travel patterns even though this was expected to reduce the workload size below the two-week level.

3.2 Stratification of PSU's

The 604 PSU's that encompass the entire area of the 50 States and the District of Columbia were grouped into strata by the process described in this section. Statistical theory shows that the efficiency of a stratified design is improved by defining strata of approximately equal size and such that the PSU's within the strata are as homogeneous as possible with respect to the important statistics to be estimated from the survey. Homogeneity of PSU's within strata can be improved by using strata as groups of PSU's with similar economic structure. This generally should reduce the variance among PSU's within each stratum for the statistics of interest to the NOES.

The disparity in size of the PSU's interfered with establishing strata of equal sizes because some PSU's were larger than the desired average stratum size. The largest of these PSU's were defined as separate strata and the remaining PSU's were grouped into strata of approximately equal size.

As shown in this section, the process produced a total of 98 strata. Twenty-six of these strata contain only one large PSU; these strata are called self-representing (SR) because the single PSU represents itself in the sample. The remaining 578 PSU's (604 minus 26) were grouped into 72 nonself-representing (NSR) strata having about equal measures of size; the term NSR is applied to these strata because one PSU is selected to represent all other PSU's in the stratum in addition to itself. The reasons for choosing this number of strata are given below.

3.2.1 Number of Strata

The original design requirement that PSU's should provide an interviewing assignment of two to four weeks for a four person team was an important consideration in determining the total number of assignments and the number of sample PSU's in the design. The number of selections and the average interview time per establishment, shown in Table 1 of Exhibit 1, showed the expected number of surveyor hours for establishments with 2,500 or more employees to be about 37 percent of the total surveyor workload. As these large cases were to be interviewed without regard to their PSU location, they did not influence the number of sample PSU's. We were concerned, therefore, with the roughly 63 percent of the 76,800 hours (about 48,400 person hours) required to survey establishments with less than 2,500 employees within the sample PSU's.

Each four person team was expected to produce 139.7 hours of productive surveying each week; this number was derived as follows:

		Hours Per Week
Supervision (40 hours x .2)	=	8.0
Leave (4 persons x 40 hours x 4/52 fraction of weeks in leave status)	=	12.3
Investigation (remaining hours of week)	=	139.7
Total (4 persons x 40 hours)	=	160.0

The total number of team weeks for surveying establishments of less than 2,500 employees would therefore approximate:

$48{,}400/139.7 \doteq 345$ team weeks

We assumed our attempts to establish PSU's having workloads of two to four weeks would result in an average workload of three team weeks over all PSU's. Then an approximate size for each stratum would be 3 weeks out of 345 (or about 1 in 115) of the total survey workload for establishments of less than 2,500 employees. This size condition could be used only to define NSR strata. The physical size of the larger SMSA's was the determining factor in defining the SR PSU's. Application of this condition resulted in defining 98 sample PSU's; 26 of these were self-representi and the remaining represented a sample of one PSU from each of 72 NSR strata.

3.2.2 Definition of Strata

A primary goal of stratification was to assign PSU's to groups that would be relatively homogeneous with respect to the statistics of interest for the survey. In practice, the procedure used was to identify groups of PSU's with significant concentrations of employees in certain key target industries that were likely to have serious and common health hazards. This worked fairly well for most of the small PSU's. However, for PSU's which contained a wide range of target industries, particularly in the larger size classes, it was not always possible to produce strata that were homogeneous in this regard. For this reason, this mode of stratification was only partially effective. Additional stratification criteria, other than employee concentration by SIC, were therefore used, such as metro vs. non-metro area, concentration of employees in large firms, and geography. The information assembled for each PSU for the stratification operation is given in detail in Exhibits 2, 3, and 4.

The complete file of data on each PSU in the frame contains counts of firms in various size classes, as provided by County Business Patterns for 1978. Aggregations were also made from this data source showing the number of employees in 21 subsets of the target industries with the subsets defined as combinations of SIC's. Additional codes were assigned for each PSU to indicate whether it was an SMSA, and the region of the country. The computer was used to group PSU's and display the distribution of PSU's and their measures of size across the following variables:

- Proportion of employees in target firms coded in manufacturing SIC's;

- Proportion of employees in firms within the PSU falling in the largest size classes;
- Concentration of employees in the petroleum and/or chemical, rubber, leather industries;
- Geography - Census region; and
- SMSA or Non-SMSA.

The above listing reflected the order of importance each variable was accorded in the formation of groups and strata. As a first step, large groups were formed comprising PSU's that were similar with respect to the employees in manufacturing and large firms. If possible, the employees in petroleum, chemical, rubber, and leather were also similarly concentrated. The measure of size in each large group determined the number of strata that should be produced from the group. If two or more strata were to be constructed, the PSU's were sorted by the five variables in the order listed above and then divided into strata, based upon total measure of size and the similarity of PSU's for each variable.

3.3 Selection of Sample PSU's

When the strata were defined, a master file was prepared listing the PSU's by stratum and showing the PSU measure of size. Prior to sampling, the strata were compared to locate pairs of strata that were composed of roughly similar PSU's. The pairing of strata was significant for mathematical purposes, as the computation of variances for the survey statistics was to employ paired stratum methods.

A sample of one PSU was selected from each stratum with the selection probability for each PSU equal to the proportion of the stratum measure of size contributed by that PSU. The selection was carried out using a random number table. Exhibit 5 shows the composition of all PSU's selected for the NOES. Also shown is the contribution to the final inflation weight attributable to the PSU selection probability. Parts of 40 States and the District of Columbia appear among the sample PSU's.

4. SAMPLING ESTABLISHMENTS

4.1 Introduction

The designation of sample facilities for field interview was performed in two steps: A sample of establishments from the DMI file was selected, and then the selected sample was interviewed by telephone to identify the facilities to be visited by the surveyors (details of this second step are contained in Section 5).

The Dun and Bradstreet (D&B) 1980 Market File, supplemented by units discovered during the telephone screening operation, served as the source for the facilities surveyed. As described in Section 4.6, other lists were considered for supplementing the DMI file, but were not used.

4.2 Composition of the Establishment Sample

The sampling rates computed from expression (6) in Section 2.2.2 apply to a design that would be feasible under the several assumptions used. The assumptions, spelled out in Section 2.1, approximated the actual survey conditions expected to apply for the NOES. However, we did not expect the assumptions to apply in every instance. For example, it was unrealistic to expect that teams would be equally proficient at all times, non-interview problems would not appear, the interviewer force would remain constant over the survey period, etc.

Some flexibility in the design was included in the planning to prepare for problems arising during the course of investigating the sample. To this end, the sample selected was subdivided into a number of random subsamples separately identified. The detailed specifications for these subsamples are presented in Exhibit 6.

4.2.1 The Initial Sample

Sampling rates for a design determined by the methods discussed in Section 2.2 were the basis of the sampling process used. The method of determining sampling rates for a design that is reasonably close to an optimum under the assumed conditions is given in Exhibit 1.

An initial sample of establishments was selected by a procedure equivalent to arranging PSU's in a convenient order, stratifying DMI establishments by size class within each sample PSU, sequencing the establishments by 4-digit SIC within size class within PSU, and designating a systematic sample of establishments across all sample PSU's independently within each size class. The procedure was intended to minimize the variability in the number of cases selected from the DMI so that the actual number of sample establishments selected in any of the size classes should be close to the expected number in the class at the national level.

If the same sampling procedure and sampling rates could be applied to the Census Bureau file of County Business Pattern establishment records for 1979, it would have produced a total of approximately 4,895 sample establishments with the sample distributed by size class as shown in column (3) of Table 4-1. The

Table 4-1. Sampling rates and expected number of establishments by size class in initial, screening, and shadow samples

Size class	Reported number of employees	From CBP[1] Initial	From DMI Listings Initial	Screening	Shadow plus screening	Over-sampling ratio fa[3]
(1)	(2)	(3)	(4)	(5)	(6)	(7)
11	N/A	-	148	185	371	1.0
1	8 - 19	1,190	1,393	1,742	3,483	1.0
2	20 - 49	914	1,073	1,341	2,681	1.593
3	50 - 99	675	785	981	1,961	3.022
4	100 - 249	838	1,003	1,253	2,507	5.464
5	250 - 499	512	604	755	1,510	9.384
6	500 - 999	344	409	511	1,023	13.576
7	1,000 - 1,499	123	163	204	407	17.235
8	1,500 - 2,499	108	142	177	355	23.785
9	2,500 - 4,999	94	124	155	309	35.984
10	5,000 and over	97 -	139 -	174 -	- 261[5]	91.110 68.330[5]
Total expected sample establishments		4,895	5,983	7,478	14,868	
Sampling interval[3] k =		k_1=199.53	k_1=199.53	k_2=159.62[4]	k_3=79.81	

[1]Expected total sample at the U.S. level assuming it could be selected from a file of CBP establishment records for 1979.

[2]Expected number of selections from the 1980 DMI file before eliminating duplications and out-of-scope cases.

[3]Sampling rate for the a^{th} size class is fa/k.

[4]For the Chicago PSU only, k_1 = 212.13, k_2 = 169.7 and k_3 = 84.85 for the initial screening, and the screening plus shadow samples respectively.

sampling rate shown in the table for the a^{th} size class is f_a/k_1; for example, for size class 7, the optimum rate would be 17.235/199.53. The same sampling rates, if applied to the 1980 DMI, would produce a total of 5,983 selections distributed by size class as given in column (4) of Table 4-1.

4.2.2 Subsamples for Workload Control

The schedule for surveying sample establishments within PSU's was based on a predicted length of stay for the team to complete the PSU assignment. The length of stay was determined by calculating the PSU person hour needs for surveying the expected sample in the PSU using the average times for survey per establishment observed during NOHS I. These calculations, prepared for each PSU, were used to establish a tentative schedule for covering all PSU's during the expected two-year period for field work. To maintain such a system, each team would have to finish each of its PSU assignments in the allotted time. Obviously, the time per establishment was not expected to be identical in all PSU's. Since the period of time that could be spent in a PSU was fixed, variable workloads were necessary for an efficient field operation. The sampling plan provided for such variable workloads.

To assure that such a plan could be implemented, we used the following system:

- The initial sample was expanded by 25 percent to an expected total of about 7,500 DMI establishments. This sample is called the "screening sample" because it comprises the firms to be screened by the telephone operation if the entire expanded sample was to be surveyed.

- Four random subsamples of the expanded sample were designated as follows:

  Subsample A = 1/2 of the expanded sample,
  Subsample B = 1/4 of the expanded sample,
  Subsample C = 1/8 of the expanded sample,
  Subsample D = 1/8 of the expanded sample.

- The team was assigned a portion of the expanded workload at the beginning of its scheduled stay in the PSU (e.g., subsamples A and C). The portions were chosen such that, over all PSU's, the total sample interviewed would approximate the number computed as the initial sample. The team was expected to complete all of the workload comprising the initially assigned subsamples for the PSU.

- With the completion of the initial assignment in the PSU, the team supervisor assigned additional subsamples where possible. All additional subsamples assigned had to be completed in the time originally fixed as the length of stay for the PSU.

This system produced a probability sample of survey firms for each PSU and allowed supervisors to react to field problems. These problems included the additional responsibility for samples of establishments with 2,500 or more employees appearing in or near the PSU, unusual concentrations of problem interviews, and surveyor staff shortages because of illness or vacation, etc.

4.2.3 Shadow Samples

Each of the establishments selected for the survey had a reserve sample establishment selected with it for use in replacing the attrition due to nonresponse. The collection of reserve establishments was called the shadow sample and was used as follows:

- If all efforts of the telephone interviewer, the surveyor, and the team leader did not succeed in obtaining cooperation at an in scope establishment, a reserve shadow was used as a substitute for the non-cooperating case. The reserve was used as a substitute for only those original in scope sample establishments currently in business.
- If the substituted shadow was found to be out of scope, or refused to cooperate, the original sample unit was retained in sample and a court order (warrant) was obtained to secure cooperation from the originally designated unit.
- Original sample establishments found at the time of the survey to be out of business, or not doing business in any of the target SIC's, were treated as out of scope and shadows were not substituted for them.

Preparatory work for an establishment in the reserve sample began only after attempted investigation in the field showed the original sample would be a non-interview. Usually the telephone screening interviews, and the contacts made by NIOSH field staff to arrange the survey visits, did not begin until after other units in the areas were surveyed. As a result, interview dates for the shadow cases fell behind other regular interview samples for the PSU.

For all size classes, except those reporting 5,000 or more employees on the D&B file, the screening sample and its reserve were designated in one operation by doubling the screening sample rate (that is, by using sampling intervals equal to half the intervals needed for the screening sample alone) and assigning alternate selections to the screening and shadow samples. The method of assigning reserve samples for D&B units showing 5,000 or more employees is discussed in Section 4.3.2.

4.2.4 Number of Sample Establishments

The number of selections in the initial sample and the expected screening and shadow samples of DMI records are summarized in Table 4-1. The table also shows the selections that would result if the optimum sampling rates could be applied to facilities in the 1980 County Business Pattern file.

The proportion of establishments to be selected in the sample in each of the size classes is determined by dividing the oversampling ratio for the class (shown as f_a in column (7) of the table) by the sampling interval for the sample (shown as k) in the table. For example, in column (3), the expected 838 selections that would result when sampling in size class 4 from the CBP file have been computed by multiplying the universe total number of CBP establishments in the target SIC's in this size class by the ratio f_a/k_1 where $f_a = 5.464$ for the class and $k_1 = 199.53$ for the optimum sample. The universe total number of CBP establishments in this size class (30,601) appears in Exhibit 1, Table 1.

The values of f_a and k were computed by the methods discussed in Section 2.2.2 and, except for classes 9 and 10, were based on a tabulation of the CBP establishment counts in the NOES target SIC's for each size class. An earlier set of CBP counts established by a clerical procedure was used to derive the sampling rates for DMI classes 9 and 10 before the more precise sampling rates based on the final CBP counts became available. For this reason, the values of f_a shown in the table for these two classes are somewhat different than would result if more precise 1980 CBP universe counts were used.

Table 4-1, column (4) shows the sample size designated from DMI, using sampling rates computed for the initial sample. The screening sample (column 5) is 25 percent larger than this optimum and occurs by reducing the optimum sampling interval, k_1, to $k_2 = (.8) * k_1$ for the screening sample.

The shadow and the screening samples were selected in one operation by using sampling intervals of

$$k_3 = (.5) * k_2 = (.5) * (.8) * k_1$$

except for size class 10; for this class, the screening sample interval was reduced to $k_2 = (2/3) * k_1$ for the reason discussed in Section 4.3.2.

4.3 Samples of Establishments Having 2,500 or More Employees

4.3.1 Selection Without Regard to PSU

The proportion to be selected into the sample from size classes 9 and 10 was so large that sample efficiency would have been impaired if confined to the sample PSU's. For example, the probability for the screening sample for establishments with 5,000 or more employees is:

$$f_a/k_2 = 91.100/159.62 = 1/1.752$$

The overall selection probability would have to be the product of the probabilities of selecting the PSU from its stratum and of selecting facilities within the PSU. (See expression (8) in Section 3.1.2.) From this it would follow that unless every sample PSU comprised at least 1/1.752 (about 57 percent or more) of the total measure of size of its stratum, it would not be possible to select a self-weighting sample. In a PSU with a measure of size less than 57 percent of its stratum measure, every establishment in size class 10 would have to be in sample and would also require an extra weight to provide unbiased estimates. The problem would be similar, although not as severe, for establishments in class 9 (2,500-4,999 employees).

We therefore ignored PSU's in selecting the sample for the two largest size classes. Many of these large firms were located in a sample PSU, or nearby, and were surveyed by a team associated with a convenient PSU. Additional travel was of course necessary in some instances.

4.3.2 Special Treatment of the Shadow Samples

The shadow sample for the class of largest establishments was designated in a somewhat different manner than for the smaller firms.

The screening sample for this size class was such a large proportion of the universe that a smaller reserve sample for shadows had to be used. The reserve for this class was half the size of the screening sample. The screening sample and the reserve for this class were selected in one operation with a systematic sampling interval 2/3 of the interval needed for the screening sample alone. Thus, two sample establishments from size class 10, adjacent on the listing of sample selections, were followed by a single reserve establishment to be shared by the two sample establishments preceding it.

4.4 Samples of Establishments with 8 to 2,499 Employees

The universe of establishments listed by DMI that reported employees in the range of 8 to 2,499 was separated into 8 size classes, each class with its own selection probability. The design restricted sample establishments from these size classes to the sample PSU's. To achieve an equal-probability sample among all firms within each size class, the selection probabilities had to take into account the probability of selecting the sample PSU from its stratum.

The selection probabilities conform to the formula presented as expression (8) in Section 3.1.2. The first term on the right of that expression represents the probability of selecting the PSU from its stratum and the second term shows the

probability of selecting establishments from the a^{th} size class within the PSU. The terms f_a are the oversampling factors that apply in the a^{th} size class. The factors f_a are constant for all PSU's so the complete set of selection probabilities is defined when the overall probability of 1/k for the class of smallest establishments is given, the constant values of f_a are specified and the PSU selection probabilities are known.

The PSU selection probabilities are given in Exhibit 5. The overall probability actually used to select the screening and reserve samples in one operation for the class of smallest firms was 1/k = 1/79.81.

In self-representing PSU's, the PSU selection probabilities are 1.0 so that sample establishments within these PSU's are designated by the overall selection probability of f_a/k. In non-self-representing PSU's, the probability of selection for establishments within the PSU is given by the formula:

$$(f_a/k) * (M_h/M_{hj})$$

where the second ratio is the inverse of the probability of selecting the PSU from its stratum.

4.5 Samples of Establishments with Fewer than Eight Employees

An early decision confined the sample from the DMI lists to those reporting eight or more employees and operating within the set of target SIC's. As the DMI employee reports were not current, rejecting facilities with seven or fewer employees as out of scope could introduce some error.

To obtain a measure of the error, we selected a sample of 200 establishments reporting seven or fewer employees for screening by the telephone operation. The objective of the screening was to determine the current number of employees and whether current activities were within the target SIC's. Cases found to have eight or more employees were not scheduled for further survey work.

Table 4-2 shows that telephone interviews of 200 establishments reporting 7 or fewer employees according to DMI uncovered 11 cases actually having 8 or more and operating within the target SIC's. This suggests a loss of about 5.5 percent in coverage of these small establishments could occur. However, the ratio estimation procedure adjusts NOES levels to be consistent with the levels of the 1981 CBP; this will reduce (although not entirely eliminate) the coverage bias from this source. Some of the five refusals shown in the table could increase the percentage.

4.6 Universe Lists

4.6.1 Introduction

In planning the survey it was understood that a current and complete listing of all establishments in the universe of study did not exist and could not be constructed with the resources and time available. Although a number of commercial lists were available, each had its weaknesses as a sole source listing. It was hoped some advantage could be realized by improving the coverage of the best of the lists through supplementary samples selected from other lists and this system was

Table 4-2. Telephone interviews of 200 establishments reporting seven or fewer employees on the 1980 DMI file

Telephone interview reports	Number
Total	200
Non-working phone	50
No answer	3
Out of business	6
Under 8 employees	123
Non target SIC	2
Refusal	5
Reported 8 or more employees	11

explored. However, the sample for this survey ended up being selected entirely from one commercial list, the Dun and Bradstreet Market file of 1980; this file is referred to in this report as the DMI or the D&B file.

This section discusses the reasons for choosing the D&B file as the initial source of the universe, our investigation of the coverage of this list and the decision to forgo supplementing this list by samples from other commercial sources.

4.6.2 The Dun and Bradstreet List

The Dun and Bradstreet list is a well-known and widely used industrial directory service. Listings generally enter the file on the basis of need to establish credit ratings. However, some listings are initiated as a result of other needs. For example, Federal agencies are required to ask contractors to supply "Dun's numbers" to agency contract officers. This is a new requirement (since October 1978) of the Federal Procurement Data System. Such contractors would automatically be added to the D&B file if not previously included. We were also informed that D&B had made a special effort to complete listings of large manufacturing establishments.

We decided to use the D&B commercial list as the basic source for the universe listing, primarily because coverage of firms appears to be quite good and in most cases adequate for the set of target SIC's. The D&B file included more than 500,000 company and establishment listings in the target SIC's. Coverage was expected to be relatively good in manufacturing, less complete in some other industries. However, we anticipated some problems in using this file because of the inclusion of erroneous or obsolete entries. We handled them by removing the selections of

obsolete entries from the sample as part of a screening process.

4.6.3 Coverage Evaluation of D&B Lists

The original proposal suggested that sampling establishments from a file covering 90 percent or more of the target universe would be adequate for the study purposes. For establishment groups that did not meet this criterion, we considered supplementing the D&B list. We adopted a further condition that supplementation would not be considered unless the problem group of establishments comprised at least 0.5 percent or so of the total 29,000,000 employees in all target establishments.

The adequacy of the D&B file was examined by comparing the total number of employees reported for target firms listed on D&B with corresponding totals from the CBP. We used the Census Bureau's 1977 County Business Pattern publication as the most authoritative source of the number of employees and establishments by size class in each of the target SIC's [1]. There were several problems in comparing CBP and D&B tabulations that had to be considered in interpreting the evaluation system used.

- The two files did not refer to the same time periods; CBP tabulations were for 1977 with establishment size classes in most cases based on the number of employees reported as of mid-March 1977. The DMI file was labelled "1980" with number of employees as carried on the most recent D&B record. Therefore, complete agreement among the files for total employment for an SIC group could not be expected even if the D&B list was currently complete.

- Establishments in scope for the study were confined to firms with eight or more employees. However, the CBP tabulations did not provide counts for the necessary establishment size classes so that approximations were required.

- SIC coding for establishments was probably not consistent for the two files. For this reason, comparisons were made initially at the 2-digit SIC levels. Where serious differences appeared at the 2-digit level, the examination progressed to 2- and 4-digit levels. This assumed coding inconsistencies would be more evident at the detailed SIC levels.

- CBP files exclude government employees, self-employed persons, farm workers, employees under the Railroad Retirement Act, and domestic service workers. About 24 percent of the total paid civilian wage and salary employment did not appear in the CBP tabulations. The absence of the self-employed was not considered a problem as we assumed them to be concentrated among firms too small to be in scope. The absence of the other categories may have accounted for some of the observed differences for the target SIC's.

Coverage of government workers in D&B was not clear although a few government installations were noted on the D&B universe lists. In some situations, we tried to evaluate D&B using counts of employees on nonagricultural payrolls by industry as given in Employment and Earnings; these figures referred to essentially the same group of employees as the CBP except that civilian government workers were included [3].

Establishments in the following SIC groups comprised a significant portion of the target universe and were found to have D&B to CBP employee ratios of less than 0.9:

- 451 and 452 - Air transportation.
- 481 - Telephone communication,
- 491 - Electronic services,
- 493 - Combination electric, gas and other services combined,
- 5541 - Gasoline service stations,
- 7231 - Beauty shops,
- 7241 - Barber shops,
- 7299 - Miscellaneous personal services.

4.6.4 Consideration of Supplementing the D&B Lists

We considered supplementing coverage of firms in SIC groups listed above by use of a second commercial list, the National Business List (NBL). However, this source could not provide the number of employees for each firm; such information would have to be determined by telephone contact at each firm selected in the supplemental sample. For the reasons given below, NIOSH decided to accept the coverage provided by the D&B files without supplementation.

In the case of SIC 5541 for example, the NBL could have supplied a list of about 126,000 service stations that were not supposed to be on the D&B list. Our sample from this additional group would have been about 790 cases which would have to be contacted by phone to screen out those with less than 8 employees; an expected 74 of these would have 8 or more employees and therefore be in scope (assuming all were still in business). The cost of adding the 74 additional cases to the sample would have been roughly $35 per case not counting the cost of telephone screening of the 790 units and the field interview cost of the 74 units.

For firms in the last three groups listed in Section 4.6.3, the sample supplementation selected from NBL would also have to be matched against the D&B universe listing to remove those already having a chance of selection from that first source. This would have been required because NBL lists for these firms would include some already on D&B.

For all of the other SIC groups listed in Section 4.6.3, the NBL lists were constructed from essentially the same sources as the D&B and were not expected to be of much help in improved coverage. We could have undertaken supplementation by performing a search for firms appearing in phone directory yellow pages for the localities in the sample PSU's. This project was considered beyond the level of NIOSH resources.

4.6.5 Establishments with Hazardous Occupations

During early survey planning, we considered oversampling establishments with employees in particularly hazardous occupations. By identifying such plants in advance and sampling them at higher rates, the reliability of estimates for hazards for all plants could be increased. For this system to be efficient, however, a relatively small subset of the universe with markedly greater concentration of risk would have to be identified.

For example, if a subset of plants (say less than 10% of all plants) could be identified as having hazard exposure rates nine times the rate of the remaining plants, it would be possible to reduce the sampling error on the estimated hazard rate for all plants by as much as 10 percent. This would require substantial oversampling of the high hazard group.

This procedure was not adopted, partly because of problems in identifying high hazard establishments and because the sample selected to optimize estimates for one hazard would also be used when estimating other characteristics for which the oversampling could be a disadvantage.

5. COMPOSITION OF THE FIELD INTERVIEW SAMPLE

5.1 Introduction

The following procedures were used to identify the establishments surveyed for the NOES.

- Sample establishments were designated from the universe file. The full ABCD sample and all potential shadows were designated in one operation for each PSU in advance of the scheduled field survey date for the PSU.
- The subsamples (ABCD) to be surveyed were determined for the PSU and telephone screening was carried out for the sample units. The units surviving screening were transmitted to NIOSH to be assigned to a survey team. Out of scope units identified during telephone screening were dropped from the survey.
- The field surveyors contacted each of the units to schedule the appointment for interview.
- A field surveyor then visited each establishment, made final determination of survey eligibility, and surveyed the establishment. Units determined to be out of scope at the time of the field survey were dropped from the study.

5.2 Assign Workload Subsamples

The D&B listings processed to determine the establishments for survey in the field comprised (as a minimum) the ABC workload subsamples in each of the 98 NOES sample PSU's and the full set of ABCD workload subsamples in size classes 9 and 10. In roughly half of the PSU's it was possible to use the full sample by including subsample D.

Table 4-1 shows both the overall sampling rate computed for the optimum (1/199.53) and the rate for the full ABCD samples (1/159.62).

Table 4-1, column (5), shows that if the ABCD sample had been used in all PSU's and size classes, it would produce an expected sample of 7,478 D&B selections. As discussed in Section 5.5, the total number of D&B listings actually selected for the NOES amounted to 7,167, a number between the expected designations given in columns (4) and (5) of Table 4-1.

5.3 Telephone Screening

Each establishment in the sample was contacted by telephone to verify (or correct if necessary) some basic information that we already had from the DMI and to obtain some additional information. Exhibit 10 is a copy of the questionnaire used for this screening.

In general, the following information was verified (or corrected):

- The establishment name,
- The street address, and
- The Standard Industrial Classification (SIC).

Additional information obtained included:

- Information on any other establishments owned or managed in the PSU,

- The name, title and telephone number of a knowledgeable contact person in each establishment, and

- The names of any unions at the establishment.

5.3.1 Screening Procedures and Results

The screening work was done by telephone from the Westat Telephone Center in Rockville, Maryland. A screening questionnaire was completed for each sample establishment by trained interviewers, most with experience on other related telephone surveys of commercial firms. The manual used in training the interviewers for this survey is available as an attachment to this report.

The interviewers were instructed to contact each establishment, verify that they were in contact with the intended establishment, explain the purpose of the survey, describe the further information to be requested by NIOSH, and request the cooperation of the firm. Written authorization for the survey was provided for firms that requested it.

Firms that refused to cooperate with the telephone interview were retained in the survey and field interviews were later scheduled for them.

The telephone interviewers were also instructed to verify the current address of the establishment, obtain an update on the number of employees, and determine whether or not the business carried out by the firm fell within the SIC industries targeted for the survey.

Firms that were conducting business in SIC's outside of the scope of the study were eliminated from the survey. However, those temporarily out of business or temporarily having fewer than eight employees were retained in the survey. Final determination of the in-business status and firm size was made at the time of the field interview.

The interviewer asked if the sample firm owned or managed other firms, and obtained names, addresses and contact persons for any owned or managed firms. In addition, the interviewer probed for information on whether the DMI listing for the firm could refer to more than one plant, or more than one facility under the same company name. The interviewer also obtained the names and phone numbers of company and union representatives to be contacted when the surveyors commenced field work.

Except for establishments selected from the strata of units reporting 2,500 or more employees, locations outside the sample PSU were dropped from the survey.

Approximately 7,400 establishments were screened. Of these only 3 percent refused to cooperate over the phone and 66 percent were determined to be in scope. Detailed screening results are presented in Table 5-1.

5.3.2 Use of Screening Information for Sample Modification

An alphabetized list of all firms on the universe file within each PSU was prepared as a tool in the screening process. Each firm in the sample was compared with the list for its PSU to determine if duplicate records for the sample firm were present

under the name selected or under one of the alternate names discovered during the telephone interview. Where duplication was noted, the weighting of the firm was adjusted to reflect its true selection probability.

An intent in the design of the survey was to consider sample firms as single plants or locations. However, a company could operate in more than one location or be composed of several plants or branches and be listed only once on DMI with a single address and employee total. Some of these branches, not listed on DMI, were located in the screening process. During screening, therefore, sample firms in size classes 1 through 8, and 11 were asked whether they owned or managed other firms in the PSU, or operated at other locations. If other firms were owned or managed, the identity and the size of these firms was recorded. The alphabetic list was then searched to determine if the new location should be treated as an addition to the sample frame.

New firms reported as managed by a sample establishment and found on the universe listing were dropped because their presence on the listing meant they already had their proper chance of selection. New firms not appearing on the list were given a chance of selection in the interview sample. This was accomplished by means of a worksheet designed to select new firms with probabilities reflecting their chance of selection had they been originally recorded on the universe listing.

5.4 Procedures When Telephoning Screening was Completed

Screening assignments were made in PSU units just prior to the planned time of field interviews for the PSU. In addition

to an allowance for the time to conduct the telephone screening, a period of about one month was set aside for the NIOSH staff to mail advance letters to the establishments and to assign the PSU to a field team.

To further these steps, at the close of the screening operation in Rockville, Westat provided the NIOSH staff the following information for each establishment requiring a field interview:

- The Westat case identification number,
- Name and address of the establishment,
- Name and title of contact person,
- SIC as reported by the establishment during screening,
- Name, address and contact person(s) for each union at the plant, and
- A six digit NIOSH survey control number.

This file was transmitted to NIOSH for use in preparing mailings to the sample establishments.

5.5 Telephone and Field Interview Samples

This section describes the process used to produce a total of 4,490 completed interviews. This process is most easily explained by referring to Table 5-1. Table 5-1 presents the construction of the telephone and field interview samples. Since column (1) of the table lists the totals of the telephone screening,

Table 5-1. Construction of Telephone Field Interview Samples NOES

Counts	Telephone Screening			Interviews Assigned		
	Total (1)	Initial (2)	Added (3)	Total (4)	COMPL. (5)	NI (6)
Telephone Screening Interviews	7392	7167	225			
Out of Scope	2542	2535	7a			
Non Working Phone	683	682	1			
Out of Business	232	230	2			
Less Than 8 Employees	979	978	1			
Non Target SIC	229	229	-			
Gov't & Administrative Office	368	365	3			
Out of PSU	51	51	-			
In Scope	4850	4632	218	4850		
Refusals	221	221	-			
Other In Scope	4629	4543	218			
Field Operations		4632	218	4850		
Out of Scope		339	7a	346e		
Out of Business		64	2	66		
Less Than 8 Employees		186	1	187		
Non Target SIC		21	-	21		
Government		11	2	13		
Administrative Office		39	1	40		
Workload Subsamples		18	1	19		
In Scope		4293	211c	4504	4490	14f
Cooperators		4293	86	4379	4367	12
Subsampled Plants			86b	86	86	-
Other Cooperators			-	4293	4281	12
Refusals			125d	125	123	2
Shadows			113	113	111	2
Warrants			12	12	12	-

a Out of scope portion of a subsample of 93 multi-facility establishments discovered during telephone screening
b In scope portion of the multi facility subsample
c Additional screening from refusals and subsample cases.
d Additional screening to replace refusals.
e Includes all out of scope cases from categories a and d.
f Could not be completed during survey period.

it is more informative to begin by discussing entries to column (2).

The first entry in column (2) shows that an initial selection of 7,167 establishments was designated from the DMI for telephone screening interviews. This is the number of DMI selections in the workload subsamples actually assigned. Column (2) also shows that 2,535 of these selections were eliminated as out of scope in the telephone operation.

The 4,632 in scope telephone selections remaining in column (2) were eligible for field interview; they constitute the major portion of the 4,850 establishments, see column (4), assigned to field surveyors. The 218 "Added" cases will be discussed later.

Column (4) shows the results of field work for the 4,850 plants. A total of 346 of the 4,850 selections, when contacted in the field, were found to be out of scope which left 4,504 establishments which required field interviews. While 4,379 of these establishments cooperated when contacted by the surveyors, 125 refused. Column (6) reports that 14 of the 4,504 expected interviews were not completed leaving the 4,490 interviews shown in column (5). There are 4,490 completed NOES questionnaires.

The 125 refusals in column (4) were handled by identifying 125 shadows and contacting them by telephone, accounting for the 125 additional contacts shown in column (3). Twelve of the 125 could not be accepted as substitutes, so warrants were obtained to require cooperation from the 12 refusals.

During telephone screening, a number of establishments were encountered which had several plants operating in the same PSU. These additional locations did not appear on the DMI. An additional sample of 93 plants was selected from this group and included in the telephone samples as shown in column (3). Of the 93, 7 were found to be out of scope and the remaining 86 were added to the field interviews.

6. ESTIMATION PROCEDURES

6.1 Introduction

A probability of selection can be associated with each of the steps undertaken in designating sample establishments for interview. Thus, the probability involved in selecting a sample from the DMI file to be contacted for telephone interview and for the additional establishments uncovered in the screening operation is known for each establishment. Unbiased procedures were developed for selecting this sample. Weights which consist of the inverse of the probabilities of selecting the sample units, can be used to prepare unbiased estimates of statistics for the DMI firms from which the sample was selected. These weights, referred to as inflation weights, comprised the first step in the estimation procedure. However, we have also incorporated two stages of ratio estimation; these additional steps have been introduced to improve the precision and reduce some of the biases.

The following sections summarize the details of the NOES estimation procedures. Further details appear in Exhibits 7 and 8.

6.2 Inflation Estimates

It is necessary to discuss the selection probabilities and associated inflation weights for the telephone sample as well as the field interview sample because estimates prepared for publication of NOES data involve inflation estimates from both of these sources.

6.2.1 Inflation Estimate of the DMI Universe

The telephone sample was selected as a systematic sample of units from the DMI universe as discussed in Section 4.2. The telephone sample weight is the inverse of the probability of selecting the unit in the telephone sample (the telephone sample weight has been identified in this report as WT). The weighted sum over all telephone sample units provides an unbiased estimate of the DMI universe. These estimates were used in later procedures to determine the final publication weight for units surveyed in the field.

6.2.2 Inflation Estimate of the NOES Universe

Processing of the telephone sample discussed in Section 5.3 defined the subset of sample units for which surveys were conducted in the field. The probability for a unit surveyed is a function of the probability the unit was selected in the telephone sample. Details of the steps that determined selection probabilities for survey units are given in Exhibit 7.

Estimates of the characteristics of establishments in the NOES universe can be made as the weighted sums of the characteristics of units surveyed in the NOES sample using, as weights, the inverse of the probability the unit was selected for field survey.

6.3 Ratio Estimation

We have incorporated two stages of ratio estimation to improve inflation estimates of the current target universe discussed in 6.2.2. Ratio estimates make use of the relationship

between estimates produced from the sample for a characteristic, for which more reliable, independent data are also available, and the independent estimates. When the characteristic is sufficiently correlated with the items for which improvement in precision is desired, use of ratio estimates will reduce the sampling errors.

We have used two sets of ratio characteristics:

- The total number of units in target industry SIC's by size class from the DMI file; this information is used in the first stage of ratio estimation.
- The number of employees (number of establishments in some instances) for target industry SIC's by size class as published in the 1981 County Business Patterns; this information is used in the second stage of ratio estimation [2].

6.3.1 First Stage of Ratio Estimation

The first stage of ratio estimation used universe counts of DMI establishments in the NOES target industries and inflation estimates of these counts from the telephone sample selections within each of the 11 size classes given in Table 4-1.

For the first stage of ratio estimation, the computer grouped sample units by area into cells within size classes and computed first stage ratio factors that became part of the adjusted estimation weight for all facilities in the cell. The computer process is summarized as follows (See Exhibit 8, Section II for details):

- One ratio cell was defined for each of the two size classes covering the largest units (the

samples selected from DMI establishments in these size classes were selected without regard to their location by PSU). The size classes used in the first stage of ratio estimation are defined by DMI employees and are identified in Table 4-1.

- Ratio cells within each of nine size classes of smaller firms were defined independently as combinations of one or more whole strata of PSU's.

- The numerator of each factor was the universe count of DMI establishments in the size class in all PSU's (sample and nonsample) in the strata comprising the ratio cell; the denominator was the telephone sample inflation estimate for DMI establishments in the corresponding strata.

- Each ratio factor had to be in the range:

  $$.3333 < \text{Ratio} < 3.0000$$

  with at least four sample plants contributing to the estimate of the denominator. A ratio cell for which the ratio factor failed either of these conditions was augmented by incorporating additional strata until the conditions were satisfied.

- The first stage ratio factors (referred to elsewhere in this report as R1 factors) were multiplied by the field interview inflation weights (W) for all surveyed units in the ratio cell and used in subsequent weighting steps.

6.3.2 Second Stage of Ratio Estimation

The ratio characteristics used in the second stage of ratio estimation involved data published in County Business Patterns for 1981 and estimates of these data based on the sample of establishments surveyed in the NOES [2].

The expense of adding the complication of a ratio estimation procedure is usually justified in terms of reductions in sampling error for estimated statistics highly correlated with

the ratio characteristic. For the NOES, this ratio estimate has a further important benefit in bringing the NOES estimates into closer consistency with current levels. The 1981 CBP provides the most current figures available in the detail required for this ratio estimate.

Section III of Exhibit 8 provides additional detail on second stage ratio estimation procedures which are summarized as follows:

- One ratio cell was defined for each of the two size classes covering the largest units.
- Ratio cells within the eight size classes of smaller firms were defined independently in terms of combinations of target SIC's summarized at 2-digit levels. Different size classes were used in the second stage of ratio estimation and were defined by numbers of employees reported by the surveyed establishment.
- The numerator of each ratio factor was the US total CBP count of number of employees in the size class in the target SIC's comprising the ratio cell. The denominator was the estimate of the numerator from the reports of sample units surveyed in the NOES for the corresponding SIC's. Counts of establishments rather than number of employees were used as the ratio characteristic for firms reporting 1,000 or more employees because the CBP does not report US estimates of number of employees in sufficient detail for establishments of this size.
- Denominators of ratio factors were estimated with the field inflation weights multiplied by the first stage ratio estimate factors R1.
- Each ratio factor had to be in the range:

  $$.3333 < \text{Ratio} < 3.0000$$

  with at least four sample plants contributing to the estimate in the denominator. A ratio cell for

which the ratio failed these conditions was augmented by incorporating data for additional SIC's until the conditions were satisfied.

- The second stage ratio factor (referred to elsewhere in this report as R2) became part of the final publication weight for the surveyed unit. The publication weight is defined:

  W * R1 * R2

  the product of the field inflation weight and the first and second stage ratio factors for the plant.

7. ESTIMATION OF SAMPLING ERRORS

7.1 Introduction

The computation of sampling errors for a survey having a complex sample design and using a variety of estimation procedures is a complicated undertaking. We have devised a system sufficiently flexible to provide measures of reliability for all tabulations planned from the NOES data. Certain simplifying assumptions and compromises are needed to reduce the task to manageable proportions.

As used in this report, the term "variance" means the square of the standard error of the estimate (or the square of what is generally referred to as the sampling error). It is more convenient to approach the problem of estimating sampling errors by first computing the variance of the estimate; the square root of the estimated variance becomes the estimated sampling error. In the sections that follow, we discuss some of the general problems in estimating variances by the method of replications and then summarize the steps.

7.2 Estimation of Variances by the Method of Replications

The sampling error of an estimate, based on any sample design with more than a trivially small sample size, using any estimation procedure no matter how complex, may be estimated by the method of replications. This method requires that the sample selection, collection of data, and estimation procedures be independently carried through (replicated) several times. The dispersion of the resulting estimates can be used to measure the variance of the full sample.

Obviously, one would not consider repeating the entire NOES several times simply to estimate sampling errors. A practical alternative is to draw a set of random subsamples from the full sample. The same principles of selection and stages of sampling used for the full sample are used in each of the random subsamples. The subsamples are referred to as replicates. By applying all of the procedures in the regular estimation process to each of the replicates and then computing the dispersion among the resulting estimates, it is possible to obtain an estimate of the sampling error of the NOES.

In addition to measuring the reliability of estimates from the sample, some measure of the impact of specific sampling and estimation procedures on the reliability of estimates can also be obtained. This is accomplished by comparing the dispersion of the replicated estimates with and without the procedure in question. For example, the effect on sampling errors caused by second stage ratio estimation can be determined by computing the variance of estimates that do, and then again for estimates that do not, incorporate the second stage of ratio estimation.

The method of replications has special advantages in reducing the complexity of variance computations experienced using other methods. For example, it may be applied to compute sampling errors for higher order statistics without the need for new variance expressions. Application of the method requires carrying out three simple steps:

- Assemble data for the sample units that make up each of the replicates; this is equivalent to making a copy of the sample data for each unit in each of the subsamples of the full sample.

- Perform the estimation procedure on each of the replicates. The same estimation procedures, prepared for the full sample, are applied separately to each of the replicates.

- Calculate the dispersion of the resulting estimates among the replicates to estimate the variance of the full sample; a relatively simple computation formula is used.

Sample selection methods used for the NOES will result in estimates that are slightly biased (usually overestimates) of the sampling errors. This will happen regardless of whether replications or other methods of variance estimation are used. We do not consider these biases to be serious. The biases will arise primarily because:

- One PSU (rather than two or more) was selected for the sample in each of the 98 PSU strata in the NOES design. As a result, it is necessary to combine nonself-representing strata to estimate the variance. This process introduces a variance component in the replicates not present in the full sample and therefore produces an overestimate of the variance [7]. A different (but less efficient) design could have been used that would select two sample PSU's from each stratum; such a design would permit unbiased estimates of sampling errors. The less efficient design was not adopted because improved efficiency was desired even if some overestimate of sampling errors had to be tolerated.

- Systematic samples of establishments were selected within each sample PSU. Defining half samples within the PSU's as alternative selections of the original sample tends to yield slight overestimates of the variance.

7.3 Defining the Replicates

The number of half samples to use in computing estimates of the variance is a function of the methods used to select the original full sample. Three separate methods were used to designate sample establishments from the DMI file. The methods used and the system for defining half samples are summarized in the following subsections (see Exhibit 9 for details).

7.3.1 Samples of Large Establishments

Establishments reporting 2,500 or more employees on the DMI file were assigned to two size classes (classes 9 and 10, see Table 4-1) and a systematic sample was selected from each class without considering the location of the establishment. The sample within each of these classes was divided into two subsamples by assigning alternate selections to the subsamples. Each of these two subsamples can be called a "pair" and the two subsamples identified as a first and second "member." These two size classes contribute two pairs.

7.3.2 Samples Within Self-Representing (SR) PSU's

Establishments located within the 26 large self-representing PSU's and reporting less than 2,500 employees on the DMI file were assigned to 9 size classes (classes 1 through 8 and 11). A systematic sample of establishments was selected independently within each size class using the overall probability appropriate for the class. Each of the 26 PSU's was named as a pair with two, within-PSU half samples identified as first and second members. Each member comprised a half sample of establishments selected from all of the nine size classes within the PSU.

7.3.3 Samples Within Nonself-Representing (NSR) PSU's

Establishments located within the 72 nonself-representing PSU's and reporting less than 2,500 employees on the DMI file were assigned to the 9 size classes. Within each class, a systematic sample of establishments was selected independently using the within-PSU selection probability necessary to yield a self-weighting sample at the same rates as facilities in SR PSU's.

The 72 NSR strata were formed into 36 pairs with the strata in each pair as homogeneous as possible. Half samples within each pair were formed using the entire sample designated in one or the other sample PSU within each pair.

7.3.4 Number of Half-Sample Replications

The total number of pairs formed is:

$$64 = 2 + 26 + 36.$$

By choosing one of the two members from each of the 64 pairs it would be possible to designate a total of

$$2^{64} = 1.845 * 10^{19}$$

different half samples. The dispersion among estimates from this large number of half samples could provide estimates of the variance. However, MacCarthy has shown [5, 6] the variances can be estimated with equivalent reliability from only 64 orthogonal half-sample replications -- a much smaller number.

To compute the variance of the publication estimate for one item would mean preparing a set of 65 estimates for the item, one for the publication estimate and the additional 64 from the half samples. To reduce costs, we have set up the variance computation process using a sample of 32 of the 64 orthogonal replications. This reduction in computing effort increases the variance of the variance estimates, but does not introduce any new bias in estimates of variance.

7.3.5 Identify the Half-Sample Replications

As the half samples within each stratum and within PSU's are identified as members of the 64 pairs, the replications can be identified by specifying the members from each pair in the replication. The full sample from which the publication estimates are prepared comprises both members of all 64 pairs. The coding scheme used to identify the 32 replicates is discussed in detail in Exhibit 9.

7.4 Estimation Procedure for Each Replicate

For each characteristic for which variance computations are to be performed, the computer prepares 32 estimates based on the half sample replicates defined in the preceding sections. The variance computations also require an estimate of the characteristic based on the full sample. (For convenience, the full sample is called replicate zero). For each characteristic, therefore, the estimation procedure is executed a total of 33 times, once for replicate zero and once for each of the 32 half-sample replicates.

Each of the 32 estimates of the characteristic based on the half samples reproduces as nearly as possible all of the steps in the estimation procedure that are used in the full sample. Thus, as publication estimates use two stages of ratio estimation, each of the half samples is prepared with the same two stage ratio estimation procedure.

The variance estimation package enables the user to identify the characteristic of interest on each of the questionnaires completed in the survey. The computer then prepares 33 weighted estimates for the characteristic; one of these (for replication zero) is the publication estimate obtained as the weighted sum of the characteristic over all sample establishments. The remaining 32 estimates are produced by the computer each of which is, like the publication estimate, a weighted sum over all establishments in the half sample. This process means that estimates from the half sample replications are based on half of the surveyed facilities having weights about twice the weight used when including the facility in the full sample; the other half of facilities have weights of zero for the replication. The weights used for each of the half samples have been computed and entered as part of the survey records.

7.5 Computations to Estimate Variances

The computation formula that provides estimates of relvariance from the 32 replicates is given by

$$v^2_{x'} = \frac{1}{32(x'_o)^2} \sum_r^{32} (x'_r - x'_o)^2 \qquad (12)$$

where

$$V^2_{x'} = \frac{\text{Variance of } x'}{(x')^2}$$

$$= (\text{coefficient of variation})^2,$$

x'_r = the estimate of the characteristic made from the r^{th} replicate ($r = 1, 2, \ldots 32$), and

$x' = x'_o$ = the estimate prepared from the full sample (the zero th replication).

The same computation formula will approximate the relvariance of any estimate prepared from the NOES data.

7.6 Calculation and Presentation of Sampling Errors

With procedures for estimating sampling errors programmed for a computer, it is possible to compute and display estimates of the sampling error for each of the tabulated statistics in a report. The VARNOES computer package allows the user to do this.

There are several problems in computing and publishing sampling errors for each statistic presented in a report.

- It is usually not possible to predict the combination of results an analyst may wish to examine beyond those presented in a report (e.g., additional ratios of, or differences between, estimates given in the report).

- Computation of sampling errors is costly, and additional costs also arise in larger tables needed to accommodate sampling errors as well as the desired estimates.

- Sampling errors have sampling errors of their own when estimated from sample data. The sampling error computed for a statistic will usually have an even larger relative sampling error than the statistic itself. This means that in addition to their lesser reliability, estimated sampling errors may also appear inconsistent for related statistics.

These issues have been met in some surveys by using prior experience to define groups of statistics known to have similar variances. For a new survey, computations are performed for a number of key statistics within the groups (usually for a small subset of the total survey items planned). Least squares methods are then used to develop a set of parameters for a variance function that has, in the past, provided fairly close agreement with observed variance computations for items in the groups. The sampling errors can then be shown as a function of estimate size in a sampling error table that provides the user with sampling errors for all items in the report from the group and need not be shown for every estimate in the report tables. This procedure has been successful for surveys that measure attributes (rather than values) for sample units that have fairly uniform tabulation weights. This system has not been proposed because statistics measured by the NOES do not meet these criteria.

Standard errors for a wide group of statistics have already been prepared and are summarized in the four tables in this section. Approximations to other items that are closely related to the ones in these tables can be made by using the information shown. The VARNOES package can be used for other statistics.

7.7 Sampling Errors for Selected Items

The following four tables (Tables 7-1 through 7-4), show estimates based on NOES data for statistics similar to tables published for the NOHS I analysis [4]. The estimates have been prepared using the two-stage ratio estimation procedure intended for the production of NOES publication estimates.

The first two of the four tables show various categories of plants along with the total number of employees at the plants. For these tables the "Sample Plants" refer to the total number of plants in the sample in the indicated target SIC's. The two categories are:

For Table 7-1	The subset of the NOES universe of plants that have industrial hygiene services.
For Table 7-2	The subset of the NOES universe of plants that require replacement physical examinations.

The last two of the four tables show the reported number of injuries in four categories of plants; note that some plants do not have records of injuries or did not report injuries. The four categories are:

For Table 7-3 Those plants in the NOES universe that have safety engineer services and the plants that have both safety engineer and industrial hygiene services. "Sample Plants" in this table show all sample plants in the indicated target SIC's (the same as in the first three tables).

For Table 7-4 All plants in the NOES universe and in the subset of plants that have unions. The numbers of sample plants reporting unions is also given for the indicated SIC's.

In addition to the estimates of NOES data, the tables show the coefficient of variation for each estimate. The coefficient of variation is the ratio of the sampling error to the estimated statistic. The coefficient of variation (shown as percentages in the tables) enable the relative levels of sampling error to be evaluated when determining whether sampling errors should be computed for similar statistics being considered for a report.

The explanation of how to interpret these coefficients of variation (CV) shown in the tables is made easier by considering the following example.

Table 7-1 shows that the NOES produces an estimated total of 76,460 plants with industrial hygiene services between 8 and 249 employees in the NOES target SIC's. The table also shows this estimate has a CV of 6.11 percent. This information

allows confidence intervals to be constructed around the estimate of plants in this category. The following shows how this information is used to construct confidence intervals:

- Assume the NOES could be conducted a large number of times with all of the surveys conducted during the same period of time, each based on independent sample of the same number of plants selected by the same sampling procedure, and with surveyors chosen and trained in the same manner. Assume an estimate is prepared for each of these surveys using the same estimation procedure. With these assumptions, we would expect roughly two out of three of these estimates to be within 6.11 percent of the 76,460 estimate given in Table 7-1.

- We would also expect about 19 out of 20 of the estimates to be within twice the coefficient of variation that is within 12.22 percent (2 * 6.11) of the estimate in the table.

- By converting the coefficient of variation to a standard error, the same confidence intervals could be stated in terms of absolute rather than relative sampling errors. The standard error of the foregoing estimate is equal to the product of the estimate and its CV; thus 6.11 percent of 76,460 is 4,672, the standard error of the estimated 76,460 plants.

Table 7-1. Final NOES estimates of number of plants and employees in plants with industrial hygiene services (corresponds to NOHS I - Table 2)

SIC	SAMPLE PLANTS	NUMBER OF PLANTS				NUMBER OF EMPLOYEES IN PLANTS			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
ALL TARGET SIC'S	4490	76460	6858	6394	89711	4385024	2325193	9248157	15958374
	CV %	6.11	5.76	4.58	5.37	4.10	5.06	4.12	2.95
MANUFACTURING	2665	40520	5027	4491	50037	2874016	1695143	6480024	11049184
	CV %	5.26	10.08	6.04	4.54	5.00	9.57	4.60	4.19
NON MANUFACTURING	1825	35940	1831	1903	39674	1511008	630050	2768133	4909190
	CV %	9.26	14.95	15.71	8.53	9.17	14.50	9.73	6.83
7	22	1249	-	-	1249	31953	-	-	31954
	CV %	44.26	-	-	44.26	46.97	-	-	46.98
13	61	1164	80	-	1244	34141	27118	-	61259
	CV %	48.08	78.33	-	45.46	57.11	78.33	-	44.14
15	148	1540	62	24	1628	65565	24108	45804	135478
	CV %	35.42	89.2	70.14	33.25	31.05	90.96	49.28	28.92
16	106	1653	109	35	1798	69478	32746	30708	132933
	CV %	30.95	90.95	129.77	30.06	34.51	90.95	94.05	35.76
17	319	3525	138	-	3664	101185	37531	-	138716
	CV %	28.50	71.9	-	27.85	35.73	68.07	-	31.72
20	216	4305	748	272	5326	292724	263210	229172	785106
	CV %	17.49	21.62	36.39	14.85	14.73	21.61	29.50	10.5
21	8	30	-	17	47	1866	-	45151	47017
	CV %	143.13	-	162.88	100.22	143.86	-	92.90	87.76
22	112	722	391	258	1371	92880	133113	222423	448417
	CV %	17.34	24.56	27.67	13.55	23.14	26.13	34.65	15.70
23	194	2211	119	84	2415	165158	35403	73689	274250
	CV %	25.72	90.49	56.42	23.09	14.01	84.58	29.52	18.41
24	91	2380	12	7	2400	190451	3241	8146	201837
	CV %	18.29	264.12	110.57	17.98	23.13	264.12	110.57	21.47
25	73	899	131	32	1061	54926	44675	48937	148539
	CV %	45.13	99.86	77.65	42.71	46.10	101.01	62.11	42.21
26	85	1554	211	122	1887	135395	75269	129097	339760
	CV %	21.91	48.03	48.68	18.24	17.56	50.51	34.82	20.35
27	197	3094	244	207	3546	172187	81809	223650	477646
	CV %	19.72	51.40	46.06	18.70	18.02	48.40	32.25	21.24

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-1. Final NOES estimates of number of plants and employees in plants with industrial hygiene services (corresponds to NOHS I - Table 2) (continued)

SIC	SAMPLE PLANTS	NUMBER OF PLANTS SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	NUMBER OF EMPLOYEES IN PLANTS SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
28	140	2458	277	337	3072	152242	100315	450959	703515
	CV %	17.04	30.02	25.98	13.46	13.78	28.59	25.25	16.68
29	19	109	120	69	298	14364	34778	90619	139762
	CV %	110.00	39.42	57.76	46.08	100.59	41.44	76.51	55.02
30	128	2934	235	186	3354	168956	85829	192174	446958
	CV %	20.14	37.25	39.93	18.67	15.91	39.24	40.25	18.78
31	26	499	119	3	622	39917	36105	4015	80036
	CV %	50.41	65.90	284.95	44.13	41.29	65.33	284.95	39.57
32	107	2461	210	144	2815	142259	81780	97961	322000
	CV %	24.41	35.59	43.49	21.28	20.18	36.99	43.25	16.01
33	114	1707	253	255	2214	139158	92965	497447	729570
	CV %	19.60	30.88	25.30	15.91	18.38	31.62	21.48	16.68
34	293	4864	472	303	5639	345009	158958	297918	801885
	CV %	10.95	17.66	27.61	9.62	9.50	17.39	28.81	12.09
35	334	4079	560	563	5202	295798	180074	881568	1357440
	CV %	13.58	23.47	11.83	11.12	10.39	23.12	14.29	9.42
36	209	2650	499	623	3772	203838	150078	1010006	1363922
	CV %	20.46	27.39	24.26	16.66	14.92	28.95	18.43	15.8563
37	139	1148	218	309	1674	99040	67406	1363747	1530193
	CV %	33.56	44.98	21.35	24.01	20.74	42.26	20.40	17.58
38	86	720	125	359	1203	71126	40571	403975	515672
	CV %	32.46	72.42	34.90	25.15	22.30	70.46	28.22	23.19
39	94	1696	82	338	2116	96723	29565	209369	335657
	CV %	45.85	228.25	64.06	39.33	98.03	206.67	57.55	39.99
40	3	-	-	-	-	-	-	-	-
	CV %	-	-	-	-	-	-	-	-
41	29	264	83	-	347	9094	22938	-	32032
	CV %	89.03	*	-	81.16	81.60	*	-	112.11
42	150	3430	33	30	3493	156837	8597	23770	189204
	CV %	28.08	90.57	90.75	27.36	22.18	90.57	90.87	19.5
44	11	39	27	-	66	1935	7631	-	9565
	CV %	160.74	142.31	-	127.05	160.74	142.31	-	128.34

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-1. Final NOES estimates of number of plants and employees in plants with industrial hygiene services (corresponds to NOHS I - Table 2) (continued)

SIC	SAMPLE PLANTS	NUMBER OF PLANTS				NUMBER OF EMPLOYEES IN PLANTS			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
45	34	987	-	137	1125	63083	-	242169	305972
	CV %	45.37	-	42.06	39.57	65.39	-	33.50	28.83
46	3	599	-	-	599	9578	-	-	9577
	CV %	91.43	-	-	91.42	91.43	-	-	91.42
47	1	-	-	-	-	-	-	-	-
	CV %	-	-	-	-	-	-	-	-
48	69	2154	29	18	2201	114174	13941	46280	174395
	CV %	49.28	153.19	75.29	48.10	47.58	153.19	72.31	36.02
49	65	2171	319	89	2579	172509	117080	104805	394395
	CV %	38.96	50.36	82.66	33.53	36.90	49.09	52.54	29.88
50	151	3792	167	-	3959	161163	50155	-	211317
	CV %	30.37	75.04	-	29.61	25.26	75.31	-	29.72
51	30	1617	-	-	1617	53214	-	-	53214
	CV %	45.73	-	-	45.73	61.85	-	-	61.85
55	56	3200	-	-	3200	76024	-	-	76023
	CV %	37.20	-	-	37.20	41.60	-	-	41.59
72	88	1678	41	13	1732	62416	10283	10670	83369
	CV %	39.10	125.38	*	37.81	30.78	125.38	*	33.26
73	135	1570	224	190	1983	102173	76304	210359	388836
	CV %	49.08	36.94	60.31	40.16	34.09	33.22	44.61	26.76
75	73	2613	-	-	2613	52843	-	-	52842
	CV %	46.47	-	-	46.46	45.41	-	-	45.41
76	39	864	36	-	900	12775	12672	-	25447
	CV %	86.52	*	-	83.4	82.03	*	-	76.28
80	229	1553	482	1366	3401	142222	188946	2053567	2384734
	CV %	26.70	30.15	14.12	14.00	19.58	29.65	12.62	10.62
84	3	274	-	-	274	17925	-	-	17925
	CV %	85.19	-	-	85.19	61.64	-	-	61.63

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-2. Final NOES estimates of number of plants and employees in plants requiring replacement physical examinations (corresponds to NOHS I - Table 10)

SIC	SAMPLE PLANTS	NUMBER OF PLANTS				NUMBER OF EMPLOYEES IN PLANTS			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
ALL TARGET SIC'S	4490	113714	8082	7653	129449	5738758	2797558	10869942	19406257
	CV %	4.23	5.80	5.90	3.79	2.95	5.71	4.64	2.81
MANUFACTURING	2665	44439	5426	4830	54695	3061589	1844779	6998693	11905061
	CV %	6.16	10.20	4.67	5.52	5.36	10.34	3.66	3.84
NON MANUFACTURING	1825	69275	2656	2823	74754	2677169	952778	3871248	7501196
	CV %	5.21	18.18	16.25	4.98	5.14	19.40	10.41	6.03
07	22	313	-	-	313	14788	-	-	14788
	CV %	80.88	-	-	80.88	93.68	-	-	93.68
13	61	3315	90	46	3451	125940	36962	31383	194284
	CV %	27.05	61.87	*	27.31	27.88	61.98	*	34.72
15	148	1185	-	59	1244	53488	-	39479	92967
	CV %	45.67	-	159.14	44.95	49.52	-	159.14	72.94
16	106	1461	108	71	1640	53378	34270	64095	151743
	CV %	43.49	104.07	79.35	42.11	37.54	104.07	62.61	41.00
17	319	4172	36	-	4208	135909	9061	-	144970
	CV %	29.17	87.98	-	28.70	31.95	87.98	-	28.99
20	216	5498	807	341	6646	391775	280713	320024	992513
	CV %	13.36	23.26	32.49	11.42	17.48	24.87	29.41	10.04
21	8	4	-	54	58	258	-	98088	98346
	CV %	133.77	-	63.46	56.50	133.77	-	67.76	67.46
22	112	771	342	247	1360	93784	111368	224095	429247
	CV %	24.57	25.26	29.57	16.39	23.43	26.82	39.56	21.54
23	194	263	74	63	400	26029	25808	62757	114594
	CV %	45.56	57.6	76.03	30.72	59.37	55.65	76.03	39.30
24	91	2283	13	19	2314	159907	3241	35767	198915
	CV %	16.76	264.13	42.16	16.67	19.21	264.13	52.53	20.23
25	73	1191	142	32	1365	84206	53536	48937	186679
	CV %	30.19	31.95	77.66	26.83	26.05	32.09	62.11	19.15
26	85	2085	247	112	2444	209515	85758	130079	425352
	CV %	15.33	43.63	43.81	14.75	11.19	46.07	33.92	17.09
27	197	897	334	184	1415	75744	102661	213453	391857
	CV %	18.07	45.04	43.78	16.94	19.59	44.84	28.59	22.16

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-2. Final NOES estimates of number of plants and employees in plants requiring replacement physical examinations (corresponds to NOHS I - Table 10) (continued)

SIC	SAMPLE PLANTS	NUMBER OF PLANTS				NUMBER OF EMPLOYEES IN PLANTS			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
28	140	4484	318	380	5182	243328	115799	476341	835468
	CV %	10.78	33.08	16.65	9.17	6.14	30.92	21.28	14.24
29	19	387	155	101	643	35144	45848	117545	198537
	CV %	55.87	36.96	41.96	37.47	56.37	38.17	59.77	39.92
30	128	2145	102	193	2440	168320	35410	216359	420089
	CV %	28.35	8812	38.27	27.26	20.20	92.63	35.06	20.49
31	26	-	33	-	33	-	9755	-	9755
	CV %	-	107.44	-	107.44	-	107.44	-	107.44
32	107	4604	184	184	4972	206552	71840	128993	407385
	CV %	22.01	35.55	38.48	20.00	15.42	37.41	38.52	11.96
33	114	2310	391	318	3019	191728	131174	583358	906260
	CV %	14.47	21.72	20.82	11.81	13.33	21.52	16.65	10.4
34	293	6583	697	223	7503	398043	231374	257854	887272
	CV %	13.33	12.53	38.89	12.06	9.16	11.10	42.55	13.94
35	334	5658	612	669	6939	380370	215578	999920	1595868
	CV %	1517	16.58	12.99	11.62	11.25	16.00	15.53	9.02
36	209	1799	591	625	3015	132425	191549	1015895	1339869
	CV %	26.54	18.65	24.55	16.31	23.72	17.66	20.26	15.96
37	139	1340	134	391	1866	107578	49198	1483381	1640157
	CV %	22.80	58.94	21.58	18.27	16.23	57.63	15.17	14.12
38	86	755	132	358	1246	79879	42735	376478	499092
	CV %	25.09	68.70	32.85	19.05	16.51	69.68	30.59	23.69
39	94	1379	118	338	1835	77004	41434	209369	327807
	CV %	46.43	208.58	64.06	39.19	97.78	207.45	57.55	42.51
40	3	-	48	16	64	-	15606	9005	24611
	CV %	-	66.63	*	56.10	-	67.72	*	56.28
41	29	2615	83	-	2698	96504	22938	-	119442
	CV %	36.08	*	-	35.62	30.60	*	-	34.57
42	150	13941	191	75	14207	477757	60699	62235	600690
	CV %	11.98	53.01	61.02	11.92	13.55	57.93	50.94	12.32
44	11	553	54	-	607	42555	19409	-	61965
	CV %	61.59	107.62	-	61.37	46.01	113.91	-	58.20

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-2. Final NOES estimates of number of plants and employees in plants requiring replacement physical examinations (corresponds to NOHS I - Table 10) (continued)

SIC	SAMPLE PLANTS	NUMBER OF PLANTS				NUMBER OF EMPLOYEES IN PLANTS			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
45	34	1129	118	147	1394	85029	47505	264255	396789
	CV %	39.19	*	38.44	34.25	53.07	*	29.39	24.64
46	3	1007	-	-	1007	43716	-	-	43716
	CV %	61.47	-	-	61.47	59.36	-	-	59.36
47	1	-	-	-	-	-	-	-	-
	CV %	-	-	-	-	-	-	-	-
48	69	2746	179	34	2958	159259	62586	70083	291928
	CV %	34.63	54.61	63.49	31.20	24.32	56.66	62.34	20.68
49	65	7276	696	173	8144	356763	258588	146565	761916
	CV %	23.57	50.51	57.43	22.07	22.29	50.97	44.74	23.60
50	151	9959	43	-	10002	366247	13054	-	379301
	CV %	20.67	63.06	-	20.65	20.86	63.37	-	20.50
51	30	5546	-	-	5546	127112	-	-	127112
	CV %	23.35	-	-	23.35	34.61	-	-	34.61
55	56	1094	-	-	1094	24687	-	-	24687
	CV %	64.59	-	-	64.59	83.35	-	-	83.35
72	88	1611	11	-	1622	39890	3746	-	43636
	CV %	53.09	*	-	52.71	42.54	*	-	40.78
73	135	3609	131	186	3925	118374	51403	210856	380633
	CV %	28.60	32.82	64.41	26.26	25.92	30.57	52.84	31.16
75	73	4024	-	-	4024	96257	-	-	96257
	CV %	28.83	-	-	28.83	23.40	-	-	23.40
76	39	991	36	-	1028	22524	12672	-	35196
	CV %	55.59	124.32	-	52.70	46.11	124.32	-	42.59
80	229	2730	832	2016	5578	236991	304279	2973293	3514563
	CV %	21.28	23.03	15.26	12.51	15.62	24.28	12.29	10.18
84	3	*	-	-	-	-	-	-	-
	CV %	*	-	-	-	-	-	-	-

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-3. Final NOES estimates of recorded injuries in plants with both safety engineers and industrial hygiene services (corresponds to NOHS I - Table 34)

SIC	SAMPLE PLANTS	SAFETY ENGINEERS SERVICES				BOTH SERVICES			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
ALL TARGET SIC'S	4490	1355926	405274	708814	2470014	597875	264630	542144	1404649
	CV %	3.92	12.54	6.47	3.87	9.57	15.12	7.78	4.31
MANUFACTURING	2665	802975	294238	478656	1575870	411649	192610	375136	979395
	CV %	5.91	16.70	6.45	5.59	9.70	16.92	6.69	5.92
NON MANUFACTURING	1825	552951	111036	230158	894145	186226	72020	167008	425254
	CV %	7.76	24.00	14.78	7.85	19.21	33.27	17.97	11.85
07	22	7617	-	-	7617	7617	-	-	7617
	CV %	63.84	-	-	63.84	63.84	-	-	63.84
13	61	35087	9356	2169	46612	505	2664	-	3169
	CV %	45.26	63.64	*	25.84	78.17	78.33	-	67.36
15	148	59331	20286	10301	89918	11686	11525	4339	27551
	CV %	24.80	60.33	52.14	20.99	45.87	101.25	69.63	45.39
16	106	47390	8780	10140	66310	13846	982.39	2024	16852
	CV %	20.23	59.10	84.76	21.36	46.69	90.96	164.84	44.80
17	319	118698	7038	-	125736	13767	4908	-	18675
	CV %	15.07	60.92	-	14.58	44.83	77.00	-	43.06
20	216	126617	53629	45255	225502	61754	44801	29044	135599
	CV %	22.87	28.47	30.92	17.44	32.01	29.52	40.87	18.76
21	8	438	9269	-	9707	438	1131	-	1568
	CV %	133.44	69.90	-	65.66	133.44	113.43	-	78.73
22	112	17891	14249	15329	47470	10310	7846	14108	32264
	CV %	20.67	46.11	40.84	24.26	27.69	34.69	38.01	15.76
23	194	23086	8042	13820	44948	11327	1794	7010	20130
	CV %	31.01	43.15	48.69	19.47	40.81	160.15	60.11	34.37
24	91	59550	2223	4237	66010	35735	1013	691	37439
	CV %	19.74	103.48	47.77	18.87	36.41	264.12	110.58	33.44
25	73	17369	19768	14766	51903	3753	6554	3056	13363
	CV %	29.45	37.52	67.86	27.82	64.39	115.05	72.86	69.81
26	85	27258	2895	8001	38154	13824	2895	5996	22715
	CV %	27.55	56.41	33.19	22.36	36.64	56.41	40.00	23.39
27	197	30256	12156	15909	58322	12065	8732	13745	34543
	CV %	23.16	53.64	34.99	23.42	31.94	65.74	36.47	31.98

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-3. Final NOES estimates of recorded injuries in plants with both safety engineers and industrial hygiene services (corresponds to NOHS I - Table 34) (continued)

SIC	SAMPLE PLANTS	SAFETY ENGINEERS SERVICES				BOTH SERVICES			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
28	140	27559	9093	12926	49578	21090	5611	12542	39243
	CV %	22.64	33.09	19.22	15.72	33.51	50.02	21.73	19.25
29	19	5108	4222	13516	22846	2234	2319	11817	16370
	CV %	60.01	45.60	64.32	44.56	104.03	50.16	77.58	59.23
30	128	37486	13750	14971	66207	20409	6883	11054	38347
	CV %	25.61	87.26	24.62	29.34	25.21	49.05	38.57	19.37
31	26	1618	5900	684	8202	1618	3821	684	6124
	CV %	52.79	55.36	284.96	49.82	52.79	71.84	284.96	59.46
32	107	40662	6690	13931	61282	20275	5849	10699	36822
	CV %	24.68	54.78	46.01	17.34	30.91	49.97	53.18	22.21
33	114	32455	17408	55694	105557	30018	8371	46098	84487
	CV %	19.55	33.10	28.3	17.70	22.09	46.24	31.85	20.55
34	293	133794	42523	22593	198910	78173	33158	20072	131403
	CV %	9.80	25.88	34.45	11.13	14.67	34.35	37.88	17.34
35	334	125279	34017	70608	229904	42801	25709	57057	125566
	CV %	15.20	22.13	17.47	9.13	23.33	34.71	15.49	11.85
36	209	30852	26794	51440	109087	17860	13346	43722	74927
	CV %	16.85	13.35	32.93	17.01	22.81	30.89	36.12	22.33
37	139	40259	6530	59892	106681	14395	3326	55360	73081
	CV %	20.67	33.32	14.29	11.79	30.02	56.71	13.56	11.63
38	86	12492	5699	14327	32518	4966	5146	9764	19876
	CV %	23.95	70.66	47.75	29.22	32.63	76.16	38.11	28.68
39	94	12944	8651	21488	43081	8604	54.36	21488	35527
	CV %	124.64	224.61	60.38	60.19	141.08	213.60	60.38	43.28
40	3	-	-	-	-	-	-	-	-
	CV %	-	-	-	-	-	-	-	-
41	29	2046	-	-	2046	47	-	-	47
	CV %	73.67	-	-	73.67	123.77	-	-	123.77
42	150	38560	4695	7510	50765	11745	1449	2649	15843
	CV %	21.36	30.59	62.51	16.84	26.48	90.57	98.96	28.39
44	11	3451	7162	-	10613	542	4650	-	5191
	CV %	142.74	108.04	-	97.06	160.74	142.31	-	134.13

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-3. Final NOES estimates of recorded injuries in plants with both safety engineers and industrial hygiene services (corresponds to NOHS I - Table 34) (continued)

SIC	SAMPLE PLANTS	NUMBER OF PLANTS				NUMBER OF EMPLOYEES IN PLANTS			
		SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
45	34	39327	1773	51875	92974	30317	-	51875	82192
	CV %	74.46	*	47.11	38.79	89.59	-	47.11	39.97
46	3	7208	-	-	7208	599	-	-	599
	CV %	68.42	-	-	68.42	91.43	-	-	91.43
47	1	-	-	-	-	-	-	-	-
	CV %	-	-	-	-	-	-	-	-
48	69	28683	2312	3088	34082	14917	734	3040	18691
	CV %	42.57	93.45	84.50	36.03	54.91	153.19	85.67	45.45
49	65	27150	8179	4203	39531	13725	8179	3559	25462
	CV %	37.99	57.31	52.52	32.52	64.69	57.31	55.19	43.92
50	151	71636	23714	-	95349	40683	23714	-	64397
	CV %	28.45	96.59	-	35.81	41.79	96.59	-	49.12
51	30	6739	-	-	6739	2408	-	-	2408
	CV %	44.29	-	-	44.29	78.70	-	-	78.70
55	56	14597	-	-	14597	4188	-	-	4189
	CV %	39.35	-	-	39.35	77.82	-	-	77.82
72	88	4471	535	-	5006	1535	535	-	2070
	CV %	47.96	125.38	-	47.30	68.46	125.38	-	62.88
73	135	14549	3407	5666	23622	8364	2549	4862	15775
	CV %	34.64	36.87	55.25	25.02	55.32	36.71	60.74	34.99
75	73	9980	-	-	9980	5811	-	-	5811
	CV %	43.65	-	-	43.65	94.38	-	-	94.38
76	39	8486	1666	-	10151	356	1666	-	2022
	CV %	43.80	124.32	-	43.36	156.93	124.32	-	107.47
80	229	7945	12136	135206	155287	3568	8467	94659	106695
	CV %	34.65	42.10	17.07	14.05	46.06	44.28	17.17	13.46
84	3	-	-	-	-	-	-	-	-
	CV %	-	-	-	-	-	-	-	-

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-4. Final NOES estimates of recorded occupational injuries in all plants and in plants having one or more unions (corresponds to NOHS I - Table 36)

SIC	PLANTS WITH UNIONS					ALL PLANTS				
	SAMPLE PLANTS	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SAMPLE PLANTS	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
ALL TARGET SIC'S	1830	960447	269009	506898	1736354	4490	1990574	466107	773559	3230240
	CV%	6.11	13.01	8.91	4.14		2.56	9.35	5.45	2.81
MANUFACTURING	1168	459715	187376	362823	1009914	2665	1128648	330445	508361	1967454
	CV%	7.66	14.56	8.50	5.54		5.16	13.41	5.49	4.84
NON MANUFACTURING	662	500732	81633	144074	726440	1825	861925	135662	265199	1262786
	CV%	10.66	29.01	20.12	9.68		5.86	20.92	12.86	6.46
07	2	3670	-	-	3670	22	13209	-	-	13209
	CV%	105.26	-	-	105.26		44.46	-	-	44.46
13	4	14560	2664	-	17224	61	72059	9356	2169	83584
	CV%	102.17	78.33	-	78.92		20.002	63.64	*	17.10
15	78	63378	12016	12785	88178	148	92011	23455	15334	130800
	CV%	26.09	46.05	43.83	23.49		17.55	52.20	44.99	17.33
16	51	31383	7587	1343	40313	106	66787	8780	10140	85707
	CV%	26.37	66.76	107.46	22.05		22.95	59.10	84.76	20.34
17	171	164353	6685	-	171038	319	209898	13396	-	223294
	CV%	15.99	72.79	-	15.38		12.37	56.63	-	11.59
20	107	75952	45380	51199	172531	216	154118	56401	55312	265831
	CV%	31.41	34.19	22.78	16.90		21.01	26.14	22.43	14.73
21	3	101	-	8522	8622	8	438	9974	-	10411
	CV%	*	-	79.61	78.57		133.44	61.45	-	57.84
22	30	11551	4337	5272	21160	112	25933	14996	17060	57989
	CV%	24.22	62.38	63.68	22.24		16.11	42.81	36.34	18.75
23	90	14894	6967	6822	28683	194	38797	14321	15188	68306
	CV%	31.83	38.43	66.21	25.75		17.37	23.59	45.95	13.98
24	25	22229	2223	2359	26812	91	80860	2223	4237	87319
	CV%	32.43	103.49	46.40	30.00		17.25	103.49	47.77	16.51
25	29	16478	1594	1350	19422	73	31946	21484	14766	68196
	CV%	38.76	127.99	145.65	33.09		21.00	30.90	67.86	21.23
26	57	28111	4197	8001	40309	85	40903	4197	8001	53101
	CV%	17.62	52.15	33.19	13.85		16.40	52.15	33.19	13.48

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

Table 7-4. Final NOES estimates of recorded occupational injuries in all plants and in plants having one or more unions (corresponds to NOHS I - Table 36)(continued)

SIC	PLANTS WITH UNIONS					ALL PLANTS				
	SAMPLE PLANTS	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SAMPLE PLANTS	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
27	69	12177	8556	16356	37089	197	40215	14680	20160	75054
	CV%	23.86	59.83	35.07	26.46		16.66	43.14	32.5	19.63
28	71	24464	6257	4673	35395	140	36234	9497	12926	58657
	CV%	24.15	36.63	34.16	20.80		18.73	31.55	19.22	13.79
29	11	1033	4222	11817	17071	19	6128	4222	13516	23865
	CV%	87.24	45.60	77.58	56.92		54.47	45.60	64.32	44.55
30	44	22839	389	10052	33281	128	62914	13750	14971	91634
	CV%	20.45	151.30	41.40	18.11		13.59	87.26	24.62	18.78
31	9	2130	2154	-	4284	26	5141	5900	902	11943
	CV%	43.39	86.00	-	47.62		33.36	55.36	221.26	33.65
32	56	17735	7554	13363	38651	107	54433	8628	13931	76991
	CV%	25.02	36.81	49.42	20.69		17.38	34.08	46.01	12.40
33	74	26770	16060	53399	96229	114	57358	19483	55694	132535
	CV%	18.21	39.74	32.37	17.54		15.62	30.13	28.30	16.50
34	130	62142	30029	10842	103014	293	183328	49208	23069	255605
	CV%	17.18	38.79	56.52	17.32		9.46	21.89	33.50	9.68
35	128	64445	24788	51354	140587	334	179002	42751	73935	295688
	CV%	13.82	26.35	24.17	10.59		9.10	17.03	17.00	6.35
36	86	14235	11632	27747	53614	209	46107	27825	55073	129004
	CV%	29.64	43.46	35.02	20.72		11.87	12.97	31.22	14.53
37	96	27778	3239	56633	87650	139	49781	6530	62086	118397
	CV%	32.00	71.06	18.07	14.53		15.25	33.32	14.34	10.85
38	25	4708	4104	5248	14060	86	15898	5699	16075	37672
	CV%	14.94	78.04	49.58	28.91		22.88	70.66	50.10	30.52
39	28	9944	3696	17813	31452	94	19116	8651	21488	49254
	CV%	124.35	142.08	74.98	37.87		107.25	224.61	60.38	60.73
40	3	-	-	-	-	3	-	-	-	-
	CV%	-	-	-	-		-	-	-	-
41	7	453	-	-	453	29	4160	-	-	4160
	CV%	108.03	-	-	108.03		44.76	-	-	44.76
42	82	34188	2942	5300	42429	150	64602	5669	8931	79201
	CV%	16.87	58.26	42.70	13.95		15.12	31.03	55.18	12.82
44	3	542	7162	-	7704	11	6528	7162	-	13690
	CV%	*	108.04	-	105.63		80.96	108.04	-	78.76

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION

Table 7-4. Final NOES estimates of recorded occupational injuries in all plants and in plants having one or more unions (corresponds to NOHS I - Table 36)(continued)

SIC	PLANTS WITH UNIONS					ALL PLANTS				
	SAMPLE PLANTS	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL	SAMPLE PLANTS	SMALL 8-249	MEDIUM 250-500	LARGE OVER 500	TOTAL
45	21	44078	-	54696	98773	34	45555	1773	54696	102023
	CV%	66.66	-	45.14	37.65		63.21	100.00	45.14	37.01
46	2	6047	-	-	6047	3	7208	-	-	7208
	CV%	81.21	-	-	81.21		68.42	-	-	68.42
47	-	-	-	-	-	1	774	-	-	774
	CV%	-	-	-	-		*	-	-	*
48	23	21897	1719	3088	26704	69	33466	2549	3088	39103
	CV%	48.22	95.99	84.50	39.90		38.96	86.65	84.50	33.39
49	42	34343	11492	4309	50145	65	44602	12707	4309	61617
	CV%	30.61	53.57	51.40	26.17		28.99	49.27	51.40	24.97
50	43	59454	23714	-	83167	151	95410	23713	-	119123
	CV%	33.57	96.59	-	40.78		22.04	96.59	-	29.58
51	7	6044	-	-	6044	30	10755	-	-	10755
	CV%	60.45	-	-	60.45		36.93	-	-	36.93
55	2	6867	-	-	6867	56	23616	-	-	23616
	CV%	89.95	-	-	89.95		40.31	-	-	40.31
72	21	2407	161	-	2568	88	6508	695	-	7203
	CV%	48.18	*	-	44.57		41.22	95.05	-	39.37
73	26	3886	941	2833	7660	135	19404	4981	6766	31151
	CV%	66.44	89.95	79.49	39.13		28.44	29.13	48.48	21.74
75	6	1827	-	-	1827	73	16995	-	-	16995
	CV%	73.21	-	-	73.21		39.10	-	-	39.10
76	7	719	1666	-	2384	39	16003	1666	-	17669
	CV%	100.91	*	-	63.62		35.33	*	-	38.24
80	61	636	2886	59722	63243	229	12375	19761	159766	191903
	CV%	99.98	51.79	20.21	19.08		26.94	30.29	14.63	11.64
84	-	-	-	-	-	3	-	-	-	-
	CV%	-	-	-	-		-	-	-	-

- NO SAMPLE PLANTS.
* ONE SAMPLE OBSERVATION.

8. INFORMATION FOR FUTURE SURVEY PLANNING

8.1 Introduction

This section reviews the assumptions made in developing the NOES sample design and offers methods of comparing these assumptions with measures of cost and variance actually observed with the design as executed. Also recorded are some of the early planning experiences in introducing programs for minimizing and measuring the contribution of the surveyor to the overall variability of the NOES design. This section is intended to preserve information useful for evaluating the NOES results and for planning a future version of the study.

8.2 Field Survey Procedures

At the time data collection began, there was some question whether the survey could be carried out over the two years that were planned. Because of this uncertainty, the intended two-year data collection phase was (for a time) restructured as two, independent surveys with data to be collected for one of these parts in the first year, with the second covered in the next year. This change would permit estimates (with reduced precision) to be published if the second year of field work could not be conducted.

We made extensive use of the workload subsamples (see Section 4.4.2) originally designated as part of the within-PSU sample selection process to reduce the originally planned sample. Later, when the survey was able to go forward as planned, the sample was restored by reinstating workload subsamples.

NOES data collection involved field visits to about 4,500 establishments located throughout the 50 States and the District of Columbia. A surveyor force varying in number, but comprising as many as 22 trained surveyors during part of the time (employing 31 different surveyors at some time during the 32-month survey period), conducted the interviews. Teams of surveyors were situated to operate independently at widely separated locations so that the opportunity for direct and continual observation of the quality of surveyors' work was difficult.

Variability in interviewer quality was recognized as an important component of the total precision of the survey. Original plans called for methods of controlling and measuring the impact of the surveyors concerning the quality of the survey. However, the initial budget did not permit the introduction of any of these methods. These plans were revived when the budgetary concerns were resolved. By that time, however, it was necessary to implement a vigorous recruitment and training program to bring the surveyor force up to planned levels. Systematic programs for interviewer control and measurement were not implemented, although the following plans were considered.

8.2.1 Minimizing Surveyor Variability

After the initial and on-the-job training had been completed, we would focus attempts to minimize surveyor variability on feedback programs for the interviewers. Three methods were considered to be used for the facilities remaining to be interviewed (at that time, approximately 40 percent of the

total). Rough cost estimates were about $45,000 for these three programs:

- Specially trained control staff would accompany surveyors into facilities to observe the conduct of the interviewer at a sample of firms.

  At completion of each interview, the observer would discuss procedures requiring improvement, emphasizing the most important problems noted. Similar systems have been useful to illustrate what is desired of field staff in other surveys.

- Telephone contact at a sample of firms could obtain general information from the respondent on the quality of the surveyors' work.

  Telephone contacts would be made from a central office for a small sample of the plants surveyed by each member of the field staff. The respondent plants would be asked questions to elicit information on the quality of the interviewer's work; for example, did the surveyor appear at the appointed time, was the survey purpose adequately explained, did the surveyor conduct the interview in a business-like manner, did the surveyor have adequate answers to questions the respondent may have asked.

  The field surveyors would be made aware of the verification procedure. The program was intended to increase our confidence that work was actually done as reported and to emphasize to the surveyor the importance of the survey effort.

- A retraining session with all surveyors and supervisors participating was the third method considered.

  This would include the experienced as well as the more recently hired surveyors. The session would give special attention to situations where surveyors must deal with interpretations and make informed judgments as to what must be recorded. The intent would be to insure that all of the surveyors, particularly the newer ones, were making interpretations consistent with the objectives of the study.

A session of this kind would permit review of problems turned up by the observation program and telephone contacts of respondents. Training sessions of this kind have been useful in other surveys to correct errors in the understanding of training, to insure consistent treatment of interview problems, and to maintain surveyor morale.

8.2.2 Measurement of Surveyor Variability

Several alternatives are theoretically possible for measuring interviewer variability. Some form of each of the three methods discussed below was considered during initial survey planning but none was implemented, largely because the complications in the survey processes that would result were considered unacceptable at the time.

The first two of these programs are based on statistical theory and have been applied to other studies. However, these two programs would have their limitations when applied to the NOES. When the decision could be made, more than half of the sample had been surveyed; thus, analyzing the outcome of the experiment would be difficult because of the restricted sample size and because the area represented by the remaining sample did not represent a cross section of the universe. However, it might have been possible to draw some conclusions about the relative importance of the contribution of the interviewers to the total variance of the design. The experiment would probably have shown evidence of important surveyor variance if it was large. If the variability was too small to be detected by the study, this in itself would have been useful information.

The three potential methods for measuring surveyor variability are described in the following paragraphs.

- Revisit a sample of firms and perform independent reinterviews of the facilities.

  A direct comparison could be made between the results produced originally by the team and by the reinterview. Teams would not be informed when reinterviews would be made so that normal performance would be unaffected. With a sufficiently large reinterview sample, this method could theoretically allow evaluation of individual surveyors as well as teams.

  This alternative was considered earlier but not adopted because a second visit to conduct a reinterview was considered an unreasonable burden on the facility.

- Designate two random subsamples of the workload in each of the remaining PSU's and make survey assignments such that all possible pairs of teams work in one or more common PSU's.

  Each team in the pair would do one of two random subsamples of the workload in the PSU; in a given PSU, for example, one team would be assigned the A workload subsample, the other team the B and C workload subsamples. Differences between the two teams within the same PSU would provide a means for assessing interviewer team variability.

  Because the study would have to be confined to the PSU's in which surveys had not been started, the sample would not have been large enough to evaluate the variability between individual surveyors. Because surveying operated as a team effort, the variability measurement would be among teams. This would mean that teams, once established, would have to be kept intact insofar as possible. Thus, when members quit the survey or new team members were introduced, a consistent method of reconstituting the teams would have to be applied. Travel costs would also be affected because teams would have half-sized assignments in twice as many PSU's. The travel schedule for each team would have to be fixed (instead of the flexible arrangement used) because each of all possible pairs of teams would have to operate in at least one common PSU with the assigned interviews completed within a restricted time period to prevent the introduction of differences caused by the time of survey.

- Derive indications of surveyor performance by analysis of questionnaire returns.

  The survey questionnaires have provision for recording the ability of the surveyor team to elicit cooperation by respondent management during the interview as well as nonresponse rates and responses to the content items. As each survey document identifies a surveyor participating (and therefore the team as constituted when the facility was surveyed), some evaluation of the team could be developed based on item omissions, number of firms completed, management cooperation, and times required to complete the interviews.

  This analysis could provide indications of interviewer variability but without the ability to quantify its importance relative to other sources of survey error.

8.3 Update of Design Parameters

8.3.1 Introduction

The design of the NOES sample began by considering an optimum sample based on a number of assumptions. It is useful to review the basis of these assumptions and consider how the design could have been improved had more current knowledge been available when the NOES design was determined. These assumptions were:

(1) The number of person hours to survey a facility would be about the same as in the NOHS I for corresponding facility sizes.

(2) A force of 21 trained surveyors would be available for a period of 24 months.

(3) The relvariance for key statistics among the facilities within a size class would be about the same in all size classes. We allowed for higher relvariances in the class of smallest and largest facilities.

(4) We planned a two-stage sample: first a sample of PSU's and second, a sample of establishments within the sample PSU's. We assumed the contribution of the variance arising from the sampling of PSU's would be small and could be ignored.

8.3.2 Review of Original Design Parameters

This section examines the assumed design parameters in light of the NOES experience.

1. Person Hours to Survey Sample Facilities

Table 8-1 displays the average number of person hours used to complete the survey of establishments in the earlier NOHS I program. These surveyor hours per establishment were important in determining the NOES design because the total sample was fixed at the number of facilities that could be surveyed in all the person hours contributed by the surveyors over the length of their work periods. Average hours observed for the NOES are also given in this table. Note that the total cost of survey work in the field must include the cost of travel, per diem, training, and other overhead costs, as well as direct observation of the sample facilities.

The table shows that NOHS I and NOES hours were about the same for the smaller facilities but NOES were much less for the larger firms. Basing the NOES sample allocations on NOHS I costs therefore meant we did not survey as many larger firms as we would have done if we had information on the actual costs.

This does not mean the sample is biased. It implies that we could have had a design with smaller sampling errors for the more important statistics.

The difference in NOHS I and NOES surveyor hours per facility shown in Table 8-1 may be partially explained by differences in survey procedures, composition of surveyor teams, and improvements in supervision of the field staff.

2. Surveyors' Resources

Table 8-2 shows the number of surveyors reported active during any part of the 32 months of the survey period. The 21 surveyors expected to be available for the planned 24-month survey period were reached during 5 of the 32 months. The increase in the number of surveyors to 17 and over occurred 16 months into the survey period and was maintained at this level, or higher, for 13 months until the survey entered the final clean-up phase.

The 461 implied surveyor-months in Table 8-2 is less than 10 percent short of the planned

$$21 * 24 = 504$$

surveyor-months. However, because Table 8-2 shows months in which any surveyor activity occurred (no matter how little) the number of person hours contributed to the survey is less than implied by the table.

3. Variation Among Plants within Size Classes

Expression (6) of Section 2.2.2 of this report shows how variances among establishments within the size classes are

used along with unit costs to determine an efficient allocation of the sample to each size class. The needed variances were not available when the NOES sample was designed and Section 2.1.3 discusses the assumptions used in developing the NOES design.

The NOES sample data and the VARNOES package provide the opportunity to prepare estimates of these variances; the results may be used to examine the validity of the assumptions used and can have impact on future designs of surveys having objectives similar to the NOES. The variances have been computed but their analysis is not covered in this report; however, they are available for analysis. (See Exhibit 11 for the specifications).

4. Unit Variances for PSU Sampling

The field survey at establishments having less than an expected 1,000 employees was confined to a sample of 98 primary sampling units; larger establishments were selected without regard to their physical location. Decisions on the number of sample PSU's and the establishment sizes confined to them were made essentially for cost and administrative convenience. It was assumed that confining the sample to 98 PSU's would not have an important impact on the reliability of the sample design. This assumption can be checked if the study is to be repeated.

The VARNOES package has been used to approximate the components of variance of NOES estimates caused by confining the surveys to the 98 sample of PSU's. This can be done by a subtractive process, by first examining the variance of the NOES publication estimate (the "total variance") and then subtracting the variance computed for sampling establishments within the PSU's (the "within PSU variance"). These results are available for analysis; the specifications are given in Exhibit 13.

8.4 Efficiency of the NOES Sample

The total variance for an estimator can be expressed, theoretically, as the sum of separate variance components with each component providing a measure of the loss or gain in precision caused by each of the stages of sampling and estimation.

The method of replications can be used to compute approximations for some of these components. The variance package has been used to prepare a set of variances that will be useful when considering improvements in the design of future surveys having objectives similar to the NOES. The analytical computations are available and have been produced following the specifications given in Exhibit 12.

8.4.1 The Estimation Procedures

The estimation procedures described in Section 6 show how two stages of ratio estimation modify the unbiased estimates. The unbiased estimates are produced as weighted tabulations of characteristics of surveyed units using the inverse of the probability of selecting the surveyed unit as the weight for each unit. The precision of the two estimates, the unbiased estimate and the two-stage ratio estimation procedure used for publication estimates, can be measured by comparing the variances of these two estimators. Variances of unbiased estimates have been prepared and are available (see Exhibit 12 for specifications); variances of publication estimates are discussed in Section 7.6.

It should be noted, however, that the reduction of sampling errors was not the only purpose of the ratio estimation procedure. Another reason (particularly for the use of County Business Patterns) was to reduce possible biases arising from

using the DMI list of establishments and assuming it constituted a complete universe frame.

8.4.2 The Sample Design

There are several stages of sampling involved in the NOES design:

- A sample of 98 PSU's out of the universe of 604 NOES PSU's was designated;
- Within the sample PSU's, the DMI establishments with less than 1,000 employees were classified into 9 size strata according to the number of employees reported by DMI and a self-weighting sample of units within each size class was designated to be contacted by telephone;
- On the basis of the telephone interview, the sample of units to be visited for the survey was identified;
- The process from step (2) for establishments with 1,000 or more employees was repeated by selecting a self-weighting sample of units within two size classes without regard to the location of the unit.

The increase in variance (the loss in precision) of NOES estimates due to the introduction of the first of these steps can be measured by the replication variance computation process incorporated in VARNOES by the process discussed above in Section 8.3.2.

Table 8-1. Average number of survey hours to complete sample establishments in NOHS I and in NOES by establishment size

Size class (employees at survey) (1)	Average number of surveyor hours NOHS I (2)	NOES (3)
8 - 19	3.18	3.85
20 - 49	4.28	4.50
50 - 99	5.82	5.58
100 - 249	8.66	7.16
250 - 499	14.36	9.60
500 - 999	26.79	14.37
1000 - 1499	49.78	21.86
1500 - 2499	66.52	
2500 - 4999	86.16	30.84
5000 +	189.25	55.85

Table 8-2. Number of surveyors available during field interview phase of NOES

Number of surveyors [a]	Number of months	Implied surveyor-months
3	1	3
4	1	4
9	1	9
10	4	40
11	5	55
12	1	12
13	2	26
14	2	28
15	3	45
17	2	34
19	4	76
20	1	20
21	1	21
22	4	88
Total	32	461

[a] Number of surveyors working during any part of a calendar month.

9. VARNOES SOFTWARE PACKAGE

9.1 Introduction

The VARNOES software package utilizes two computer procedures that Westat developed for the National Accident Sampling System (NASS). NASSTIM produces national estimates of user-selected characteristics and displays the number of missing values and weighted missing values. NASSVAR uses the method of half-sample replication to produce for each user-selected characteristic:

1. National estimate,
2. Number missing for replication zero,
3. Weighted missing for replication zero,
4. Variance,
5. Relvariance,
6. Coefficient of Variation (%),
7. Standard Error,
8. Lower 95 percent confidence interval, and
9. Upper 95 percent confidence interval.

The user-selected characteristics may include totals, ratios and many other arithmetic functions of data items appearing on the questionnaires completed at the NOES sample facilities.

These procedures require the use of the Statistical Analysis System (SAS). Consequently, the VARNOES package includes, in addition to NASSTIM and NASSVAR, a COBOL program that generates the SAS program that executes the two procedures and an IBM Job Control Language (JCL) procedure that ties the entire process together. The user need only invoke the JCL procedure with a set of data and procedure specifications for the program and a file of questionnaire items that are to be analyzed.

9.2 Using VARNOES

VARNOES requires two computer files as input: a file of weights and questionnaire data values, and a file of program specifications.

9.2.1 Data File

The data file must consist of fixed length records, one record for each NOES sample facility. Each record must contain a field for each questionnaire data item that is to be analyzed. Each record must also contain the publication weight (replication zero) for the facility that the record represents. This is sufficient if only the estimation procedure, NASSTIM, is to be used. If the variance procedure, NASSVAR, is to be used, all of the half-sample weights must also be included, one for each of the 32 half-sample replications.

9.2.2 Specification File

VARNOES requires two types of specifications: data specifications that describe the format of the data file records and procedure specifications that direct the processing of the data, both the manipulation of the data and calls to the NASSTIM or NASSVAR procedures. The specification file must consist of 80 byte records with the following format:

Col. 1	Specification code, "D" for data and "P" for procedure;
Cols. 2-72	Specification, free format as described below;
Cols. 73-80	Record numbers or any other identification.

1. Data Specifications

Each data specification record must contain the following elements, in the order indicated but in any columms up to and including column 72 (the only exception to this is the specification code which must be in column 1):

D @ position name format

where position = beginning position of the field,

name = data item name, up to 8 characters,

format = w.d or $CHARw.

where w.d indicates a numeric field of total width w and for which the right most d positions represent a decimal fraction, and where $CHARw. indicates a character field of width w. The periods must be included as shown in either format.

For example,

```
D    @8   WESTID    $CHAR14.
D    @23 W0         7.3
```

specifies that a field for character data named WESTID starts in column 8 and is 14 bytes long, and that a field for numeric data named W0 starts in column 23, is 7 bytes long and the right most 3 bytes represent the decimal fraction portion of the field, i.e., the field is interpreted as PIC 9(4)V9(3).

As a convenience, arrays may be specified using name lists and format lists:

```
D    @ position (firstelement - lastelement) (format)
```

For example,

```
D    @31  (W1-W32)  (7.3)
```

specifies that an array W starts in column 31, consists of 32 7-byte fields, and the right most 3 bytes of each field represent the decimal fraction portion of the field, i.e., the fields are interpreted as PIC 9(4)V9(3) OCCURS 32 TIMES.

Exhibit 13 contains the copy of a listing of a sample run of VARNOES. The data specifications that were used for that run are as follows:

```
D    @1   FACNUM    $CHAR6.
D    @8   WESTID    $CHAR14.
D    @23  W0         7.3
```

```
D    @31  (W1-W32)  (7.3)
D    @256 SIC2DGT   2.0
D    @259 EMPS      5.0
D    @265 INJWL     3.0
D    @269 INJNWL    3.0
D    @279 HYGSERV   1.0
D    @281 SAFSERV   1.0
```

These data items will also be referred to in the following section.

2. Procedure Specifications

Procedure specifications are statements consisting of clauses, each statement ending in a semicolon(;). More than one statement may be written on a single specification record and a statement may be continued on one or more consecutive records; the specifications are written in free format with the only requirement that every procedure specification record have a "P" in column 1. Comments may be entered between "/*" and "*/" any place in the specification field that a blank is allowed.

If only the input variables are to be processed and if no manipulation of the input variables is to be done, then NASSTIM and/or NASSVAR may be called using the following formats:

```
P    PROC NASSTIM   DATA=DATA1    BEST;
P         VARIABLES variablenames;
```

```
P         WEIGHT    W0;
P         TITLE     title or 'title';
```

and

```
P    PROC NASSVAR   DATA=DATA1   BEST TOTAL;
P         VARIABLES          variablenames;
P         WEIGHT             W0  W1-W32;
P         TITLE              title or 'title';
```

where variablenames is the list of all variables that are to be analyzed by the called procedure, where title may contain up to 40 characters and where the single quote marks for title are required only when title has one or more semicolons or equal signs embedded in it.

For example,

```
P    PROC NASSTIM   DATA=DATA1     BEST;
P         VARIABLES HYGSERV SAFSERV;
P         WEIGHT    W0;
P         TITLE     PLANTS WITH HYGIENE/SAFETY SERVICES;
```

provides national estimates, number of missing values, and weighted missing values, of facilities offering hygiene services and of facilities offering safety services.

```
P    PROC NASSVAR   DATA=DATA1   BEST TOTAL;
P         VARIABLES EMPS;
P         WEIGHT    W0  W1-W32;
P         TITLE     'U.S. TOTALS; TOTAL PAYROLL';
```

provides the nine statistics mentioned in Section 9.1 for total number of employees.

If any manipulation of the input data is to be done, or if any auxilary variables are to be created, it must be done before the first call to NASSTIM, NASSVAR, or any other SAS procedures (examples are given below). An auxilary variable is created simply by using it in a procedure specification. It is not declared in a data specification. Data specifications are used only for input variables. For example,

```
P  TOTINJ = INJWL + INJNWL; /* TOTAL INJURED */
```

defines a variable TOTINJ that will contain the sum of data items INJWL and INJNWL for each facility. TOTINJ may now be used in any following procedure specification or in the variable list of any call to NASSTIM, NASSVAR, or any other SAS procedure. Note the use of a comment.

An IF or an IF-THEN statement may be used to selectively keep or delete a facility, or to selectively modify or define input or auxilary variables. For example,

```
P    IF   EMPS > = 50;
```

selects only those facilities with 50 or more employees for analysis, while

```
P    IF   SIC2DGT < 15 THEN DELETE;
```

deletes all facilities with two-digit SIC less than 15 from the analysis. An ELSE statement may be used with an IF-THEN statement, and may in fact contain another IF-THEN statement, i.e., IF-THEN and ELSE statements may be nested.

For example,

```
P    IF EMPS  < 500 THEN SIZE = 1;
P    ELSE IF EMPS  < 2500 THEN SIZE = 2;
P         ELSE SIZE = 3;
```

If a THEN clause or an ELSE statement contains more than one statement, a DO-END group must be used to contain those statements.

For example,

```
P    IF   HYGSERV = 1 OR SAFSERV = 1
P    THEN DO;
P         HYGSAFTY = 1;
P         HYGSFEMP = EMPS;
P         HYGSFINJ = TOTINJ;
P         END;
P    ELSE DO;
P         HYGSAFTY = 0;
P         HYGSFEMP = 0;
P         HYGSFINJ = 0;
P         END.
```

Note that the IF-THEN statement was continued to the second line and, consequently, the first line did not end in a semicolon. Note also that the auxilary variable HYGSAFTY is conditionally assigned a value of one for each facility, i.e., on the condition that the facility offer hygiene services or safety services.

The national estimate of HYGSAFTY would then yield the estimated number of facilities offering one or the other or both of those services. If a variable were unconditionally assigned a value of one for each facility, for example,

```
P    PLANTS = 1;
```

then the national estimate for that variable would be the estimated number of facilities. Of possible interest would be the ratio of these two estimates. Using NASSTIM or NASSVAR to obtain this ratio will yield the additional statistics mentioned in Section 9.1. Note that it is the ratio of the estimates that is wanted and not the weighted total of the ratios since for each facility, HSRATIO defined as

```
P    HSRATIO = HYGSAFTY/PLANTS;
```

would always yield HYGSAFTY because PLANTS is always equal to one. To obtain the ratio after the estimates are computed, the OUTPUT, OUTDATA, COMPVAR and OUTVAR statements must be used and the ratio defined in the actual call to NASSTIM or NASSVAR:

```
P    PROC NASSVAR   DATA=DATA1  BEST TOTAL OUTPUT OUTDATA=REPS
P         VARIABLES HYGSAFTY PLANTS;
P         WEIGHT    W0 W1-W32;
P         TITLE     HYGIENE AND SAFETY SERVICES;
P         COMPVAR   HYGSAFTY PLANTS;
P         OUTVAR    HSRATIO;
P         HSRATIO = HYGSAFTY/PLANTS;
```

The OUTPUT clause requests that the replication level estimates be saved; the OUTDATA clause names the internal data set that contains them. The COMPVAR statement names the variables whose estimates are to be used in a computation. The OUTVAR statement names the result or results of such computations. The statement or statements following the OUTVAR statement define the calculation of the results. The OUTPUT and OUTDATA clauses are required for such calculations involving estimates, but may also be used in the absence of such calculations, i.e., when the COMPVAR and OUTVAR statements are not being used. To display the replication level estimates, one may use the SAS procedure:

```
P    PROC      PRINT      DATA=REPS;
P              TITLE      REPLICATE LEVEL ESTIMATES;
```

To obtain estimates and variances on the basis of categories of facilities one must define categorical variables, called BY variables, sort the data on the basis of these variables, and indicate these variables in the call to NASSTIM or NASSVAR.

For example, the auxilary variable SIZE defined above could serve as a BY variable. Assuming that it had been so defined, we could then specify:

```
P    PROC SORT      DATA=DATA1   BY SIZE;
P    PROC NASSTIM   DATA=DATA1    BEST;
P         VARIABLES EMPS;
P         BY        SIZE;
P         WEIGHT    W0;
P         TITLE     TOTAL PAYROLL BY FACILITY SIZE;
```

More than one BY variable may be used. For example,

```
P    PROC SORT      DATA=DATA1    BY SIZE SIC2DGT;
P    PROC NASSVAR   DATA=DATA1    BEST TOTAL;
P         VARIABLES HYGSERV SAFSERV;
P         BY        SIZE SIC2DGT;
P         WEIGHT    W0  W1-W32;
P         TITLE     HYGIENE AND SAFETY SERVICES;
```

would yield the statistics of Section 9.1 computed for each two digit SIC within size category. Replacing BY SIZE SIC2DGT; with BY SIC2DGT SIZE; would yield those statistics for each size category within two digit SIC.

The label statement may be used to associate labels of up to 40 characters to one or more variables. This allows one to give more meaning in the output listing than is possible with the variable names alone, which are limited to 8 characters. For example,

```
P    LABEL     PLANTS=U.S. FIRMS
P              EMPS=EMPLOYEES
P              SIC2DGT=TWO DIGIT SIC CODE
P              HYGIENE=PLANTS WITH HYGIENE SERVICE
P              HYGEMPS=EMPLOYEES AT HYGIENE PLANTS
P              TOTINJ=TOTAL INJURED
P              W0=PUBLICATION WEIGHT;
```

Note that the above is a single statement, therefore, the semi-colon appears only at the end of the last line.

9.3 Example

Exhibit 13 is a listing of a sample computer run using the VARNOES procedure with many of the specifications described in Section 9.2. The complete specifications used in that run are displayed in Figure 9-1.

A listing of the VARNOES procedures itself is in Figure 9-2. A model of the minimum JCL required to run VARNOES is given in Figure 9-3. The Parklawn Computing Center requires that the names of cataloged procedures begin with the computer account initials. Therefore in the example of Exhibit 13 the procedure name was changed from VARNOES to BITVAR. If VARNOES is to be run as a cataloged procedure under a different set of account initials, the procedure name would have to be changed again to conform to the computer center requirements. The changes would occur in line 20 of Figure 9-2 and line 30 of Figure 9-3. If the name of the VARNOES library is changed, that change would have to be made in line 180 of Figure 9-2.

Figure 9-4 shows how to use VARNOES as an in-line procedure. There are no special requirements for the names of in-line procedures. A change in the VARNOES library would necessitate a change in line 200 of Figure 9-4.

In both forms of the procedure, the output of the computer run is directed to the standard print queue (SYSOUT = A, HOLD = NO) by default. If examination of the output is a terminal via LISTJES is desired, the print queue of the output may be changed in line 30 of Figure 9-2 or line 370 of Figure 9-4 to:

```
//GO  EXEC  VARNOES,SYSOUT=T,HOLD=YES
```

```
D @1    FACNUM        $CHAR6.                                          00000010
D @8    WESTID        $CHAR14.                                         00000020
D @23   W0             7.3                                           00000030
D @31   (W1-W32)      (7.3)                                          00000040
D @256 SIC2DGT         2.0                                           00000050
D @259 EMPS            5.0                                           00000060
D @265 INJWL           3.0                                           00000070
D @269 INJNWL          3.0                                           00000080
D @279 HYGSERV         1.0                                           00000090
D @281 SAFSERV         1.0                                           00000100
P TOTINJ = INJWL + INJNWL ;          /* TOTAL INJURED */             00000110
P IF    EMPS >= 50 ;                                                 00000120
P IF    SIC2DGT < 15 THEN DELETE ;                                   00000130
P IF    EMPS < 500 THEN SIZE = 1 ;                                   00000140
P ELSE IF   EMPS < 2500 THEN SIZE = 2 ;                              00000150
P       ELSE SIZE = 3 ;                                              00000160
P IF    HYGSERV = 1 OR SAFSERV = 1                                   00000170
P THEN DO ;                                                          00000180
P       HYGSAFTY = 1 ;                                               00000190
P       HYGSFEMP = EMPS ;                                            00000200
P       HYGSFINJ = TOTINJ ;                                          00000210
P       END ;                                                        00000220
P ELSE DO ;                                                          00000230
P       HYGSAFTY = 0 ;                                               00000240
P       HYGSFEMP = 0 ;                                               00000250
P       HYGSFINJ = 0 ;                                               00000260
P       END ;                                                        00000270
P PLANTS = 1 ;                       /* COUNT FACILITIES */          00000280
P LABEL PLANTS = U.S. FIRMS                                          00000290
P       EMPS = EMPLOYEES                                             00000300
P       SIC2DGT = TWO DIGIT SIC CODE                                 00000310
P       HYGSFEMP = EMPS AT PLANTS WITH HYGIENE SERVICE               00000320
P       HYGSFINJ = INJURED AT HYGIENE PLANTS                         00000330
P       HYGSAFTY = PLANTS WITH HYGIENE OR SAFETY SERVICE             00000340
P       TOTINJ = TOTAL INJURED                                       00000350
P       W0 = PUBLICATION WEIGHT ;                                    00000360
P PROC NASSVAR DATA=DATA1 BEST TOTAL OUTPUT OUTDATA=REPS ;           00000370
P       VARIABLES EMPS HYGSFEMP HYGSFINJ HYGSAFTY PLANTS ;           00000380
P       WEIGHT    W0 W1-W32 ;                                        00000390
P       TITLE     'HYGIENE AND SAFETY: U.S. TOTALS' ;                00000400
P       COMPVAR   HYGSAFTY PLANTS ;                                  00000410
P       OUTVAR    HSRATIO ;                                          00000420
P       HSRATIO = HYGSAFTY / PLANTS ;                                00000430
P PROC PRINT DATA=REPS; TITLE REPLICATE LEVEL ESTIMATES ;            00000440
P PROC SORT DATA=DATA1 ;  BY SIZE ;                                  00000450
P PROC NASSTIM DATA=DATA1 BEST ;                                     00000460
P       VARIABLES EMPS PLANTS ;                                      00000470
P       BY        SIZE ;                                             00000480
P       WEIGHT    W0 ;                                               00000490
P       TITLE     TOTAL EMPLOYEES & PLANTS BY PLANT SIZE ;           00000500
```

Figure 9-1. Specifications for Exhibit 13

```
//*
//VARNOES PROC ENTRY=SAS,OPTIONS=,SYSOUT=A,HOLD=NO                      00000010
//*                                                                     00000020
//SASGEN  EXEC PGM=SASGEN                                               00000030
//STEPLIB  DD  DSN=USH.BIT.VARNOES.LOAD,DISP=SHR                        00000040
//SYSOUT   DD  SYSOUT=&SYSOUT,HOLD=&HOLD                                00000050
//SYSPRINT DD  SYSOUT=&SYSOUT,HOLD=&HOLD                                00000060
//SAS1     DD  DSN=&&SAS1,DISP=(NEW,PASS),                              00000070
//             UNIT=SYSDA,SPACE=(80,(500,100)),                         00000080
//             DCB=(RECFM=FB,LRECL=80,BLKSIZE=3120)                     00000090
//*                                                                     00000100
//SASGO   EXEC PGM=&ENTRY,PARM='&OPTIONS',REGION=384K                   00000110
//LIBRARY  DD  DSN=&LIBRARY,DISP=(MOD,PASS),                            00000120
//             UNIT=SYSDA,SPACE=(CYL,(1,,20))                           00000130
//STEPLIB  DD  DSN=PCC.SAS823.LIBRARY,DISP=SHR                          00000140
//             DD  DSN=SYS1.PLIBASE,DISP=SHR                            00000150
//             DD  DSN=*.LIBRARY,VOL=REF=*.LIBRARY,DISP=(OLD,PASS)      00000160
//             DD  DSN=USH.BIT.VARNOES.LOAD,DISP=SHR,DCB=BLKSIZE=19069  00000170
//SASHELP  DD  DSN=PCC.SAS823.SASHELP,DISP=SHR                          00000180
//WORK     DD  UNIT=SYSDA,SPACE=(6160,(500,200),,,ROUND)                00000190
//FT11F001 DD  SYSOUT=&SYSOUT,HOLD=&HOLD,                               00000200
//             DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)                    00000210
//FT12F001 DD  SYSOUT=&SYSOUT,HOLD=&HOLD,                               00000220
//             DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)                    00000230
//FT13F001 DD  SYSOUT=B,DCB=(RECFM=F,BLKSIZE=80)                        00000240
//FT14F001 DD  DUMMY                                                    00000250
//FT15F001 DD  UNIT=SYSDA,SPACE=(400,(100,300)),                        00000260
//             DCB=(RECFM=FB,LRECL=80,BLKSIZE=400,BUFNO=1)              00000270
//MAPS     DD  DSN=PCC.SAS823.MAPS,DISP=SHR                             00000280
//ZIP      DD  DSN=PCC.SAS.ZIPCODES.COUNTIES,DISP=SHR                   00000290
//SYSIN    DD  DSN=&&SAS1,DISP=(OLD,DELETE)                             00000300
//*                                                                     00000310
                                                                        00000320
```

Figure 9-2. VARNOES procedure

```
//  JOB  CARD                                                                  00000010
/*  JES2 CARDS                                                                 00000020
//GO       EXEC VARNOES                                                       00000030
//SASGEN.SPEC1 DD  DISP=SHR,DSN=    (DATA & PROCEDURE SPECIFICATIONS)         00000040
//SASGO.DATA1  DD  DISP=SHR,DSN=    (WEIGHTS & DATA)                          00000050
//                                                                           00000060
```

Figure 9-3. Model JCL for VARNOES

```
//  JOB  CARD                                                                  00000010
/*  JES2 CARDS                                                                 0000002^
//*                                                                            0000003
//VARNOES PROC ENTRY=SAS,OPTIONS=,SYSOUT=A,HOLD=NO                            0000004v
//*                                                                            00000050
//SASGEN  EXEC PGM=SASGEN                                                      0000006
//STEPLIB  DD  DSN=OSH.BIT.VARNOES.LOAD,DISP=SHR                               0000007
//SYSOUT   DD  SYSOUT=&SYSOUT,HOLD=&HOLD                                       00000080
//SYSDBOUT DD  SYSOUT=&SYSOUT,HOLD=&HOLD                                       0000009^
//SAS1     DD  DSN=&&SAS1,DISP=(NEW,PASS),                                     0000010
//              UNIT=SYSDA,SPACE=(80,(500,100)),                               0000011v
//              DCB=(RECFM=FB,LRECL=80,BLKSIZE=3120)                           00000120
//*                                                                            0000013
//SASGO   EXEC PGM=&ENTRY,PARM='&OPTIONS',REGION=384K                          0000014
//LIBRARY  DD  DSN=&LIBRARY,DISP=(MOD,PASS),                                   00000150
//              UNIT=SYSDA,SPACE=(CYL,(1,,20))                                 00000160
//STEPLIB  DD  DSN=PCC.SAS823.LIBRARY,DISP=SHR                                 0000017
//           DD  DSN=SYS1.PLIBASE,DISP=SHR                                     0000018
//           DD  DSN=*.LIBRARY,VOL=REF=*.LIBRARY,DISP=(OLD,PASS)               00000190
//           DD  DSN=OSH.BIT.VARNOES.LOAD,DISP=SHR,DCB=BLKSIZE=19069           0000020^
//SASHELP  DD  DSN=PCC.SAS823.SASHELP,DISP=SHR                                 0000021
//WORK     DD  UNIT=SYSDA,SPACE=(6160,(500,200),,,ROUND)                       0000022v
//FT11F001 DD  SYSOUT=&SYSOUT,HOLD=&HOLD,                                      00000230
//              DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)                          0000024
//FT12F001 DD  SYSOUT=&SYSOUT,HOLD=&HOLD,                                      0000025
//              DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)                          00000260
//FT13F001 DD  SYSOUT=B,DCB=(RECFM=F,BLKSIZE=80)                               0000027^
//FT14F001 DD  DUMMY                                                           0000028
//FT15F001 DD  UNIT=SYSDA,SPACE=(400,(100,300)),                               0000029v
//              DCB=(RECFM=FB,LRECL=80,BLKSIZE=400,BUFNO=1)                    00000300
//MAPS     DD  DSN=PCC.SAS823.MAPS,DISP=SHR                                    0000031
//ZIP      DD  DSN=PCC.SAS.ZIPCODES.COUNTIES,DISP=SHR                          0000032
//SYSIN    DD  DSN=&&SAS1,DISP=(OLD,DELETE)                                    00000330
//*                                                                            00000340
//         PEND                                                                0000035
//*                                                                            0000036
//GO       EXEC VARNOES                                                        00000370
//SASGEN.SPEC1 DD  DISP=SHR,DSN=     (DATA & PROCEDURE SPECIFICATIONS)         0000038^
//SASGO.DATA1  DD  DISP=SHR,DSN=     (WEIGHTS & DATA)                          0000039
//                                                                             0000040v
```

Figure 9-4. VARNOES used as an in-line procedure

REFERENCES

1. U.S. Bureau of the Census. County Business Patterns, 1979, United States: Appendix B. Washington, DC: Government Printing Office, 1980.

2. U.S. Bureau of the Census. County Business Patterns, 1980, United States. Washington, DC: Government Printing Office, 1982.

3. U.S. Department of Labor, Bureau of Labor Statistics. Employment and Earnings (Table B-2). Washington, DC: Government Printing Office, January 1981.

4. U.S. Department of Health, Education and Welfare. National Occupational Hazard Survey: Vol. III Survey Analysis and Supplemental Tables. Prepared by the Public Health Service Center for Disease Control, National Institute for Occupational Safety and Health (Division of Surveillance Hazard Evaluations and Field Studies, Cincinnati, Ohio). Washington, DC: Government Printing Office.

5. McCarthy, Philip J. Replication: An Approach to the Analysis of Data from Complex Surveys. Prepared for the U.S. Department of Health, Education and Welfare, National Center for Health Statistics. Washington, DC: Government Printing Office, Ser. 2, No. 14, 1966.

6. ____. Psuedoreplication: Further Evaluation and Application of Balanced Half-Sample Technique. Prepared for the U.S. Department of Health, Education and Welfare, National Center for Health Statistics. Washington, DC: Government Printing Office, Ser. 2, No. 31, 1969.

7. Hansen, M.H.; Hurwitz, W.N.; Madow, W.G. Sample Survey Methods and Theorys: Volume II Theory. New York: John Wiley and Sons, 1953, p. 218.

EXHIBITS

WESTAT
An Employee-Owned Research Corporation

MEMORANDUM

TO: J. Waksberg July 22, 1980

FROM: R. Hanson

SUBJECT: Optimum sample size for NIOSH NOHS II (Revised*)

I. Introduction

The purpose of this memorandum is to record a method for determining the size and the distribution of the sample for the NOHS II survey and to discuss some elements of the sampling procedure implied by this method.

The following broad assumptions are incorporated in the proposed method; we introduce further and more specific assumptions later in the derivation:

(1) The major objective of the survey is to provide information on the characteristics of all production employees in a set of industries called target industries. Important, although lower priority, interest lies in providing information for employees in subsets of target industries, in selected occupation categories across target industries, and information on the physical description of the work locations.

(2) The study covers a wide range of hazards, and it is not possible to define a subset of industries in which the hazards are concentrated. As a result, the target industries will not be put into groups according to assumed classifications of risk of exposure to hazards.

(3) The data collection procedure is to be accomplished by methods very similar to the NOHS I process, and we expect relative costs (in terms of person-hours) for investigation in that earlier study to be repeated in NOHS II.

*Revised to incorporate tabulations of 1979 CBP records in Table 1.

(4) The costs per PSU in sample are expected to have little, if any impact on deriving the optimum sample within PSU's. Investigator team travel between PSU's has been ignored (this travel is to be performed on weekends), and the cost of selecting sample establishments is assumed free of a PSU component.

II. Formulas for an Optimum Sample

A. Notation

The following notation is used:

Let N_a be the total number of establishments in the U.S. in all target industries in the a^{th} employee size class;

n_a the number of establishments selected with equal probability from N_a;

C_a the average cost (in person-hours) to investigate a sample establishment in the a^{th} employee-size category;

$C = \sum_a n_a C_a$ cost of investigating the $n = \sum_a n_a$ sample establishments in terms of person-hours.

$\bar{y}'_a$ the estimated average value of the y characteristic per establishment in the a^{th} size class based on the sample of n_a facilities in that class; and

$y' = \sum_a N_a \bar{y}'_a$;

$s^2(\bar{y}_a)$ the estimated population variance of $\bar{y}'_a$. Then,

$$\sigma^2(y') = \sum_a N_a^2 (1/n_a - 1/N_a) s^2(\bar{y}_a) \qquad (1)$$

is the variance of the estimated total y'.

B. Optimum Sample Sizes

With the given notation, the formulas for the optimum values of n_a are determined in the usual way by writing

$$\phi = \sigma^2(y') + \lambda \left(\sum_a n_a C_a - C\right) \tag{2}$$

taking partial derivatives of ϕ with respect to each n_a, setting them equal to zero, and solving the resulting simultaneous equations for $\sqrt{\lambda}$ and then for n_a. The result is

$$n_a = \frac{N_a s(\bar{Y}_a)}{\sqrt{C_a}} \frac{C}{\sum_a \sqrt{C_a}\, N_a s(\bar{Y}_a)} \tag{3}$$

III. Sampling Procedure

A. Measure of Size for Sampling Purposes

The previous section shows how to compute n_a/N_a, the sampling fraction for selecting establishments in the a^{th} size class. One of these fractions will be the smallest, let this fraction n_1/N_1 be defined $1/k = n_1/N_1$. This fraction is likely to be (although not necessarily) for the class made up of the large number of smallest establishments. Define ratios of sampling fractions for each size class:

$$n_a/N_a = f_a/k,$$

then,

$$n = \sum_a n_a = \frac{1}{k} \sum_a N_a f_a.$$

We now define a measure of size at the U. S. level for all target industries:

$$M = \sum_a N_a f_a \qquad (4)$$

From this it follows that $1/k = n/M$ will be the sampling fraction for the class having the smallest proportion in the sample and for other classes, the sampling fraction is given by f_a/k.

B. Sampling Within Sample PSU's

The told measure of size defined in (4) can be defined for each of the PSU's in each stratum of PSU's by

$$M_{hj} = \sum_a N_{hja} f_a$$

where N_{hja} is defined as for N_a but for the j^{th} PSU in the h^{th} stratum. The measure M_{hj} and the measure of the h^{th} stratum

$$M_h = \sum_j M_{hj}$$

are used to determine the probability of selecting the PSU from the stratum. A self-weighting sample of establishments for the size class having the smallest sampling fraction is defined using probabilities that satisfy the expression:

$$\frac{1}{k} = \left(\frac{M_{hj}}{M_h}\right) \times \left(\frac{M_h}{M_{hj}} \times \frac{1}{k}\right).$$

For the a^{th} size class, the probabilities in the following expression define the sampling system:

$$\frac{f_a}{k} = \left(\frac{M_{hj}}{M_h}\right) \times \left(\frac{M_h}{M_{hj}} \quad \frac{f_a}{k}\right). \qquad (5)$$

The first term on the right of (5) is the probability of selecting the j^{th} PSU out of the h^{th} stratum, and the second term is the probability to be used for selecting establishments in the a^{th} class from the hj^{th} sample PSU.

IV. Application of procedure to NOHS II

In this section, we compute the total sample n, the sampling fractions f_a/k and the expected sample sizes n_a. This will provide guidelines for the sampling of establishments in the certainty PSU's. These computations require us to make certain approximations. In the paragraphs that follow, we explain the source of these approximations.

In a later memorandum, we discuss the additional problems that arise in determining the number of sample first-stage units.

A. Discussion of Table 1

Application of the theory discussed in Section III produces the results summarized in the attached Table 1. The information in the table is interpreted as follows:

(Column 2) Target establishments, N_a

The numbers of establishments by size class for the target industries are based on the 1979 County Business Patterns (CBP) publications. For this survey, we will be applying the resulting sampling fractions to the 1980 Dun and Bradstreet (D&B) files, but the CBP data are used in these computations because the levels of the counts are assumed to be nearer to the current levels (i.e., duplicates and dead records appearing in D&B are excluded). CBP establishments in the 8-19 employee class are approximated by allocating of the count of the CBP published 5-9 employee establishments according to the proportion of firms having 8-9 employees in D&B plus the establishments in the published 10-19 employee size class.

(Column 3) Survey Cost (hours), C_a

These numbers come from a special tabulation of NOHS I information prepared by NIOSH and show the average number of person hours used in investigating accidents of the given employee size class in that earlier survey.

The total cost available to investigate all establishments in the sample is the number of person hours to be obtained from 21 investigators working for two years. The investigators will be grouped into five teams, each with a team leader who produces 80 percent of the production of the other team members (the other 20% of the leader's time is taken up by supervisory responsibilities). Each team member is assumed to work a 40-hour week for 48 weeks of the year, the remaining four weeks are taken up by annual and sick leave and holiday time. The hours contributed by the five team leaders will be $5 \times 40 \times 48 \times .8 \times 2 = 15{,}360$ hours and from the others $16 \times 40 \times 48 \times 2 = 61{,}440$ hours, for a total of 76,800 hours.

(Column 4) Average employees per establishment

These numbers have been computed from published CBP employee totals for size classes below 1,000. For larger classes the average was taken at 0.4 of the size of the interval plus the lower bound. For the class of 5,000 and over, an average was taken such that the total employees in all size classes approximated the published totals for all classes.

(Column 5) Variance Factor

As shown in expression (3) of Section II, the term $s(\bar{y}_a)$ appears in both the numerator and denominator of the formula for n_a; hence, it is sufficient to have an approximation to the relative sizes of the $s(\bar{y}_a)$ rather than their absolutes. That is,

$$V^2(\bar{y}_a) \doteq s^2(\bar{y}_a)/(\bar{y}_a)^2 \doteq \text{constant; from which it}$$

follows that

$$s(\bar{y}_a) \doteq (\bar{y}_a)\sqrt{\text{constant}}.$$

We have assumed that, except for the smallest and largest size classes, the relvariances are about the same for the size classes. This assumption is often made when actual numbers are not available; the assumption usually produces reasonably good approximations to the sample sizes computed from more precise information. We have asked for computed values of $V^2(\bar{y}_a)$ for a set of characteristics produced in the NOHS I study; they will be available later to check this assumption.

(Column 6) Facilities in sample, n_a

The sample has been computed using expression (3) of Section II.

(Column 7) Sampling fraction, 1 in N_a/n_a

This is the sampling fraction n_a/N_a expressed as "1 in (N_a/n_a)".

(Column 8) f_a

The term f_a is the ratio of the sampling fraction for the a^{th} class to the fraction for the first class, i.e., 1 in 199.53. Thus, for the second class (1/125.25)/(1/199.53) = 1.593.

B. Comments

(1) Many uncertainties have been accepted in determining the total sample size at an expected 4,895 CBP establishments. We are not sure whether a sample of this or a larger number is needed to keep the investigators fully occupied. We therefore intend to prepare a larger sample than shown in Table 1 and assign the sample in a random order such that the establishments for which the investigation is completed will constitute a random sample. We intend to increase the total sample size by 25 percent for this purpose to yield a total expected CBP sample of about 6,000 establishments.

(2) The question of how to define the universe of first stage sampling units, deciding the number of first stage units to have in the sample, and whether to sample establishments without regard to location within sample first stage units must be decided before the general survey outline is complete. These issues are discussed in a separate memorandum.

cc: J. Edmonds
J. Blais
C. McElrea
M. Tevelow
D. Ward

Table 1. Computation and distribution of sample of 4,895 CBP establishments for NOHS II

Employee size class	Target establishments N_a [1]	Survey cost (hrs.) C_a [2]	Average employees per facilities	Variance factors	Facilities in sample n_a	Sampling fraction: 1 in N_a/n_a:	f_a
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
8-19	237,445	3.18	12.4	2	1,190	199.53	1.0
20-49	114,508	4.28	32.4	1	914	125.25	1.593
50-99	44,567	5.82	71.7	1	675	66.03	3.022
100-249	30,601	8.66	158.1	1	838	36.52	5.464
250-499	10,887	14.36	349.9	1	512	21.26	9.384
500-999	5,056	26.79	690.1	1	343	14.70	13.576
1000-1499	1,424	49.78	1,200	1	123	11.58	17.235
1500-2499	906	66.52	1,900	1	108	8.389	23.785
2500-4999	520	86.16	3,500	1	94[3]	5.545	35.984
5000+	212	189.25	9,250	2	97[3]	2.190	91.110
Total	446,125				4,895		

[1] Based on tabulation of CBP county summary records for 1979.

[2] Person hours per establishment required to investigate facilities in NOHS I.

[3] Proportion of total investigator time to cover establishments of this size
= (94 x 86.16 + 97 x 189.25)/76800 = .344

WESTAT
An Employee-Owned Research Corporation

MEMORANDUM

TO: J. Blais August 1, 1980

FROM: J. Edmonds J.E.

SUBJECT: Initial Summaries of Firms from CBP by County and Size Class

Prior to forming PSU's for the NIOSH sample, it is necessary to construct county summary records which will be used to define PSU's and calculate measures of size. No PSU's can be formed until a county summary file has been produced which contains, for each county covered on the CBP tape, the following data:

- State Code, and County Code;
- Number of CBP firms, by size class (only counts of firms with 8 or more employees), for all target SIC;
- Sequence numbers which will be used to mechanically link counties when PSU's are formed;
- Assorted census demographic information.

The initial programming for these county summary records should be directed at producing records with the first two categories of information, leaving space so that the other data may be added later. A possible format for the initial summary record is given below.

Col. 1-2	State code
Col. 3-5	County code
Col. 6-14	# firms with 8-19 employees
Col. 15-22	# firms with 20-49 employees
Col. 23-29	# firms with 50-99 employees
Col. 31-38	# firms with 100-249 employees
Col. 39-46	# firms with 250-499 employees
Col. 47-54	# firms with 500-999 employees
Col. 55-62	# firms with 1000-1499 employees
Col. 63-70	# firms with 1500-2499 employees
Col. 71-78	# firms with 2500-4999 employees
Col. 79-86	# firms with 5000+ employees

In each of these totals, there would be three implied decimal places.

It is not possible to produce these desired county records through direct aggregation of the CBP File. Two problems prevent the direct aggregation.

1. The CBP File does not have the 8-19 employee category. It does have 5-9 and 10-19. Thus, to form the 8-19 aggregate, it will be necessary to obtain the CBP 5-9 count, reduce it by multiplying by a specified fraction based on D&B counts (which will be different for each target SIC), and add the result to the CBP count for the 10-19 group.

2. The CBP File does not contain some target codes at the 4-digit level. For these, it will be necessary to aggregate a broader (2 or 3 digit) code for each size class, and then reduce the count by a specified fraction based on D&B counts.

The attached tables are to be used to obtain the SIC codes to be located by the computer for each county. Also included are the reduction fractions to be used to multiply the 5-9 size counts, prior to summing them with the 10-19 class. Finally, the fractions needed to reduce the counts from the overly broad screening codes are shown.

The computer programming to create these county summary records could proceed as described below.

1. Create an extract file from the CBP tape, which contains only the SIC codes of the attached table, for all counties. Also screen out all records pertaining to size class 1-4 employees, which will not be used. When the program locates records for the overly broad screening codes, the size class of the record should be determined by the machine. If the size class is 20-49 or greater, the record count should be multiplied by the appropriate reducing fraction.

2. From this abbreviated file, a sort should be made by state and county. Then the county aggregate may be created, being certain to multiply each 5-9 class record by its appropriate reduction, and forming the 8-19 class. For overly broad codes, the resultant 8-19 count must be further reduced as in paragraph 1 for the classes 20-49 or greater. For example the entry to approximate the (8-19) count for the sum of target SIC's 4212 and 4214 is computed by the following:

   [.257 (CBP count of 5-9) + (CBP count 10-19)] x .588

3. The table shows that one SIC type is to be aggregated in a different fashion. For each county, the count of the 7218 code must be stored and deducted from the 7200 count. The result of this subtraction is tabulated in the usual fashion.

JE:rb

cc: J. Waksberg
T. McKenna
R. Hanson
C. McElvea
B. Tevelow
D. Ward

Computer to search for SIC	Target SIC where different form (1)	RATIO (8-9)/(5-9)	8-19	20-49	50-99	100-249	250-499	500-999 and all higher classes
(1)	(2)	(3)	(4)					
0720	0723 0724	.233	.656	.785	.891	.895	.895	.895
0740	0742	.230	.888	.932	1.000	1.000	1.000	1.000
0780	0782 0783	.208	.486	.488	.457	.610	.714	.400
1300		.246	None					
1500		.209	None					
1600		.236	None					
1700		.236	None					
2000		.273	None					
2100		.167	None					
2200		.281	None					
2300		.256	None					
2400		.266	None					
2500		.271	None					
2600		.315	None					
2700		.247	None					
2800		.268	None					
2900		.233	None					
3000		.270	None					
3100		.242	None					
3200		.269	None					

Computer $ search for SIC	Target SIC where different from (1)	RATIO (8-9)/(5-9)	8-19	20-49	50-99	100-249	250-499	500-999 and all higher classes
5093		.245	None					
5160		.234	None					
5170		.244	None					
5520		.141	None					
5530		.211	None					
5540		.166	None					
7200 -7218	7200 7218	.213	None - Subtract out 7218 before multiplying by .213					
7330		.221	None					
7340		.256	None					
7391		.232	None					
7395		.238	None					
7397		.288	None					
7399		.219	None					
7510		.223	None					
7530		.202	None					
7540		.251	None					
7600		.206	None					
8060	8062	.253	.422	.690	.803	.874	.905	.899
8070		.245	None					
8090		.266	None					
8400		.264	None					

Computer to search for (SIC)	Target SIC where different from (1)	RATIO (8-9)/(5-9)	8-19	20-49	50-99	100-249	250-499	500-999 and all higher classes
3300		.296	None					
3400		.278	None					
3500		.274	None					
3600		.274	None					
3700		.261	None					
3800		.249	None					
3900		.250	None					
4110		.289	None					
4120		.291	None					
4150		.327	None					
4170		.268	None					
4210	4212 4214	.257	.588	.453	.326	.208	.180	.273
4220		.238	None					
4230		.342	None					
4400		.252	None					
4500		.221	None					
4600		.248	None					
4740	4742	.148	.667	.750	.818	.571	.857	.333
4780		.297	None					
4800		.336	None					
4900		.265	None					
5010		.248	None					
5030		.270	None					
5050		.246	None					

Exhibit 3

WESTAT
An Employee-Owned Research Corporation

MEMORANDUM

TO: J. Blais August 7, 1980

FROM: R. Hanson RH

SUBJECT: Initial Summaries of Firms from CBP by County and Size Class, Revised

This memorandum provides specifications for additional information that should be included in the county summary records that are to be combined into PSU summary records. This memorandum should be made a part of the specifications given in John Edmonds' memo to you dated 8/1/80, subject as above.

The additional information requested in this memo will be used in classifying PSU's into strata (after the PSU's have been defined).

This memorandum requires you, for each county, to compute the approximate number of employees in establishments within 21 subsets of the target industries. This will be done by: (1) Establish counts of the number of establishments by size class for each of 21 specified sets of 2-digit SIC's within the target SIC's, (2) For each establishment size class, within each SIC group, multiply the number of establishments in the class by an average number of employees per establishment in the size class, (3) Compute the sum of these products over all size classes for the SIC subgroup, and (4) Show this estimated total number of employees for the SIC group.

This means that the county (and eventually the PSU) summary record must be expanded to include the 21 new counts. Note that the 21 employee counts are to be summarized along with the other data fields when the county records are summarized into PSU records.

The 21 groups of 2 digit SIC's for which an approximate employee count is required are given in Table 1. The average number of employees to be assumed for each size class is given in Table 2; one set of these averages will be used for all SIC groups.

RH:rb

cc: J. Waksberg
T. McKenna
J. Edmonds
C. McElrea
B. Tevelow
D. Ward

Table 1. Combinations of NIOSH Two Digit SIC's for which Approximate Employee Estimates are Required.

Combination number	Two digit target SIC's in the combination*
1	13
2	15, 16, 17
3	20
4	21, 22, 23
5	24
6	25, 26
7	28
8	29, 30, 31
9	32
10	33
11	34
12	35
13	36
14	37
15	38, 39
16	42
17	48
18	49
19	50, 51, 55
20	73
21	80

*The following two-digit target SIC's do not appear in these combinations: 07, 27, 40, 41, 44-47, 72, 75, 76, 84

Table 2. Approximate Number of Employees per Establishment by Employee Size Class Among NIOSH Target Establishments.

Class	Employee size class	Approximate employees per establishment
11	NA	11.5
1	8-19	11.5 a/
2	20-49	28.8 a/
3	50-99	64.8 a/
4	100-249	143.5 a/
5	250-499	329.8 a/
6	500-999	700 b/
7	1000-1499	1200 b/
8	1500-2499	1900 b/
9	2500-4999	3500 b/
10	5000+	zero c/

a/ Based on tabulations of all (target and non-target SIC) establishments, 1980 D&B Market file.

b/ Value taken at 40 percent of the range

c/ Establishments of this size will be sampled without regard to PSU.

Exhibit 4

WESTAT
An Employee-Owned Research Corporation

MEMORANDUM

TO: J. Blais July 31, 1980

FROM: R. Hanson RH

SUBJECT: Construction of County and PSU summary tapes NIOSH

This memorandum specifies a set of computer procedures that define the universe of primary sampling units (PSU's) to be used in the NIOSH sample. Each PSU, a potential sample assignment area, must meet certain requirements, one of which is to have a large enough body of establishments to supply the needs of the sample that will eventually be sampled from it.

The procedure begins by defining a set of county summary records that carry County Business Pattern (CBP) counts of establishments in the NIOSH target industries, some limited data from the County Data Book (CDB), and a sequence number assigned to array counties in a geographically contiguous order.

The metro county records are combined into a summary record for each SMSA; we assume SMSA's will be acceptable assignment areas. Non-metro counties are tested to see whether they could provide the needed sample if selected; if not, contiguous counties are combined to meet this condition.

We have decided to construct PSU's that will supply a sample expected, on the average, to require a team about three weeks to complete. The team, assumed to comprise three investigators plus a supervisor, is expected to contribute an average of 139.7 person hours each week to productive investigation (allowing for leave and only 80% of the supervisor's time).

The procedure allows us to review the machine-prepared results for other desirable PSU characteristics and to change specific PSU compositions if we desire to.

1. Produce summary records for counties

1.1 Produce a summary record for each county in the 50 States. The record should include information from the following sources:

1.1.1 Sequence number

A 4-digit number has been clerically assigned so that, when sequenced by this number, adjacent counties in the order will be contiguous. For the record, the following interpretations may be given to the blocks of numbers.

0001 - 0285 Assigned to counties within the 285 SMSA's. All counties within an SMSA are given the same number.

1000 - ____ Assigned to non-metro counties in the NE census region.

2000 - ____ Non-metro counties in NC

4000 - ____ " " " " South

7000 - ____ " " " " West

9002 - 9058 " " " " Alaska

9500 - 9504 " " " " Hawaii

1.1.2 County Business Patterns (CBP)

Include the following information from the 1979 CBP:

- State and County code
- Number of target establishments by size class (approximate*); the size classes are:

Class	Number of employees
1	NA
2	8-19
3	20-49
4	50-99
5	100-249
6	250-499
7	500-999
8	1000-1499
9	1500-2499
10	2500-4999
11	5000+

*The methods of approximating the CBP counts by establishment size by county are recorded in a memo by J. Edmonds dated 8/1/80.

1.1.3 County Data Book

Include the following information from the 1977 CDB:

- o Land area (square miles)
- o Population (1975)
- o Population (1970)

1.2 Make the following computations and record the information on the county summary record.

1.2.1 Probability measures of size

$M_c = \sum_a f_a N_{ca}$ where

N_{ca} is the number of CBP establishments in the a^{th} size class in the c^{th} county and f_a is the oversampling ratio for the a^{th} size class. See column (8) of Table 1 of my memo to Waksberg dated 7/22/80.

Show three values of M_c:

M_c as defined above where the summation on a is over all size classes in the county,

M_{c1} computed by omitting all contributions from size class 11, and

M_{c2} computed by omitting all contributions from size classes 10 and 11

1.2.2 Cost measure of size

$K_c = \sum_a C_a f_a N_{ca}$ where

C_a is the average person hours to investigate an establishment selected from the a^{th} size class, see column (3) of Table 1 of my 7/22/80 memo to Waksberg.

Show three values of K_c

K_c as defined above where the summation on a is over all size classes in the county,

K_{c1} computed by omitting all contributions from size classes 11, and

K_{c2} computed by omitting all contributions from size classes 10 and 11.

2. Sequence and list county records

2.1 Sequence the county summary records by the clerically assigned 4-digit sequence number; first make any sequence number corrections required for the reasons given in paragraph 2.2.

2.2 Make necessary changes in the sequence number.

This instruction will not apply the first time the records are sequenced. However, later review of the PSU's that have been prepared may require changes in the sequence numbers for a few counties so that a different set of PSU's are generated and reviewed. The changes in sequence numbers may arise because of the need to generate a different PSU by (for example):

- o Adding a non-metro county to an SMSA
- o Making more than one PSU out of an SMSA
- o Combining counties not adjacent in the sequence

2.3 List the county records.

With the county records in the sequence specified in 2.1, list the county records.

3. Combine metro counties into SMSA summary records.

3.1 Summarize into a single record the information from the counties within SMSA's; sequence numbers for this group will be in the range 0001 - 0999 and all counties in the SMSA will carry the same sequence number. This will produce one record for each SMSA (or pseudo - SMSA in the event we decide to revise the SMSA definition for sampling purposes). Each of the items listed under section 1 should be summarized and shown for the SMSA summary record.

3.2 The combined record will be identified by the sequence number common to the component parts (e.g. component counties). The record must also carry with it a list of the counties (State and county codes) covered by the combined record.

3.3 Examine the cost-measure condition

3.3.1 Compute Rc for each combined SMSA record

Compute for the c th record the value of

$$Rc = (279.4)\ \tilde{fac}$$

The term $\tilde{fac}$ is the value of fa for the highest numbered size class on the c th record having a non zero count of establishments. You are to ignore all counts that may appear in size classes 10 or 11 in determining Rc.

3.3.2 Compare Rc and Kc_2

The term Kc_2 is defined in section 1.2.2.

- If $Kc_2 \leq$ Rc enter a flag in the records so that it will be identified as an unsatisfactory PSU (metro PSU is too small) when the combined records are listed; enter Rc on the record.
- If $Rc < Kc_2$, enter Rc on the record but do not show the unsatisfactory PSU flag.

4. Combine non-metro records to generate PSU's.

In this operation we intend to start by examining the record for the first county in the sequence. If the record fails to pass a certain test, we want to combine records with those following in sequence, testing the combinations until one is made to pass the test. This will identify PSU's (a county or a group of continguous counties) that satisfy the test condition. We wish to have the PSU's listed for review; as a result of the review we may change some of the sequence numbers and ask to have the program repeated. The operation begins with the records sequenced as in section 2.1

The instructions are described in terms of a loop.

4.1 Compute Rc for the record

The "record" refers to the area being tested (a county or a combination of counties). The term Rc is defined in section 3.3.1. The next operation is specified in section 4.2.

4.2 Compare Rc and Kc_2

The term Kc_2 is defined in section 1.2.2.

4.2.1 If $Rc \geq Kc_2$ create a new record by combining the current record with the record following it in sequence and then go to section 4.1 to execute the test on the combined record. The entries on the combined record are to be:

- Sequence number same as for current record.
- Data fields, including all M's and K's are equal to the respective sums of data fields on the records being combined.
- State and County identifications must be shown for every county covered by the combined record.

4.2.2 If $Rc < Kc_2$ the record is assumed to define an acceptable PSU; this record is to be listed; enter Rc on the record. Bring in a new record and go to section 4.1 to begin testing records to define the next PSU.

RH:rb

cc: J. Waksberg
T. McKenna
J. Edmonds
C. McElvea
B. Tevelow
D. Ward

Exhibit 5

WESTAT
An Employee-Owned Research Corporation

MEMORANDUM

TO: Diane Ward December 23, 1980

FROM: R. Hanson

SUBJECT: Characteristics of sample PSU's for NOHS II

The attached table lists all of the sample PSU's for the NOHS II. The table shows for each PSU the identification codes, the pair number, approximations to the expected team workload by year of investigation, the PSU weight, and the PSU description. I have tried to spell out the meaning of the information shown for the PSU's in order to make this document self-contained.

The expected number of team weeks for the PSU's shown on this table include the investigation time required to cover the sample establishments in the two largest size classes that are located in or near the sample PSU's. The table does not reflect the additional time needed for investigating sample large firms that have not been associated with a specific sample PSU; see table 3 of D. Ward's letter to Sundin dated 11/19/80.

PSU Number

A three digit code is assigned to each PSU using the following system:

For self-representing PSU's:

Hundreds digit (Census region)

1 Northeast
3 North Central
5 South
7 West

Tens digit

One digit assigned to SR PSU's within the Census region

Units digit

0 PSU to be interviewed during both of the first and second years

1 Interviews in PSU to be completed in year 1

2 Interviews in PSU to be completed in year 2

For non self representing PSU's

Hundreds digit (Census region)

2 Northeast
4 North Central
6 South
8 West

Tens and units digits

Two digit serial number assigned to NSR PSU's within the region.

Pair Number

The NSR strata have been paired and a random (equal probability) selection of one of the sample PSU's in each pair has been designated

for interview in the first year, the other sample PSU is investigated during the second year. Pair numbers ending in A and B identify the PSU's to be investigated in year one and year two respectively. The smaller SR PSU's have been paired and designated for investigation in years one and two and assigned pair numbers in a similar manner.

The larger SR PSU's are to be interviewd in both of the two years; the investigation time for the portions of the within-PSU sample to be covered during each of the two years is displayed in the table. The method of determining the PSU workloads is discussed under the description of expected team weeks, below.

Expected team weeks

An approximation to the number of person-hours required to investigate the expected sample NOHS II firms in each PSU is given by the entry for expected team weeks. The number of person-hours is reported in units of 139.7 hours; this is the number of effective investigator hours supplied by a four person team in one week. The total person-hours for a PSU is based on the sum of the products of the expected number of sample firms by size class and the person-hours per case observed during NOHS I. The number of sample firms in this computation is consistent with a sample that would sum over the PSU's to approximate the optimum number of 4,895 CBP establishments; the number of firms does not include the additional 25 percent increment to the optimum sample that makes up the expanded sample

The workload for a PSU comes from two groups of sample establishments:

1) Sample firms reported to have 2500 or more employees (size classes 9 and 10).

Firms of this size are selected for the sample without considering their location. As a result they may or may not lie in a sample PSU. A sample for screening has been selected from the class 9 and 10 D&B listings, and presumably each sample case represents an actual establishment (in contrast to an expected number that would result by sampling the count of firms on the CBP county summary records as done for size classes 1 through 8).

Most of these sample large firms have been assigned to specific sample PSU's, either because they are actually located within the PSU or nearby, so that they could be conveniently reached from the PSU. The person-hours required to investigate all of these sample large firms are included in the team weeks shown for the PSU in the table. A small number of additional sample large firms are not located convenient to any of the sample PSU's and their investigation times are not included in the table.

2) Sample firms reported to have less than 2500 employees (size classes 1 through 8)

All of these sample firms lie within the sample PSU's. The total number of establishments expected in sample from all sources of universe listings for a PSU is estimated by applying the within PSU sampling intervals by size class to the number of CBP firms in the PSU.

The sample delivered to NIOSH is the full expanded sample which is defined as 125 percent of the sample computed as the optimum (see exhibit 2). The numbers of person-hours contributed to the team week entry for each PSU for firms in size classes 1 through 8 have been multiplied by the factor 0.8 to bring the expected workload to the level consistent with a sample size of the optimum.

The number of team weeks shown in the table for the PSU depends on the proposed method of investigating firms during the first and second years of field investigation. The expected sample in the PSU is considered in terms of the four subsamples of the expanded sample referred to in my memorandum of 8/20/80 to J Blais (see exhibit 6) but reduced to remove the 25 percent increment that is included in the expanded sample.

The workload subsamples assigned for investigation in the NDHS II PSU's are summarized as follows:

- For large SR PSU's to be visited for investigation in both years, the subsamples assigned in years 1 and 2 are:

Assignment for year	Firms in size class: 1 thru 8	9 and 10
1	B,C,D	B,C
2	A	A

- For all other sample PSU's (small SR and all NSR PSU's) to be investigated in one year, the subsamples are:

	A,B,C,D	A,B

The total over all PSU's of the number of person-hours shown in the table will exceed the 76,800 hours used in computing the optimum (the total in the table is about 587 team weeks or 587 x 139.7 ÷ 82,000 hours). The excess comes from using the counts obtained by sampling the D&B listings which yield 110 and 127 selections from class 9 and 10 vs 94 and 97 selections expected by sampling CBP for these size classes.

PSU probability

The probability of selecting the sample PSU from its stratum is the ratio of the PSU to the stratum measures of size. The probabilities are shown as "1 in.." (1/2.854 for PSU 201, for example). These probabilities are also given in J. Waksberg's letter to Sundin of 10/20/80 as amended by J. Edmond's letter of 11/7/80.

Composition of the PSU

The State and counties making up each sample PSU are shown. The composition is the same as given in J. Waksberg's letter of 10/10/80 as amended by J. Edmonds letter of 11/7/80.

cc: J. Waksberg
J. Edmonds
C. McElrea
T. McKenna

NOHS II SAMPLE PSU'S

PSU number	Pair number	Expected team-weeks*	PSU probability 1 in:	Composition of PSU	
				State	Counties
NON-SR PSU's					
201	13A	5.78	2.854	NY	Albany, Greene, Montgomery, Rennselaer, Saratoga, Schenectady
202	13B	5.44	1.885	RI	Bristol, Kent, Newport, Providence, Washington
203	17A	5.51	1.201	NY	Erie, Niagara
204	17B	7.29	6.531	CT	New London, Windham
205	19A	2.98	8.046	ME	Hancock, Nennebec, Knox, Lincoln, Waldo, Washington
206	19B	4.74	11.984	PA	Blair
207	21A	2.54	7.375	NY	Cattaraugus, Chautauqua
208	21B	5.26	3.164	PA	Lancaster
209	22A	7.13	1.973	CT	Fairfield
210	22B	3.32	2.017	PA	Lackawanna, Luzerne, Monroe, Wyoming
211	28A	2.49	2.882	NJ	Passaic, Sussex
212	29B	3.85	5.954	NJ	Mercer
213	31B	3.97	5.189	PA	Columbia, Montour, Schuylkill, Sullivan
214	36A	3.66	2.227	NJ	Middlesex
401	3A	7.37	8.879	MI	Genesse, Shiawassee
402	8A	9.61	2.073	IN	Boone, Hamilton, Hancock, Hendricks, Johnson, Marion, Morgan, Shelby

NOHS II SAMPLE PSU'S

PSU number	Pair number	Expected team-weeks*	PSU probability 1 in:	Composition of PSU	
				State	Counties
403	8B	2.53	2.872	IA NE	Pottawattamie Douglas, Sarpy
404	9A	2.51	13.305	MN	Benton, Sherburne, Stearns
405	12B	3.25	7.077	WI	Brown
406	15A	2.30	16.050	KS	Douglas, Franklin, Leavenworth, Miami
407	23A	2.34	8.835	OH	Guernsey, Harrison, Tuscarawas
408	23B	5.05	1.787	OH	Delaware, Fairfield, Franklin, Madison, Pickaway
409	24A	4.61	2.739	MI OH	Monroe Fulton, Lucas, Ottawa, Wood
410	24B	5.84	3.762	IN	Adams, Allen, DeKalb, Wells, Whitley
411	25A	2.49	13.362	MO	Audrain, Boone, Callaway Howard, Randolph
412	29A	1.74	16.814	KS MO	Allen, Anderson, Bourbon, Coffey, Linn, Woodson Barton, Bates, Henry, St. Clair, Vernon
413	30A	3.89	8.535	WI	Racine
414	31A	6.14	16.327	OH	Knox, Marion, Morrow
415	33A	5.18	11.979	MI	Hillsdale, Lenawee
416	33B	2.62	7.908	IN OH	Lagrange, Steuben Defiance, Henry, Paulding, Williams
417	35A	5.22	16.355	IN	Dubois, Knox, Pike, Spencer
418	36B	8.84	2.768	OH	Portage, Summit

NOHS II SAMPLE PSU'S

PSU number	Pair number	Expected team-weeks*	PSU probability 1 in:	Composition of PSU	
				State	Counties
601	1A	1.78	11.849	TX	Bee, Brooks, Dimmit, Duval Frio, Goliad, Jim Hogg, Jim Wells, Karnes, Kenedy, Kinney, Kleberg, LaSalle, Live Oak, Maverick, McMullen, Starr, Uvalde, Willacy, Zapata, Zavala
602	1B	1.99	2.507	FL	Broward
603	4B	7.24	1.418	LA	Jefferson, Orleans, Plaquemines, St. Bernard, St. Charles, St. Tammany
604	5A	2.26	16.870	TX	Atascosa, Bandera, Blanco, Bosque, Burnet, Caldwell, Comanche, Erath, Gonzales, Hamilton, Kerr, Medina, Mills, San Saba, Somervell, Wilson
605	5B	2.11	13.643	TX	Austin, Bastrop, Colorado, Fayette, Jackson, Lavaca, Lee, Matagorda, Wharton
606	6A	2.55	4.856	MS	Hinds, Madison, Rankin
607	6B	1.67	9.920	TX	Clay, Montague, Wichita
608	9B	2.75	1.196	FL	Hillsborough, Pasco, Pinellas
609	10A	3.80	1.993	AR MS TN	Crittenden DeSoto Shelby, Tipton
610	10B	7.96	2.052	OK	Creek, Mayes, Osage, Rogers, Tulsa, Wagoner
611	11A	4.58	7.073	AL	Autauga, Elmore, Montgomery
612	14A	4.70	4.703	SC	Lexington, Richland
613	14B	2.57	3.564	AK	Pulaski, Saline
614	15B	4.59	3.621	DE MD NJ	New Castle Cecil Salem

NOHS II SAMPLE PSU'S

PSU number	Pair number	Expected team-weeks*	PSU probability 1 in:	Composition of PSU	
				State	Counties
615	16A	4.99	17.158	VA	Dinwiddie, Prince George, Cities of Colonial Heights, Hopewell, Petersburg
616	16B	7.24	15.921	AL	Choctaw, Clarke, Conecuh, Monroe, Washington
617	18A	3.95	20.721	SC	Clarendon, Georgetown, Williamsburg
618	18B	4.03	12.059	NC	Johnson, Wilson
619	20A	3.82	18.318	KY	Bath, Elliot, Fleming, Johnson, Laurence, Lewis, Magoffin, Martin, Mason, Menifee, Montgomery, Morgan, Nicholas, Robertson, Rowan, Wolfe
620	20B	5.73	2.292	SC	Greenville, Pickens, Spartanburg
621	25B	3.01	14.522	MD	Somerset, Wicomico, Worcester
622	26A	5.33	1.920	NC	Davidson, Davie, Forsyth, Guilford, Randolph, Stokes, Yadkin
623	26B	2.77	3.461	GA TN	Catoosa, Dade, Walker Hamilton, Marion, Sequatchie
624	27A	4.39	9.234	AL	Calhoun, Etowah
625	28B	4.52	21.775	VA	Bedford, Franklin, Rockbridge, Cities of Bedford, Buena Vista, Lexington
626	30B	3.73	10.201	OH WV	Washington Wirt, Wood
627	32A	5.96	12.052	NC	Caswell, Granville, Person, Rockingham
628	32B	5.50	21.284	MS	Clay, Lowndes, Webster

NOHS II SAMPLE PSU'S

PSU number	Pair number	Expected team-weeks*	PSU probability 1 in:	Composition of PSU	
				State	Counties
629	34A	2.90	14.542	GA	Dawson, Fannin, Gilmer, Habersham, Lumpkin, Murray, Pickens, Rabun, Towns, Union
630	34B	4.59	16.618	TN	DeKalb, Putnam, White
631	35B	4.46	18.029	KY	Anderson, Bracken. Carroll, Franklin, Gallatin, Grant, Harrison, Henry, Owen, Pendleton, Shelby, Spencer, Trimble
801	2A	2.29	2.969	CA	Placer, Sacramento, Yolo
802	2B	1.83	7.163	CA	Kern
803	3B	1.41	28.990	AK	Divisions of: Upper Yukon, Fairbanks, South East Fairbanks
804	4A	2.68	5.363	NV	Clark
805	7A	3.89	2.177	CA	Riverside, San Bernadino
806	7B	2.39	4.933	CA	Fresno
807	11B	5.17	1.871	OR	Clackamas, Multnomah, Washington, Yamhill
				WA	Clark
808	12A	2.74	6.501	CO	El Paso, Pueblo, Teller
809	27B	8.07	1.170	CA	Santa Clara
SR-PSU'S TO BE COMPLETED IN A SINGLE YEAR					
142	37B	5.49	1.0	NY	Nassau, Suffolk
371	38A	8.22	1.0	WI	Milwaukee, Ozaukee, Washington, Waukesha
381	39A	11.51	1.0	IN	Dearborn
				KY	Boone, Campbell, Kenton
				OH	Brown, Clermont, Hamilton, Warren

NOHS II SAMPLE PSU'S

PSU number	Pair number	Expected team-weeks*	PSU probability 1 in:	Composition of PSU	
				State	Counties
392	39B	6.24	1.0	KS	Johnson, Wyandotte
				MO	Cass, Clay, Jackson, Platte, Ray
511	37A	3.79	1.0	MD	Calvert, Charles, Frederick, Montgomery, Prince Georges
				VA	Arlington, Fairfax, Loudoun, Prince William, Cities of: Alexandria, Fairfax, Falls Church, Manassas, Manassas Park
				DC	Washington
542	38B	5.77	1.0	MD	Anne Arundel, Baltimore, Carroll, Harford, Howard, City of Baltimore
552	40B	4.73	1.0	GA	Butts, Cherokee, Clayton, Cobb, DeKalb, Douglas, Fayette, Forsyth, Fulton, Gwinnett, Henry, Newton, Paulding, Rockdale, Spaulding, Walton
561	40A	3.60	1.0	FL	Dade, Monroe
731	41A	7.76	1.0	CA	Orange
742	41B	3.19	1.0	CA	San Diego
752	42B	5.64	1.0	CO	Adams, Arapahoe, Boulder, Denver, Douglas, Gilpin, Jefferson
761	42A	4.09	1.0	WA	King, Snohomish
SR-PSU'S TO BE COMPLETED OVER TWO YEARS					
110	43A	10.85	1.0	NJ	Bergen
	43B	13.94		NY	Bronx, Kings, New York, Putnam, Queens, Richmond, Rockland, Westchester
120	44A	4.66	1.0	NJ	Burlington, Camden, Gloucester
	44B	5.27		PA	Bucks, Chester, Delaware, Montgomery, Philadelphia

NOHS II SAMPLE PSU'S

PSU number	Pair number	Expected team-weeks*	PSU probability 1 in:	Composition of PSU	
				State	Counties
130	45A 45B	6.67 9.38	1.0	MA	Barnstable, Dukes, Essex, Middlesex, Nantucket, Norfolk, Plymouth, Suffolk
				NH	Rockingham
150	46A 46B	4.92 2.95	1.0	NJ	Essex, Hunterdon, Morris, Somerset, Union
160	47A 47B	5.40 9.47	1.0	PA	Allegheny, Beaver, Washington, Westmoreland
310	48A 48B	14.77 13.66	1.0	IL	Cook, Dupage, Kane, Lake, McHenry, Will
320	49A 49B	10.51 15.81	1.0	MI	Lapeer, Livingston, Macomb, Oakland, St. Clair, Wayne
330	50A 50B	5.52 8.23	1.0	IL	Clinton, Madison, Monroe St. Clair
				MO	Franklin, Jefferson, St. Charles, St. Louis, city of St. Louis
340	51A 51B	3.34 2.11	1.0	MN	Anoka, Carver, Chisago, Dakota, Hennepin, Isanti, Ramsey, Scott, Washington, Wright
				WI	St. Croix
350	52A 52B	6.94 5.46	1.0	OH	Cuyahoga, Geauga, Lake, Medina
520	53A 53B	6.52 5.04	1.0	TX	Collin, Dallas, Denton, Ellis, Hood, Johnson, Kaufman, Parker, Rockwall, Tarrant, Wise
530	54A 54B	3.41 4.02	1.0	TX	Brazoria, Chambers, Fort Bend, Harris, Libert, Montgomery, Waller
710	55A 55B	12.95 12.95	1.0	CA	Los Angeles
720	56A 56B	2.46 4.44	1.0	CA	Alameda, Contra Costa, Marin, San Francisco, San Mateo

*Expected time to complete the sample of firms with less than 2500 employees located in the sample PSU plus time to complete sample of larger firms located in or near the sample PSU.

Exhibit 6

WESTAT
An Employee-Owned Research Corporation

MEMORANDUM

TO: J. Blais August 21, 1983

FROM: R. Hanson

SUBJECT: Specifications to Select Sample Establishments from the D&B File for NIOSH

These specifications revise my earlier memos on this subject. The present revision modifies specifications pertaining to listings to be provided to the screening operation for sample cases in each PSU and also corrects and updates Table 1. We have also rewritten the memo to incorporate the size class definitions used in other specifications memos.

The sampling instructions given here should, when applied to all sample PSU's, produce a sample about 25 percent greater than we expect the surveyors can cover assuming their average times for investigation are about the same as in NOHS I.

1. Sampling Procedures within Self Representating PSU's

1.1 Establish the file for D&B establishments for each of the SR PSU's. Sequence the file by 4-digit target SIC within the PSU.

1.2 With the file in the specified sequence, count the establishments and select sample establishments independently within each of the employee size classes given in Table 1. Follow the detailed sampling instructions given in paragraphs 1.3 and 1.4.

1.3 The sampling instructions to be followed independently for size classes 1 through 8 and for class 11 are summarized in Table 1 and are to proceed by the following steps:

- For the employee size class identified in column (2) of Table 1, enter the random start from column (3) into a counter. This step applies only for the first of the 98 PSU's to be sampled; random starts for subsequently sampled PSU's are computed, see Section 3.2.

- The integer portion of the number in the counter specifies the establishment in sample; count the establishments in the size class until this establishment is reached, identify it as the sample case.

Table 1. Instructions for sampling establishments from D&B for counties in each of the sample PSU's

Class (1)	Employee size class (2)	Random starts (3)	Sampling intervals (b) Chicago sample (4) (a)	Sampling intervals (b) Balance of the U.S. (5)	Over-sampling ratio f_a (6)
11	NA	19.60	84.844	79.81	1.0
1	8-19	23.22	84.844	79.81	1.0
2	20-49	17.20	53.264	50.10	1.593
3	50-99	14.08	28.076	26.41	3.022
4	100-249	1.550	15.528	14.61	5.464
5	250-499	7.260	9.542	8.505 (c)	9.384
6	500-999	5.784	6.250	5.879	13.576
7	1000-1499	1.852	4.922	4.631	17.235
8	1500-2499	1.937	3.568	3.355	23.785
9	2500-4999	2.007	-	2.218	35.985
10	5000+	.035	-	1.168 (d)	91.110

a Sample for Chicago SMSA selected before final sampling rates were determined.

b Includes the shadow sample.

c For example, for size class 5:

$$\frac{1}{8.505} = \frac{2}{2(8.505)} = \frac{1}{17.010} + \frac{1}{17.010}$$

= (screening and interview) + (shadow).

d For size class 10 only:

$$\frac{1}{1.168} = \frac{1.5}{1.5(1.168)} = \frac{1}{1.5.(1.168)} + \frac{.5}{1.5(1.168)}$$

$$= \frac{1}{1.752} + \frac{1}{3.504}$$

= (screening and interview) + (shadow).

- Add the sampling interval from column (5) to the counter. The integer portion of the sum in the counter designates the next sample establishment (use the sampling interval given in column (4) to designate the sample in the Chicago PSU).

- Continue, by repeating the instructions of the preceding step, until there are no more establishments of the size class in the PSU.

1.4 The sample for size classes 9 and 10 will be selected wherever located, even if not in a sample PSU:

- Establish a D&B file at the U.S. level covering all target industries with 2,500 or more employees. Sequence this file by zip within 4-digit SIC.

- Designate samples in each of size classes 9 and 10 by applying the instructions given in Table 1 without regard to the location of the establishment. The sampling intervals given in column (5) of Table 1 are to be used.

2. Maintain Sampling Records for Each PSU

The following sampling records must be kept by the computer for each employee size class for each PSU. The sampling procedures are to use these records to compute a new set of Random Starts and Sequence numbers for each PSU as the sampling process goes from one PSU to the next.

When sampling has been completed for a PSU, retain the following information for the PSU for each size class:

- The random start used in the a^{th} size class that began the sampling operation in the PSU = RS(1a);

- The sampling interval used for the a^{th} size class in the PSU = I(1a);

- The number of sample selections made in the a^{th} size class in the PSU = m(1a);

- The value of the counter when the last selection was made in the a^{th} class = D(1a). As a check,

  D(1a) = RS(1a) + (m(1a) - 1) * I(1a); and

- The total number of all establishments in the a^{th} size class in the PSU = M(1a).

3. Sampling within NSR PSU's

PSU's will be sampled in an order determined for the convenience of field survey operations.

Use the probability of selecting the PSU from its stratum, the basic sampling interval, and the sampling records for the preceding PSU to determine the parameters needed for sampling in the next PSU in the selection order.

3.1 The overall probability of selecting establishments from size class 1 determines the basic sampling interval. This interval is 1/79.81 for all PSU's (except 1/84.844 is used for the Chicago PSU). The intervals to use for other size classes in the PSU are obtained by dividing the basic interval by f given for the size class in column (6) of Table 1. These intervals have been computed for SR PSU's and appear in column (5) of the table.

The basic interval for a nonself-representing PSU is computed by the formula:

> PSU basic interval = 79.81 * (probability of selecting the PSU from its stratum)

This formula shows that the PSU basic interval is 79.81 for each SR PSU because the probability of selecting the PSU from its stratum is 1.0 for these PSU's.

For NSR PSU, for example for PSU 201 made up of Albany, Greene, Montgomery, Rennselaer, Saratoga, and Schenectady counties in New York (which has a PSU selection probability of 1./2.854), the basic sampling interval is:

> 79.81 * (1/2.854) = 27.964.

Using the values of f_a in Table 1, the sampling intervals to use in PSU 201 are:

Class	Intervals for PSU 201		
1	27.964		
2	17.554	=	27.964 / 1.593
3	9.253	=	27.964 / 3.022
4	5.118	=	27.964 / 5.464
5	2.980	=	27.964 / 9.384
6	2.060	=	27.964 / 13.576
7	1.623	=	27.964 / 17.235
8	1.176	=	27.964 / 23.785
11	27.964	=	27.964 / 1.0

3.2 The random starts to use in the a^{th} size class for the PSU that follows in the selection order is given by the following formula. Let RS(2a) be the random start needed for the a^{th} class for the PSU that follows, then:

$$RS(2a) = (D(1a) + I(1a) - M(1a)) * I(2a)/I(1a).$$

where the terms with subscript 1 are defined in paragraph 2 for the PSU immediately preceding in the selection order.

For example, assume PSU 201 is to be sampled immediately following PSU 381 (Cincinnati) for which the sampling records are assumed to show (for illustration only):

RS(1a)	=	3.193
I(1a)	=	79.81
m(1a)	=	17
D(1a)	=	1280.153
M(1a)	=	1345.

Then the random start for PSU 201 would be:

$$(1280.153 + 79.81 - 1345) * 27.964 / 79.81 = 5.243.$$

4. Assign identification codes to the sample establishments.

4.1 Assign a 4-digit serial number to each sample case selected in each size class.

The number series begins at 0001 for the first sample establishment selected in each size class for Chicago (the first PSU in the selection order) and continues by numbering the sample cases in one unbroken sequence in each size class independently across all sample PSU's (or across the entire U.S. file for cases selected in Section 1.4).

The sample being selected by this process includes the shadow sample. For all classes except class 10, the sample case having the odd number will be initially assigned for telephoning and investigation; if a shadow is needed, the even numbered following case will be used. Thus each case designated for the sample will be followed by its shadow in classes 1 thru 9 and in class 11. For size class 10, every third selection will be a shadow, thus in each group of three numbers, the first two will be designated for telephoning and investigation and the third case will be a single shadow for the two sample cases.

4.2 For each of employee size classes 1 thru 9 and for class 11, assign a 2-digit serial number modulo 16 to each sample case in each size class.

The numbering is to be carried over from PSU to PSU independently in each of these size classes without interrupting the sequence.

Use these numbers to assign all sample cases in each PSU to four random assignment groups for the PSU as follows:

	Two-digit serial numbers
Group A	01, 05, 09, 13
B	03, 11
C	07
D	15

4.3 For sample cases selected from class 10, assign a 2-digit serial number modulo 12, carry the numbering system through all sample selections without interruption.

Use these serial numbers to assign all sample cases in the U.S. to four random assignment groups as follows:

	Two-digit serial numbers
Group A	01, 04, 07, 10
B	02, 08
C	05
D	11

5. Listings to be provided for screening and telephone.

5.1 For each PSU provide the following listings of the D&B universe:

5.1.1 Alphabetic by company name within the PSU

5.1.2 Alphabetic by company name within zip within the PSU

5.2 For each sample PSU (and for the class 9 and 10 samples produced in step 1.4), provides a listing of the D&B sample cases:

5.2.1 For the screening sample (i.e., excluding shadows) by the four random assignment groups; by serial number within size class;

5.2.2 For the screening sample; alphabetical by company name showing the identification codes; and

5.2.3 For all sample selections (including the shadows); list the selections in sequence by serial number within size class.

5.3 Provide listings of the U.S. universe for size classes 9 and 10:

5.3.1 Alphabetic by company name; and

5.3.2 Alphabetic by company name within zip.

cc: J. Waksberg
T. Mckenna
J. Edmonds
D. Ward
B. Tevelow
C. McElrea

Exhibit 7
Memorandum I
July 27 1983

MEMORANDUM TO: J. Waksberg

FROM : R. Hanson

SUBJECT : NOES Estimation Memorandum I--
Inflation Estimation

This memorandum documents the first of the several factors in the final tabulation weight for the NOES sample. Memorandum II provides specifications for ratio estimation; specifications for preparatory steps for variance computations are presented in memorandum III of this series. A common set of tables (tables 1 through 6) is referred to in these three memoranda.

As all of the operations described in this memorandum are performed by Westat, there are no restrictions on our choice of program language.

I INTRODUCTION

A. Summary of the NOES sample

The procedures leading to the visit of surveyor teams to the sample of establishments for the completion of the questionnaires may be summarized as follows:

1. A sample of establishments (the "telephone sample") is selected from establishments reported by D&B as members of the NOES target SIC's and having 8 or more employees; a sample of firms with number of employees reported as NA is also included. In general, the sample is self weighting within establishment size strata. As the probabilities are known, the D&B information for the sample establishments can be the basis of unbiased estimates of the D&B universe.

2. The telephone sample establishments are contacted by telephone to identify those still in business and in scope; the in scope cases are assigned to a survey team for field visit and interview. At the time of the initial telephone contact, establishments found to be temporarily out of business and those indicating refusal are treated as in scope cases. All other out of scope cases are retired from further consideration in the survey.

3. The sample establishments treated as in scope as a result of the telephone interview are assigned for field interview. The teams are to complete a field interview for each establishment found to be in scope at the time of the field visit. A shadow sample is substituted for a refusal at the time of field interview; if the shadow case also refuses or is found to be out of scope, NOES personnel may obtain a court order to require cooperation of the original sample establishment.

B. Summary of the NOES weighting procedure

The estimation procedure for NOES develops weights to be assigned to the questionnaires completed for sample establishments in the field. Publication estimates of totals from the NOES sample will be prepared by summing the weights assigned to completed questionnaires having the characteristic of interest.

There are a number of counts and estimates involved in developing the tabulation weights; the definitions of these components are summarized in table 1 and are discussed as needed in the sections that follow.

The final weight assigned to an interviewed establishment for preparing publication estimates is the product of three types of weights:

- An inflation weight, based on the inverse of the probability of selecting the establishment in the telephone interview sample; two versions are used-- one applies to the telephone sample as originally selected, the other to the sample of establishments interviewed in the field.

- A first stage ratio estimate factor, based on universe counts of D&B establishments by size class and estimates of these universe counts from the telephone sample, and

- A second stage ratio estimate factor based on universe counts of employees (or counts of establishments in some instances) from the 1981 County Business Patterns for selected groups of target SIC's by size class and estimates of these counts from the interviewed sample using weights equal to the product of the inflation weight and the first stage ratio estimate factor.

This memorandum describes the inflation weights used to prepare the unbiased estimates involved in the weighting system.

II DEFINITION OF INFLATION WEIGHTS

A. Two inflation weights

There are two inflation weights involved in the NOES estimation procedure; both of these weights involve the inverse of the probability of selecting an establishment as a member of the telephone sample. In this memorandum the weights are identified as follows (refer to table 1):

1. The telephone inflation weight - used in preparing unbiased estimates of the D&B universe of establishments from which the telephone sample was originally selected; this weight is represented by the symbol WT in table 1.

2. The field interview inflation weight - used in preparing unbiased extimates of the current study universe from establishments at which field interviews were assigned; this weight is represented by the symbol W in table 1.

B. Survey operations reflected in the two inflation weights

Several survey operations have an impact on the two inflation weights. These operations involve:

- Assignment of different workload subsamples to the PSU's,
- Sampling establishments with certainty in some PSU's,
- Results of the telephone interview,
- Substitutions of shadow sample cases for refusals, and
- Duplicate listings in the D&B universe.

C. Definition of the telephone sample inflation weight WT

The telephone sample weight WT is the inverse of the probability used to select the D&B establishment in the telephone interview sample as modified by the first two of the operations listed in paragraph B.

D. Definition of field interview inflation weight W

The field interview inflation weight W is the inverse of the probability of selecting the D&B establishment in the sample assigned for telephone interview as modified by all five of the operations listed in paragraph B.

E. SIC and size class definitions

The sample design requires weights to be defined in terms of establishment size classes and SIC definitions. All references to size class and SIC in this memorandum refer to the information given for the establishment on the original D&B universe file. The information on records for the sample must therefore include the original information as reported on the 1980 D&B market file for the establishment. (Additional steps in weighting-- discussed in Memorandum II-- will involve current establishment size and SIC determined during interview of the establishment in the field.)

III TELEPHONE INFLATION WEIGHTS

The sample selected for telephone interview was intended to be self weighting and independent within each of the 11 establishment size classes defined in table 2. This means that, theoretically, every telephone sample establishment in a given size class should have the same telephone inflation weight irrespective of the PSU from which it has been selected. However, two situations introduce changes in these theoretical weights. A given PSU can be affected by either or both of these situations. Note that the size class, the identification of the PSU, and the exceptions discussed below in sections A and B will be sufficient to determine the telphone sample weight for an establishment.

A. Assignment of workload subsamples to PSU's

To permit flexibility in workload assignments, the telephone sample was divided into four subsamples such that A = 1/2, B = 1/4, and C = D = 1/8 of an expanded sample. The expanded sample was set to be 25 percent larger than an optimum sample; the optimum number of interviewed establishments was estimated at 4900 and the full expanded sample was calculated at about 6100 interviewed establishments.

The sample to be interviewed in each PSU was assigned in terms of these four workload subsamples. All PSU's were assigned subsamples ABC. However, in many PSU's it became possible to assign the full ABCD sample for field interview; these PSU's are identified by an asterisk (*) in tables 3, 4 and 5.

Table 2 shows weights that apply for sample establishments by size class for the two classes of PSU; those in which workload subsamples ABC and ABCD have been assigned.

The telephone sample consistent with the full expanded sample (ABCD) was assigned for telephone interview in size classes 9 and 10. The telephone sample weights appearing in table 2 reflect the use of the full ABCD sample for these size classes. Note that the sample from these two classes has been selected without regard to sample PSU.

A separate set of weights is shown for PSU 310 (Chicago) to reflect the revised sampling probabilities used in that one PSU.

B. Sampling with certainty within certain PSU's

The probability of selecting an establishment within a sample PSU is illustrated by considering samples selected from size class number 5. The overall probability of selecting an ABCD telephone sample for this size class is given in table 2 as 1 in 17.010. However, in all PSU's a shadow sample equal to the full telephone sample was also selected at the same time so that the selection probabilities must satisfy the following relationship:

2 * 1/17.010 = (selection of sample PSU) * (selection of establishments within the sample PSU).

This implies that if a sample PSU is selected with probability of say 1/3, then (within this PSU) establishments in this size class must be selected at

3 * 2 /17.010

(that is at 1/2.835) to conform to the above relationship.. This method of sampling can cause problems if the PSU selection probability is very small. For example, we know from other sources that PSU 206 was selected from its stratum at 1 in 11.984. Substitution of this probability in the foregoing relationship means the following sampling procedure must operate within PSU 206 for size class 5:

2/17.010 = (1/11.984) * (2 * 11.984/17.010).

As the latter probability (the fraction within the second parenthesis on the right) is greater than 1.0, all establishments in size class 5 in PSU 206 must be in sample with certainty; half of these units comprise the ABCD telephone sample and the other half make up the shadow sample.

The theoretical weights shown in table 2 imply the result for size class 5 also applies for size classes 6, 7, and 8 in this PSU. The telephone sample of establishments in PSU 206 selected from size class 5 (and 6, 7, and 8) must therefore comprise half of all such establishments in the PSU and each of the telephone sample cases from size classes 5, 6, 7, and 8 would have telephone sample weights of 2 * 11.984 = 23.968 (not 17.010 or 11.757 or 9.261 or 6.711 as given in table 2). Information for these weight adjustments is given in table 5 for all PSU's affected by this problem.

Note that a blank cell in table 5 means the inflation weight for that cell is to be derived as described in paragraph A, above.

IV FIELD INTERVIEW INFLATION WEIGHT

The definition of the field interview inflation weight for an establishment as given in section II D above means the weight can be specified by first determining the telephone sample inflation weight and then incorporating any changes that follow from among the last three survey operations given in section II B. Changes in the inflation weight from information determined during the telephone interview must be examined on a case by case basis. The changes resulting from these situations are:

A. Results of the telephone interview

1. Establishments found to be permanently out of business or out of scope (reporting fewer than 8 employees or with primary activity in a non target SIC) are removed from the interview sample by assigning field inflation weights of zero.

2. Establishments reported as temporarily out of business, refusals, and all in scope establishments are to be assigned field interview inflation weights equal to their telephone sample weights. Note that later steps may require changes in these weights; at the time of field interview, establishments found to be out of scope are assigned field inflation weights of zero and not included in the survey.

3. The telephone screening interview questioned each firm about other plant locations in the same PSU. If additional plant locations were reported that did not appear on the DMI file, they were given a chance of selection. The selection probability was determined for each of these plants by assuming they appeared on the DMI in the size class given by the number of employees reported by the telephone interview. The cases selected by these assumed probabilities were assigned field inflation weights equal to the inverse of these computed selection probabilities. Telephone sample weights for these additional selections remain as zero.

B. Substitution of shadow samples for refusals

An establishment assigned for a field interview may refuse to cooperate. A new shadow sample establishment is then introduced and, after its telephone interview, may be assigned for field interview. If the shadow sample substitute is in scope and cooperates, its telephone sample weight is assigned as its field interview weight and the weight of the original sample unit (the refusal) is set to zero. Note that the weight for the shadow sample can require the modifications discussed in the sections preceding; note also that a shadow-- ordinarily the unit selected immediately after the refusal case-- does not necessarily come from the same PSU as the refusal.

If the shadow sample case is out of scope or does not cooperate, the originally designated establishment remains as the sample unit and NOES personnel may seek a court order to require cooperation. The field interview weight for the original establishment is used if the court order brings a completed questionnaire in time for tabulation; otherwise, the case is assigned a field interview weight of zero (later steps in the estimation procedure will cause the establishment to be treated as a non interview and an adjustment will be made for it in the second stage of ratio estimation).

C. Duplicate listings in the D&B universe

Some establishments appear more than once in the D&B file and therefore affect the probability of selection. To make the proper adjustment of the probabilities, the D&B universe file is searched to identify other listings that, if selected, would lead to an already designated field interview sample establishment. The weight adjustments needed for this purpose must be determined on a case by case basis.

1. The unduplication process has been performed under the following rules:

- The search for a duplicate listing of an interviewed sample case is confined to its sample PSU.
- For a given sample establishment, the search seeks additional listings of the same establishment (including units having alternate legal names reported during telephone interview).
- The appearance of a parent firm in D&B does not imply that subordinate establishments are duplicated.
- Additional telephone contacts may be made to resolve questions for ambiguous listings.

2. The rules for adjusting the weight for a sample field interview establishment are as follows:

- If duplicates are located and the size class of the establishments (based on number of employees reported on the D&B records) is the same for all of the "d" duplicates within the PSU, adjust the field interview weight by the factor 1/d.

- If the size class of the duplicate (non sample) establishment is less than reported for the interviewed establishment on its original D&B record, do not adjust the field interview weight for the sample unit.

- If the size class of the duplicate (non sample) establishment is greater than reported for the interviewed establishment on its original D&B record, change the field interview weight for the sample unit to the PSU inflation factor for the PSU. This rule is a compromise equivalent to treating the sample unit as though it had been selected with certainty within the PSU (theoretically, it should be deleted from the sample).

Incl:
Tables 1, 2, 3, 4, and 5

cc:
D. Ward
J. Edmonds
J. Escatell
r. Martindale

Exhibit 8
Memorandum II
July 31 1983

MEMORANDUM TO: J. Waksberg

FROM : R. Hanson

SUBJECT : NOES Estimation Memorandum II--
Ratio Estimation and Final Weighting

This memorandum records specifications for the first and second stages of ratio estimation of the weighting procedures for the NOES. The methods of preparing inflation weights, necessary inputs to ratio estimation, are given in Memorandum I of this series. Memorandum III summarizes the preparatory steps necessary to compute sampling errors using the NOES version of the NASS variance package. A common set of tables (tables 1 through 8) are referred to in these three memoranda.

Westat has the choice of program language for all of the operations described in this memorandum.

I INTRODUCTION

The objective of the NOES estimation procedure is to develop weights to be assigned to the questionnaires completed in the field for sample establishments. Publication estimates of totals based on the NOES sample are to be prepared by summing the weights assigned to questionnaires from interviewed establishments having the characteristic of interest. There are a number of counts and estimates involved in developing the tabulation weights; the definitions of these components are summarized in table 1 and are discussed further as needed in the sections that follow.

The weight assigned to an interviewed establishment to prepare publication estimates is the product of three weights:

- An inflation weight, based on the inverse of the probability of selecting the establishment in the telephone interview sample; two versions are used-- one applies to the telephone sample as originally selected, the other to the sample establishments interviewed by the teams in the field.

- A first stage ratio estimate factor, based on universe counts of D&B establishments by size class and estimates of these counts from the telephone sample, and

- A second stage ratio estimate factor based on universe counts of employees or establishments from County Business Patterns for selected groups of target SIC's by size class and estimates of these counts from the interviewed sample using the product of the inflation weight and the first stage ratio estimate factor.

This memorandum specifies the preparation of the first and second stage ratio estimate factors. The specifications assume:

- We have a record for each establishment originally selected for telephone screening,

- We have a questionnaire for each completed field interview (including substitute questionnaires from the shadow sample where needed),

- All records have been edited and are now ready for tabulation.

- Inflation weights are recorded for each sample establishment according to the specifications provided in Memorandum I of this series.

Section II of this memorandum records specifications for computing first stage ratio factors and Section III the preparation of the second stage ratio factors.

II FIRST STAGE OF RATIO ESTIMATION

A. Introduction

Universe counts of establishments are available from the 1980 DMI file for the target SIC's. It is also possible to prepare estimates of these counts from the establishments designated for the telephone sample. These universe counts and estimates can be prepared by size class (based on number of employees shown on the DMI record for the establishment) and summarized for sets of areas.

The first stage ratio estimation procedure computes ratios of universe counts of establishments to estimated counts of establishments by size class and area (see table 1). Estimates prepared with a first stage tabulation weight for each establishment (defined as the product of the telephone inflation weight for the interviewed establishment and its first stage ratio factor) are expected to have greater precision than estimates prepared with inflation weights alone.

1. Employee size classes used in definitions

An independent sample of D&B establishments was selected within each of 11 size classes for the telephone sample. The number of employees reported on the D&B record determines the employee size class of the establishment (Note that size classes used in the second stage of ratio estimation have different definitions and are based on number of employees reported at interview). The size classes used in the first stage of ratio estimation are given in table 2. The samples selected from size classes 1 through 8 and 11 were selected within a sample of 98 PSU's; 26 of these PSU's are self representing and the balance of 72 are non self representing.

The independent samples selected from D&B establishments in size classes 9 and 10 were designated without regard to the location of the establishment with respect to sample PSU.

2. Ratio procedures depend on size classes

For the sample selcted from size classes 1-8 and 11, ratio estimates are prepared within areas defined as combinations of strata of PSU's. An iterative process (summarized in paragraph B, below) is used to define the areas. The areas will differ by size class.

A fixed set of ratio areas is used for the samples selected from size classes 9 and 10. For each of these two size classes, one area is defined for ratio estimation; each will comprise all 50 States and the District of Columbia.

B. Summary of the ratio procedure for size clases 1-8,11

The derivation of first stage ratio factors is an iterative procedure applied separately for each of the size classes 1 through 8 and class 11. The method is illustrated by the steps that produce a first stage ratio factor for a specific size class; details of these steps are presented in the balance of section II.

1. Establish a rational ordering of the strata.

2. Define a trial first stage ratio estimate area for the size class beginning with the first stratum in the ordering.

3. Compute the first stage ratio factor for the trial ratio area, where:

a. The numerator is the D&B total for the area, and

b. The denominator is an estimate of the numerator prepared from establishments designated for the telephone sample within the area.

4. Test the trial area:

a. If the trial area is acceptable, it is defined as the ratio area for the size class. Assign the ratio to each of the sample establishments selected within the area. To construct the next ratio area, begin at paragraph 2, above.

b. If the area is not acceptable, form a new trial area by adding the next stratum to the area; then follow the process at paragraph 3, above.

C. Order of strata for size classes 1-8,11

The 98 NOES strata have been arranged in two sequences, one for NSR strata and the other for SR PSU's. The sequences are shown in tables 3 and 4 and specify the order of combining strata and their sample PSU's into ratio estimate areas for each size class separately.

The SR PSU's shown in table 4 are arranged so that if a small PSU is not an acceptable weighting area for the size class, the procedure combines it with a larger SR PSU relatively close by and similar in character.

The ordering of NSR PSU's in table 3 attempts to achieve homogeniety among adjacent PSU's in the listing. Homogeniety is considered in terms of proportion of employees in manufacturing, in large firms (1000 or more employees), in petroleum-chemicals-rubber, and geography. The first concern in the ordering is to achieve homogeniety of the strata within pairs of strata (pairs are later used in variance computation) and then consider the criteria in sequencing the pairs. The NSR PSU's are later identified by "pair" number and by "first" or "second" member within the pair.

D. Ratio estimate areas for size classes 1-8, 11

A ratio estimate area for a size class is determined by incorporating information from the strata in the order given by table 3 or 4. A trial area is examined to see if it meets certain conditions. If the conditions are met, the ratio for the trial area becomes the first stage ratio factor and is assigned to the sample establishments. If the conditions are not satisfied by the trial area, an augmented area is constructed by adding the next stratum in the sequence and the conditions are tested again. Except for size classes 9 and 10, ratio areas are made up of all SR PSU's or all NSR strata.

1. Determine the trial ratio area

The ratio areas for each size class are to be constructed in the order implied by the listing of strata in tables 3 and 4. For example, the first trial area for sample establishments in NSR PSU's in each size class will begin with stratum 601 and its sample PSU. If the procedure indicates the ratio area is acceptable for the size class, definition of the next ratio area begins by treating stratum 602 and its sample PSU as the trial area. On the other hand, if this first trial area is unsatisfactory, the trial area is redefined to include 601 and 602 together (and if necessary, 801 and so on) until an acceptable area is determined.

A new area is always started by designating the following stratum in the order as the trial area. A special problem arises in determining the final ratio area in the SR or in the NSR sequence. If a trial area includes the final stratum listed in table 3 or table 4 (i.e., includes NSR stratum 418 or SR PSU 710) and is not a satisfactory ratio area, define the new area by combining the problem area with the preceding ratio area. For example, if an acceptable ratio area for a size class consisting of NSR strata 631 + 214 has been defined and the next trial area (stratum 418) for the same size class is not satisfactory, define the final ratio area to consist of strata 631 + 214 + 418.

2. Information needed to process a trial ratio area for a size class

When a trial area is determined, assemble the following information:

a. The ratio numerator: the universe total of all establishments in the target SIC's in the area in the size class as given by D&B. This will be the sum over all PSU's (sample as well as non sample) in the strata in the ratio area.

b. The ratio denominator: the sample estimate of the numerator obtained by a weighted tabulation of the telephone sample establishments in the size class in all sample PSU's in the strata in the ratio area. To prepare the sample estimate, use the telephone inflation weight WT as defined in Memorandum I of this series.

Note that the telephone sample comprises all establishments originally designated for telephone interview including those found during telephone screening to be out of busines, non respondent, or ineligible for other reasons. All of these original selections are members of the telephone sample including those determined to be duplicated in the universe. The original SIC or size class of a telephone sample firm is not to be altered on the basis of information learned from the telephone screening interview.

c. The unweighted count of telephone sample establishments used in preparing the estimated ratio denominator immediately above.

d. The ratio of the numerator to the denominator rounded to four decimals.

3. Conditions for an acceptable ratio area

To be an acceptable first stage ratio area, both of the following tests must be satisfied:

- The ratio of the numerator to the denominator lies in the rage:

  $$.3333 < \text{Ratio} < 3.0000, \text{ and}$$

- The denominator of the ratio is based on a minimum of four unweighted sample cases.

4. Consequences of the tests

a. If both of the preceding tests are satisfied, the ratio computed for the trial area is the first stage ratio factor for all sample establishments in the size class in the sample PSU's in the area. The ratio, identified as R1, should be entered on the records for all sample establishments in the size class in these sample PSU's.

b. If one of the tests is not satisfied, the trial area is augmented by including the information from the next stratum in the sequence. The instructions for this and subsequent steps begin in paragraph 1, above.

E. Display and Review of Factors

The preparation of ratio factors has been designed for execution by the computer. The process will be performed also for each of the half samples used for variance estimation as described in Memorandum III. We do not intend to intervene to modify ratio factors on an individual basis but there is interest in inspecting the factors produced to be satisfied the specifications produce reasonable results. For this purpose, prepare a review tabulation that shows the following information for each defined ratio area within each size class:

- The sample PSU's in the ratio area for the size class,
- The factor rounded to four decimals,
- The factor numerator,
- The weighted denominator of the factor,
- The unweighted count of telephone sample establishments in the denominator of the factor.

Display the above information in a systematic manner so that, during a visual review, each factor can be associated with the factor definition. The format of this display is left to the convenience of the programmer.

F. First stage tabulation weight

The first stage tabulation weight for a sample establishment is defined as the product of its field interview inflation weight W and the appropriate first stage ratio estimate factor. The first stage tabulation weight is used in the subsequent estimation procedures.

III SECOND STAGE OF RATIO ESTIMATION

A. Introduction

County Business Patterns (CBP) information provides counts of number of establishments and number of employees in establishments by size class by SIC using information from the the 1980 Economic Censuses and the Internal Revenue Service quarterly 941 payroll file. We intend to make use of these data at the national level.

Estimates of the CBP counts can be prepared by tabulating information from NOES interviewed sample establishments using first stage tabulation weights. The SIC and number of employees determined for the sample unit at the time of interview would be used to define the size class and SIC used in these estimates. Note that only those interviewed establishments for which completed questionnaires have been received would be used in the preparation of these estimates.

The ratio of CBP counts to estimates from the NOES will be used as second stage ratio factors; with their use, we expect to improve the precision of the NOES estimates. The final tabulation weight used to prepare NOES publication estimates will be the product of the first stage tabulation weight and the appropriate second stage ratio factor. The source of the ratio numerators will be the CBP, the denominators will be estimates based on the NOES interviews using the first stage tabulation weights.

1. Employee size classes in the second stage

The employee size classes used in the second stage of ratio estimation are determined by number of employees reported at field interview. The classes are defined as follows (note that these size classes differ from the size classes used in the first stage of ratio estimation):

Size class	Number of employees at interview
1	8-9
2	10-19
3	20-49
4	50-99
5	100-249
6	250-499
7	500-999
8	1000-1499
9	1500-2499
10	2500-4999
11	5000 +

2. Second stage ratio methods

We have adopted several methods of second stage ratio estimation to take advantage of the published CBP information that is available. The methods used for a specific sample establishment will depend on the SIC and the number of employees reported for the unit at the time of interview. These specifications therefore assume all sample units show the SIC code and number of employees as reported at interview.

Three groups of sample establishments are considered; separate ratio estimation procedures are used within each of the groups. The following summarizes the procedures; table 7 also provides a summary of the types of procedure used in the groups.

- Group 1: Establishments reporting 10-999 employees and a target SIC listed as Group 1 in table 8. For these establishments, the CBP counts of employees for the sum of all target SIC's within 2-digit SIC's can be determined from CBP publications.

  For units in Group 1, the ratio estimator will use the published CBP total numbers of employees for each of the six size classes numbered 2 through 7 for target SIC's summarized at the 2-digit level. As discussed in section B below, an iteration procedure will be used within each size class to define ratio estimate cells that meet certain criteria; for a ratio cell that fails to meet the criteria, the cell will be expanded by incorporating all target establishments within the 2-digit SIC listed next in table 8.

- Group 2: Establishments reporting 1000 or more employees and a target SIC listed as Group 2 in table 8. For these establishments, the CBP counts of establishments for the sum of all target SIC's within the 2-digit SIC can be determined from CBP publications.

  For units in this group, the ratio estimator will use the published CBP total numbers of establishments in each of the four size classes of larger establishments (8 through 11) for target SIC's summarized at the 2-digit level. An iteration procedure, similar to the process for Group 1 units, will be used to define ratio cells within each size class for this group.

- Group 3: All other NOES in-scope sample establishments

  For units in this group, average ratio factors are used. The average factors are computed and assigned without the iteration process referred to for Groups 1 and 2. Note that the "averages" are computed as the sum of ratios divided by the number of ratios in the sum.

  - For units reported in each of the four largest size classes 8 through 11, use the average over all of the ratios computed within the corresponding size class in Group 2.

  - For units reported in each of size classes 2 through 7, use the average of the factors assigned within the corresponding size class in Group 1.

  - For units reporting 8 or 9 employees, use the average factor for size class 2 (10-19 employees) as computed immediately above.

B. Ratio procedures for Group 1 establishments

The method of constructing ratio factors for Group 1 establishments parallels the system used for the first stage. A rational ordering is established, a trial cell is defined and subjected to an acceptance test. If the trial cell meets the test, the ratio is assigned to all units in the cell; if the test fails, augmented cells are constructed and tested until an acceptable cell is defined. All of these operations are conducted separately within each size class.

1. Order of constructing cells

The ordering of sample records for use in constructing second stage ratio estimate cells is based on combinations of SIC's. Table 8 shows the ordering of 2-digit SIC's (the "chaining order") as it applies to units in group 1. The table implies cells will be constructed in combinations of establishments in target SIC's that have been summarized to 2-digit SIC levels.

2. Define a trial second stage ratio cell

Ratio cells are defined separately and independently within each of the size classes given above in section III A.1. Begin the definition of a ratio cell by using the data from the next 2-digit SIC; for the first ratio cell in a size class, use information for all target SIC's within the 2-digit SIC group 15. If the procedure indicates this ratio cell is acceptable for the size class, the definition of the next ratio cell begins by treating SIC 16 as the trial cell. On the other hand, if this first trial cell is unsatisfactory, the trial cell is redefined to include units in SIC groups 15 and 16 together (and, if necessary, the next SIC group and so on) until an acceptable cell is determined.

A new cell is always started by designating establishments in the SIC group that follows the previous cell. A special problem arises in determining the final ratio cell in the size class. If a final trial cell includes the last SIC group and is unsatisfactory, combine the trial cell with the immediately preceding cell in the same size class.

3. Information needed for a trial ratio cell

When a trial cell is determined, assemble the following information:

a. The ratio numerator: The CBP universe count of employees published for the establishments in the target SIC group comprising the trial cell. In most cases these employee counts are available at the U.S. level in the CBP publication.

In a few situations, the publication shows (D) instead of the desired employee counts; this symbol means publication of the number is withheld because of "disclosure". That is, information for an individual establishment could be revealed if the data were published. If the CBP publication for a size class in an SIC shows a disclosure, assign the average of the ratio factors computed for the other (non disclosure) groups in the size class. This may mean skipping over a 2-digit SIC to define an acceptable ratio cell. The assignment of the average of other ratio factors is discussed more completely in section III D, below.

b. The ratio denominator: the sample estimate of the numerator obtained by tabulating the establishments for which a completed field interview questionnare is available using the first stage tabulation weights. Note that the estimate for the denominator must be defined in the same manner as the numerator.

Do not include field interview questionnaires in estimates of the denominator for cells involving disclosure.

c. The ratio of the numerator to the denominator rounded to four decimals.

d. The number of completed field interview questionnairs used in estimating the denominator.

4. Conditions for an acceptable ratio cell

To be an acceptable second stage ratio cell, both of the following tests must be satisfied:

- The ratio of the numerator to the denominator lies in the range:

  .3333 < Ratio < 3.0000, and

- The denominator of the ratio is based on a minimum of four unweighted sample establishments for which a field questionnaire is available (non-interview establishments are therefore not included in this count).

5. Consequences of the tests

a. If both of the preceding tests are satisfied, the ratio computed for the cell is the second stage ratio factor for all sample establishments in the cell. The ratio, identified as R2, should be assigned to all establishments in the cell.

b. If one (or both) of the tests is not satisfied, the trial cell is augmented by including the information from the same size class in the 2-digit SIC next in the sequence given in table B. The instructions for this and subsequent steps begin in paragraph III B1, above.

C. Ratio procedures for Group 2 establishments

The ratio estimation procedures for establishments in Group 2 are the same as for Group 1 with one exception:

The numerators and denominators of the ratios are constructed in terms of number of establishments (rather than number of employees as in Group 1). The procedures are also simplified somewhat in that CBP counts of establishments are never suppressed because of disclosure.

The chaining order of 2-digit SIC's and the iterative procedure for defining ratio cells is the same as for Group 1 establishments.

D. Ratio procedures for Group 3 establishments

Use a single ratio factor in each of the ten size classes numbered 2 through 11 in section III A1, above for all Group 3 establishments. The procedures do not require testing of the ratios by cell as only one factor is assigned to all Group 3 establishments in each size class.

Compute the factors for establishments in Group 3 after the factors have been prepared for Groups 1 and 2. Note that average factors as computed in this section are also used to resolve the disclosure issue discussed in section III B 3a, above.

1. The factors for size classes 2 through 11 are computed as follows:

Within each of size classes 2 through 11, compute the average of all assigned ratio factors in Groups 1 and 2; the average is the sum of the ratios divided by the number of ratios in the sum. Do not include in the average any of the cells in Group 1 in which the disclosure problem has appeared.

2. Assign the average factor computed for size class 2 to all firms that are in size class number 1.

3. Assign the average ratios to field interview questionnaires in size classes within the 2-digit SIC group where the disclosure problem has occurred.

E. Display and review of factors

Display the following information in a systematic manner so that during a visual review, each factor can be associated with the factor definition. Prepare review tabulations as follows:

1. For factors in establishment Groups 1 and 2:

   - o The factor rounded to four decimals,
   - o The factor numerator,
   - o The factor denominator,
   - o The unweighted count of completed field interview questionnaires for sample establishments appearing in the denominator of the factor.

2. For factors from establishment Group 3:

   - o The factor rounded to four decimals,
   - o The unweighted count of completed questionnairs.

F. Final weight for publication

The final weight used to prepare estimates for publication (also called the "replication zero" weight) is defined as the product of the field interview weight W, the first stage ratio factor (R1), and the second stage ratio factor (R2) = W * R1 * R2.

cc:
D Ward
J Escatell
J Edmonds
R Martindale

Exhibit 9
Memorandum III
May 24 1983

MEMORANDUM TO: J. Waksberg

FROM : R. Hanson

SUBJECT : NOES Estimation Memorandum III--
Preparation for Variance Estimation

This memorandum describes the one-time preparatory steps we must perform to compute variance estimates. Memorandum I of this series specifies the computation of inflation weights; Memorandum II, the ratio estimation factors and the final weights for tabulation.

Westat has the choice of program language for the operations in this memorandum.

I Introduction

A. Summary of the NOES variance estimator

For the NOES, we define a set of random subsamples from the full sample with each subsample equal to 50 percent of the full sample. The same principles of selection and stages of sampling used for the full sample are used in each of the subsamples. The subsamples are referred to as replicates. By applying all of the procedures in the regular estimation process to each of the replicates and then computing the dispersion among the estimates from the replicates, we can obtain an approximation of the sampling error of estimates prepared from the complete NOES sample.

The method consists of the following three steps, this memorandum describes the first two steps, the third step is executed by use of the VARNOES program package:

- Identify the sample units that make up each of the half sample replicates.
- Determine the weights of the establishments in each of the replicates. The same estimation procedure, prepared for the full sample, is applied separately to each of the replicates.
- Compute the dispersion of the estimates and approximate the variance of the full sample; VARNOES uses a relatively simple computation formula that does not depend on the form of the estimate for which the variance is to be approximated.

B. Preparation of data for the VARNOES program package

The operations discussed in this memo are meant to enumerate steps required to prepare the NOES data file for processing in the variance estimation package. These preliminary steps will involve defining a number of replicates of the NOES sample and determining the weights for the sample in each of the replicates. For estimating the total variance, 32 weights in addition to the final weight for tabulation must be provided for each interviewed sample establishment; for estimating variance components, an additional 32 weights will be used for each unit.

1. Weights for each sample establishment required for estimating total variance are described as follows:

- The weight used in preparing the estimate for which the sampling error is to be approximated (the "replication zero" weight). Usually this will be the weight used to produce publication estimates and would have all of the first and second stage weight computation steps outlined in Memoranda I and II of this series.

- Weights defined as above but for use with 50 percent subsamples of establishments. For estimating the total variance (e.g., to approximate the sampling error of a publication estimate) we intend to use 32 different 50 percent subsamples of the establishments in the complete sample. The half samples (replicates) are identified by numbers 1 through 32.

2. Weights for each sample establishment required for estimating variance components are prepared by procedures that parallel the preparation of weights for estimating total variance. Several estimation processes may be investigated by the use of these weights; for example, one process would require:

- The replication zero weight described in paragraph 1 above.

- Weights computed for a second set of 32 replications where each of these replications involves subsamples within all of the sample PSU's (in contrast to the procedure in paragraph 1 where half samples in some cases are obtained by sampling PSU's).

II Estimation of the total variance

We have defined 32 half sample replicates for computing estimates of the total variance. A record for each of the interviewed establishments must accomodate 33 weights for this purpose, one weight for replicate zero and a weight for each of the 32 half sample replications. If an establishment is not a member of a given replicate, it is assigned a zero weight; if it is in a replicate, it is assigned a non zero weight by the instructions presented later in this memorandum.

In general, half samples are constructed by treating first stage sampling units in pairs and dividing the total sample within each pair into two members such that a sample of one member from each pair will reflect the survey sampling process. Each self representing PSU is considered a pair and half samples of establishments within each size class are defined within each PSU. NSR PSU's are treated in pairs and half samples are defined in terms of whole PSU's. Size classes 9 and 10 are each considered pairs and half samples of establishments are defined within each of these size classes.

The pairs are assigned numbers; each of the 26 SR PSU's is assigned a number, 1 through 26; pairs numbered 27 and 28 are assigned to size classes 9 and 10; the 36 pairs of strata in NSR are numbered 29 thru 64. Half samples within each pair are identified as the "first member" (code 1) or the "second member" (code 0) of the pair. This method of defining half samples within pairs is used with table 6 to define replications.

A. Definition of half samples

1. Half samples in self representing PSU's

Two half samples within each of the 26 SR PSU's are defined by assigning alternate selections of the original telephone sample units within size classes within the PSU. Estimation of the total variance requires this method of defining half samples to be done only for the SR PSU's; however, when components of variance are estimated, two half samples are also defined in each NSR PSU in the same way. The instructions that follow therefore apply to all PSU's even though only the results for SR PSU's are required to compute the total variance.

A list of the SR PSU's is given in table 4. Two half samples are to be designated within each SR PSU, pair numbers from 1 to 26 have been assigned to the SR PSU's. Thus two half samples are selected within pair 1 (PSU 110), and within pair 2 (PSU 150), etc. The following procedure is used to designate half samples within the PSU's:

a. Assemble all of the original telephone sample records for the PSU's; records for all of the initial selections are required including units deleted by telephone or field interview information (out of business, out of scope, refusals, etc.). The records must show the telephone inflation weight (prepared by the specifications in Memorandum I) as it will be needed in later steps. Exclude telephone D sample records in those PSU's within which D sample cases were not assigned for telephone interview (i.e., PSU's without asterisks in tables 3 or 4).

b. Sequence the telephone sample records by Westat ID number. This sequencing should order the records as follows:

Major: PSU ID (98 different PSU's),

Intermediate: Size class (9 classes, 1 thru 8 and 11) within PSU,

Minor: Order of selection within size class within PSU.

c. With the sequence as above, identify two half samples within each size class within the PSU; one half sample consisting of the 1st, 3rd, 5th, etc. case in the sequence, the other consisting of the remainder (the 2nd, 4th, etc.).

The telephone sample units in the two half samples in the PSU are now assigned codes; code 1 will identify the "first member", code 0 the "second member" of the pair of half samples within the PSU. The following procedure identifies the first and second members within each PSU. The procedure is applied separately within each size class.

- Form the sum of the five digits in the PSU and size class identifications (the 3 digits in the PSU ID plus the 2 digits in the size class ID). For example, the desired sum for size class 01 in PSU 150 would equal 7.

- If the sum is odd, assign code 1 (denoting "first member") to each of the telephone sample firms in the size class listed first, third, fifth, ..etc. and assign code 0 (denoting "second member") to the firms listed second, fourth, sixth, ..etc. in the sequence within the size class.

- If the sum is even, assign the codes as immediately above but interchange the 0 and 1 codes.

- Repeat for all size classes in the PSU.

d. This method assigns a "first member" or "second member" identification to every telephone sample unit designated in the PSU. All sample units in the PSU with code 1 from all size classes comprise one of the half samples and those coded 0 comprise the other. This process identifies the two half samples of telephone sample establishments within the pair (i.e., within the PSU).

e. Establishments interviewed in the field are assigned to half samples by associating them with the telephone sample units in the half samples. If shadow sample establishments have been substituted, they are to be associated with the original telephone sample case to determine the half sample identification.

2. Half samples in size classes 9 and 10

All telephone sample establishments within size class 9 are members of pair number 27 and all telephone sample establishments within size class 10 are members of pair number 28. The following instructions, applied separately for size class 9 and then for 10, define the required half samples within these pairs. Note that samples from size classes 9 and 10 were originally selected without regard to establishment location so that PSU identifications are not involved in the definition of these half samples.

a. Assemble all of the original telephone sample records for the size class. All of the original selections are required including cases deleted by telephone or field interview information (e.g., out of scope, out of business, refusals, etc.). All of workload subsamples A, B, C, and D are to be included.

b. Sequence the telephone sample records as follows:

Major: Size class (2 classes, 9 and 10),

Minor: Order of selection within size class.

c. With the telephone sample records in the above sequence, assign code 1 ("first member") to the 1st, 3rd, 5th...etc records in the size class; assign code 0 ("second member") to the balance of the records,i.e., to the 2nd, 4th, 6th...etc records in the size class. These steps assign two half samples for size class 9; repeat the process in the same way for records in size class 10. Note that this is a simplification of the system used to identify members of pairs within PSU's given in paragraph II A1c, above.

3. Half samples in NSR PSU's

The system for assigning half samples witin pairs for estimating total variance is much simpler for NSR PSU's. In table 3, the 72 NSR PSU's have been put into 36 pairs and identified as pair numbers 29 thru 64. This table also identifies in each pair the PSU chosen as the "first member" and the "second member". All telephone sample records and records for establishments interviewed in the field for all size classes within a NSR PSU are assigned the pair number and member identification given in table 3.

A different method is used to define half samnples in NSR PSU's when estimating some variance components (section III B3).

B. Replications for estimation of total variance

In this section, 50 percent subsamples of the total NOES sample are defined by choosing one member (the first or second) from each of the 64 pairs defined above in section A. Each of these 50 percent subsamples is called a replication or a replicate. To compute approximate sampling errors of publication estimates, we shall define a total of 32 of these replications.

It is possible to prepare somewhat more precise estimates of the sampling error by basing the computations on a larger number of replications. For estimating the total variance, 64 replications and for approximating the within PSU variance, 100 replications could be used. The number 32 has been selected for both purposes to reduce costs.

Table 6 shows the members selected from each of 64 pairs to define 32 half sample replications for estimating the total variance. For example, the entries on the first row of table 6 show how replicate number 1 is constructed; it consists of the half samples made up of the second members of pairs 1, 2, 3, 4; the first member of pair 5; the second member of pair 6;...etc. Table 6 is also used, with the somewhat different instructions given in section III of this memorandum, to define replicates for estimating variance components.

Incorporating information from tables 3 , 4 and 6 provides a shorthand way of identifying all of the replications we shall use for estimating total vaiance. For example, the half samples for replicates 5 and 6 comprise all establishments identified as follows:

1. From the 26 SR PSU's (see tables 4 and 6), sample units in the first or second members within the SR PSU's as follows--

Pair number	SR PSU	Replicate 5	Replicate 6
1	110	First	Second
2	150	Second	First
3	120	Second	Second
4	142	First	Second
5	130	Second	First
:	:	:	:
25	731	First	First
26	710	First	First

2. From size classes 9 and 10 (see tables 4 and 6), sample units in the first or second members within the size classes as follows:

Pair number	Size class	Replicate 5	Replicate 6
27	9	Second	First
28	10	Second	Second

3. From the 72 NSR PSU's (see tables 3 and 6), all sample units in the following NSR PSU's

Pair number	Replicate 5	Replicate 6
29	601	602
30	801	801
31	803	401
32	603	603
33	604	605
:	:	:
63	631	417
64	418	418

C. Determine weights for sample units in each replicate

Each half sample replicate is to be subjected to the same weighting procedure that has been applied to the full sample. Theoretically, all of the steps in defining weights discussed in Memoranda I and II are to be applied. However, some modifications are required because the replicate samples are only half as large. Each sample establishment will have 32 additional weights assigned to it, one weight for each of the 32 replications. About half of these 32 weights for an establishment will be zero (a zero weight will indicate the unit is not in sample in the replicate) and the rest of the weights will be about twice as large as the weight for replication zero.

The following modifications are added to the appropriate portions of the weighting instructions in Memoranda I and II to produce the required weights for the establishments in each half sample replicate:

1. Telephone inflation weights

Double the original telephone inflation weight assigned to the unit if the establishment is in sample for the replicate. Assign zero as the telephone sample weight if the establishment is not in sample for the replicate.

2. Field interview inflation weight

The telephone units determined to be in sample for a given replicate determine the field interview units in sample for the replicate. Associate the record for each completed field interview with its corresponding telephone sample record. Double the field interview inflation weight for the establishment and assign it as the field interview weight for each replicate for which it is in sample. Assign zero as the field interview inflation weight for the replicates for which the establishment is not in sample.

3. First stage ratio factors for replicates

a. Compute first stage ratio estimate factors

Perform all of the steps specified for first stage ratio estimation in Memorandum II, sections II A through II D. All of these steps are to be performed separately for each replicate.

b. Display and review of factors

Display and review the factors for each replicate as specified for replicate zero in Memorandum II, section IIE.

c. First stage tabulation weight

The first stage tabulation weight for a replicate is given by the definition in Memorandum II, section II F using the appropriate weights for the replicate.

4. Second stage ratio factors

a. Numerators of second stage ratio factors

The numerators of second stage ratio factors are the same as for replicate zero, see Memorandum II, section III B.

b. Denominators of second stage ratios

Construct the denominators of the second stage factors for each replicate using the instructions of Memorandum II, section III A with the weights appropriate for the replicate.

c. Display and review of ratio factors

The display and review of the second stage factors will parallel the system to be determined for replication zero.

d. Final tabulation weight

The final tabulation weight for a replicate is constructed as the product of the field interview inflation weight times the first stage ratio factor times the second stage ratio factor. All weights in this product are derived from the replicate.

III ESTIMATION OF VARIANCE COMPONENTS

A. Introduction

The variance computation process can be used for purposes other than approximating the precision of estimates published for the NOES. The procedures can also be used to study the efficiency of estimation procedures and to measure the efficiency of other aspects of the sample design. This can be done by estimating the variance under alternative assumptions. We offer two examples of this system; the variance computation package will permit both of these analyses.

1. Efficiency of estimators

The impact of each of the two stages of ratio estimation can be evaluated. This is done by recomputing variances for estimates that do not have one (or neither) of the ratio estimates incorporated in the NOES weighting system. The effect of a ratio estimator is eliminated by assigning ratio factors of 1.0. These analyses can be accomplished without the need to define additional half sample replications.

2. Efficiency of the sample design

The allocation of the total sample to the size classes can be reviewed by analysing variances contributed by the sample selected from each of the D&B size classes. This analysis also does not require defining additional replications.

It is possible to measure the contribution to the variance caused by confining field interviews to a sample of PSU's. This process requires a variance computation that treats all PSU's (including NSR PSU's) as though they were self representing. This computation requires defining a different set of half samples within the 72 NSR PSU's and combining these half samples with the 28 pairs already defined in SR and for size classes 9 and 10 for the purpose of estimating the total variance. A new set of 32 half sample replicates will have to be assigned within the NSR PSU's.

B. Definition of half samples

This section specifies the method of defining half samples for the purpose of estimating the variance component caused by confining the interviews to a sample of (rather than all) establishments within the PSU's. This variance component is called the "within PSU variance". The half samples needed for this purpose include the same half samples defined in SR and for size classes 9 and 10 by the instructions in sections II A1 and I1 A2; however, the half samples differ from those designated in NSR PSU's for estimation of total variance. We should insure that the different data files constructed for these purposes cannot be confused with each other.

1. Half samples in SR PSU's

Use the same half samples for SR PSU's as defined for estimation of total variance (see section II A1).

2. Half samples in size classes 9 and 10

Use the same half samples for these size classes as defined for estimation of total variance (see section II A2).

3. Half samples in NSR PSU's

Define half samples within NSR PSU's for estimating the within PUS variance component as follows:

- Apply the instructions of section II A1 to each NSR PSU separately. The effect of this is to treat each NSR PSU as though it is self representing; as a result, we can identify first and second half sample "members" of the sample within each NSR PSU.

- Observe the pairs of NSR PSU's identified in table 3. Form a new first member of the sample within a pair of NSR PSU's by combining the "first members" (determined above) of the two NSR PSU's in the pair. The new second member is the combination of the "second members" of the two NSR PSU's.

C. Replications for estimation of variance within PSU's

In this section, replications of the full NOES sample are defined for approximating the variance within PSU's.

Table 6 will be used to define 32 half sample replications. This means that the half samples already identified from pairs 1 through 28 by the instructions in sections II B1 and II B2 will be part of the corresponding 32 replications defined here.

The following instructions define the samples from pairs 29 through 64 for each replication; these half samples should be combined with the half sample replicates already prepared for pairs 1 through 28. The process is illustrated by the construction of replicates 5 and 6. These replicates are to comprise;

1. The half samples from the 26 SR PSU's identified for replicates 5 and 6 in section II B1.

2. The half samples from size classes 9 and 10 identified for replicates 5 and 6 in section II B2.

3. The half samples from the 72 NSR PSU's for which the first and second pair members identified in section III B3 and combined as directed in table 6:

Pair number	Replicate 5	Replicate 6
29	First	Second
30	First	First
31	Second	First
32	Second	Second
33	First	Second
:	:	:
63	Second	First
64	Second	Second

D. Determine weights for sample units in each replicate

The sample units in each of the 32 half sample replicates defined in section III C above are to be subjected to the complete NOES weighting procedure given in Memoranda I and II. The weighting process must be adapted, however, because the replicate samples are only half as large as the full sample.

Assign weights to the units in each of the 32 replications by following the instructions given above in section II C.

E. Computation of the within and between PSU variance components

Run the VARNOES package on the 32 replications produced in section III D; this operation approximates the within PSU variance component of the total variance. The difference of the total variance (approximated by the process given in section II) and the within PSU variance is an approximation of the between PSU variance.

It will not be necessary to supply a COBOL interfaced VARNOES program for computation of the within PSU variance or for computing the difference of the total and the within PSU variance estimates. We intend to do all of these computations at Westat. However, it will be necessary to have a COBOL interfaced program for the computation of the total variance.

Incl:
Tables 1, 3, 4, 6

cc:
D. Ward
J. Edmonds
J. Escatell
R. Martindale

Table 1

Components of weights used in the NOES estimation procedure

	Notation
INFLATION WEIGHTS	
Telephone sample weight: Assigned to each telephone sample establishment and based on the inverse of the probability of selecting D&B units in the sample; used to prepare unbiased estimates of the D&B universe.	WT
Field interview weight: Assigned to each sample establishment interviewed in field, based on adjusted telephone sample weight.	W
RATIO ESTIMATE FACTORS	
First stage factor	R1
Numerator: D&B universe establishment counts by size class and area.	
Denominator: Estimates of numerator from telephone sample using weights = WT.	
Second stage factor	R2
Numerator: County Business Pattern universe employee counts (establishment counts for larger firms) by current size, and SIC.	
Denominator: Estimates of numerator from interviewed establishments using weights = W * R1	
SUMMARY OF WEIGHTS TO USE FOR:	
Unbiased estimates of D&B telephone universe	WT
Unbiased estimates of universe from field interviews	W
First stage ratio estimates of universe from field interviews	W * R1
NOES estimates for publication	W * R1 * R2

July 26 1983

Table 2

Establishment size class and
theoretical telephone sample weights
NOES

Size class	Number of D&B employees a	Theoretical weights for telephone sample ABC PSU's	ABCD PSU's	Chicago
1	8-19	182.42	159.618	169.687
2	20-49	114.52	100.205	106.527
3	50-99	60.365	52.819	56.152
4	100-249	33.386	29.213	31.056
5	250-499	19.440	17.010	19.083
6	500-999	13.437	11.757	12.499
7	1000-1499	10.584	9.261	9.845
8	1500-2499	7.670	6.711	7.135
9 b	2500-4999	-	4.436	-
10 b	5000+	-	1.752	-
11	NA	182.42	159.618	169.687

---

a Size classes defined by number of employees reported by D&B; size classes used in second stage of ratio estimation are different.

b ABCD samples selected from size classes 9 and 10 are not confined to sample PSU's.

Table 3

Order for combining NSR PSU's into first stage ratio groups and collapsed pairs of NSR PSU's for variance estimation

Chain order	Pair number	Psu's in pair: First member	Psu's in pair: Second member
1-2	29	601*	602*
3-4	30	801*	802
5-6	31	401	803*
7-8	32	804*	603*
9-10	33	604	605*
11-12	34	606	607*
13-14	35	805*	806*
15-16	36	402*	403
17-18	37	404*	608*
19-20	38	609	610
21-22	39	611*	807*
23-24	40	808*	405
25-26	41	201	202
27-28	42	612	613
29-30	43	406*	614
31-32	44	615	616
33-34	45	203	204
35-36	46	617*	618*
37-38	47	205	206*
39-40	48	619*	620
41-42	49	207	208*
43-44	50	209	210*
45-46	51	407	408*
47-48	52	409	410
49-50	53	411	621
51-52	54	622*	623
53-54	55	624	809*
55-56	56	211	625
57-58	57	412	212
59-60	58	413	626*
61-62	59	414*	213*
63-64	60	627	628*
65-66	61	415	416
67-68	62	629	630*
69-70	63	417	631*
71-72	64	214	418*

* Workload subsamples ABCD assigned in the PSU (workload subsamples ABC in all other PSU's).

Table 4

Order for combining PSU's in first stage ratio groups
Self representing strata

Chain order	Pair number	PSU's in group
73	1	110
74	2	150
75	3	120
76	4	142
77	5	130
78	6	160*
79	7	552*
80	8	542
81	9	381*
82	10	350
83	11	320*
84	12	371
85	13	# 310*
86	14	520*
87	15	330
88	16	752*
89	17	392*
90	18	340*
91	19	720*
92	20	761*
93	21	530*
94	22	511
95	23	561*
96	24	742*
97	25	731*
98	26	710*
99	27	999* &
100	28	999* &

---

* Workload subsamples ABCD interviewed in the PSU (workload subsamples ABC in all other PSU').

Within PSU selection probabilities differ from other PSU's, see text.

& Pairs 27 and 28 refer to size classes 9 and 10, respectively.

Table 5

Telephone sample weights for establishments
in PSU's having size classes
sampled with certainty

SIZE CL-> / WT CUTS->	3 26.410	4 14.617	5 8.498	6 5.893	7 4.627	8 3.373
PSU						
204				14.928	14.928	14.928
205				18.391	18.391	18.391
* 206			23.968	23.968	23.968	23.968
207				16.857	16.857	16.857
212				13.609	13.609	13.609
* 213					10.378	10.378
401			20.295	20.295	20.295	20.295
* 404			26.61	26.61	26.61	26.61
405				16.176	16.176	16.176
* 406		32.1	32.1	32.1	32.1	32.1
407			20.194	20.194	20.194	20.194
410						8.599
411			30.542	30.542	30.542	30.542
412		38.432	38.432	38.432	38.432	38.432
413			19.508	19.508	19.508	19.508
* 414		32.654	32.654	32.654	32.654	32.654
415			27.38	27.38	27.38	27.38
416				18.075	18.075	18.075
417		37.383	37.383	37.383	37.383	37.383
* 601			23.698	23.698	23.698	23.698
604		38.56	38.56	38.56	38.56	38.56
* 605			27.286	27.286	27.286	27.286
606					11.099	11.099
* 607			19.84	19.84	19.84	19.84
* 611				14.146	14.146	14.146
612					10.75	10.75
613						8.146
614						8.277
615		39.218	39.218	39.218	39.218	39.218
616		36.391	36.391	36.391	36.391	36.391
* 617		41.442	41.442	41.442	41.442	41.442
* 618			24.118	24.118	24.118	24.118
* 619		36.636	36.636	36.636	36.636	36.636
621			33.193	33.193	33.193	33.193
623						7.911
624			21.106	21.106	21.106	21.106
625		49.771	49.771	49.771	49.771	49.771
* 626			20.402	20.402	20.402	20.402
627			27.547	27.547	27.547	27.547
* 628		42.568	42.568	42.568	42.568	42.568
629			33.239	33.239	33.239	33.239
* 630		33.236	33.236	33.236	33.236	33.236
* 631		36.058	36.058	36.058	36.058	36.058
802				16.372	16.372	16.372
* 803	57.98	57.98	57.98	57.98	57.98	57.98
* 804					10.726	10.726
* 806					9.866	9.866
* 808				13.002	13.002	13.002

* Workload subsamples ABCD assigned in the PSU.

July 20 1983

Table 6

Definition of replicates for variance estimation
NOES sample design

REPLICATE	PAIR NUMBER *															
	1		2		3		4		5		6					
	1234	5678	9012	3456	7890	1234	5678	9012	3456	7890	1234	5678	9012	3456	7890	1234
PSU'S-->	1111 1155 3333 3537 3377 5557 77 1524 3654 8527 1235 9426 3164 31## (...SEE TEXT FOR DEFINITION OF NSR PAIRS...) 0002 0022 1001 0002 2001 0112 10															
1	0000	1010	1110	1100	0111	1100	1101	0010	0000	1010	1110	1100	0111	1100	1101	0010
2	1000	0101	0111	0110	0011	1110	0110	1000	1000	0101	0111	0110	0011	1110	0110	1000
3	0100	0010	1011	1011	0001	1111	0011	0100	0100	0010	1011	1011	0001	1111	0011	0100
4	0010	0001	0101	1101	1000	1111	1001	1010	0010	0001	0101	1101	1000	1111	1001	1010
5	1001	0000	1010	1110	1100	0111	1100	1100	1001	0000	1010	1110	1100	0111	1100	1100
6	0100	1000	0101	0111	0110	0011	1110	0110	0100	1000	0101	0111	0110	0011	1110	0110
7	1010	0100	0010	1011	1011	0001	1111	0010	1010	0100	0010	1011	1011	0001	1111	0010
8	1101	0010	0001	0101	1101	1000	1111	1000	1101	0010	0001	0101	1101	1000	1111	1000
9	0110	1001	0000	1010	1110	1100	0111	1100	0110	1001	0000	1010	1110	1100	0111	1100
10	0011	0100	1000	0101	0111	0110	0011	1110	0011	0100	1000	0101	0111	0110	0011	1110
11	1001	1010	0100	0010	1011	1011	0001	1110	1001	1010	0100	0010	1011	1011	0001	1110
12	1100	1101	0010	0001	0101	1101	1000	1110	1100	1101	0010	0001	0101	1101	1000	1110
13	1110	0110	1001	0000	1010	1110	1100	0110	1110	0110	1001	0000	1010	1110	1100	0110
14	1111	0011	0100	1000	0101	0111	0110	0010	1111	0011	0100	1000	0101	0111	0110	0010
15	1111	1001	1101	0100	0010	1011	1011	0000	1111	1001	1101	0100	0010	1011	1011	0000
16	0111	1100	1110	1010	0001	0101	1101	1000	0111	1100	1110	1010	0001	0101	1101	1000
17	0011	1110	0111	0101	0000	1010	1110	1100	0011	1110	0111	0101	0000	1010	1110	1100
18	0001	1111	0011	1010	1000	0101	0111	0110	0001	1111	0011	1010	1000	0101	0111	0110
19	1000	1111	1001	1101	0100	0010	1011	1010	1000	1111	1001	1101	0100	0010	1011	1010
20	1100	0111	1100	1110	1010	0001	0101	1100	1100	0111	1100	1110	1010	0001	0101	1100
21	0110	0011	1110	0111	0101	0000	1010	1110	0110	0011	1110	0111	0101	0000	1010	1110
22	1011	0001	1111	0011	1010	1000	0101	0110	1011	0001	1111	0011	1010	1000	0101	0110
23	1101	1000	1111	1001	1101	0100	0010	1010	1101	1000	1111	1001	1101	0100	0010	1010
24	1110	1100	0111	1100	1110	1010	0001	0100	1110	1100	0111	1100	1110	1010	0001	0100
25	0111	0110	0011	1110	0111	0101	0000	1010	0111	0110	0011	1110	0111	0101	0000	1010
26	1011	1011	0001	1111	0011	1010	1000	0100	1011	1011	0001	1111	0011	1010	1000	0100
27	0101	1101	1000	1111	1001	1101	0100	0010	0101	1101	1000	1111	1001	1101	0100	0010
28	1010	1110	1100	0111	1100	1110	1010	0000	1010	1110	1100	0111	1100	1110	1010	0000
29	0101	0111	0110	0011	1110	0111	0101	0000	0101	0111	0110	0011	1110	0111	0101	0000
30	0010	1011	1011	0001	1111	0011	1010	1000	0010	1011	1011	0001	1111	0011	1010	1000
31	0001	0101	1101	1000	1111	1001	1101	0100	0001	0101	1101	1000	1111	1001	1101	0100
32	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

* See text for definition of pairs
code 1: First member,
code 2: Second member.

\# Pairs 27 and 28 are for size classes 9 and 10.

May 23 1983

Table 7

Summary of second stage ratio estimation methods in three groups of NOES sample establishments

Number of employees at establishment when interviewed									
8–9	10–19	20–49	. . .	500–999	1000–1499				5000+
	Group 1 Establishments Ratio factors based on published counts and estimates of number of CBP employees. Ratio cells determined by iteration.				Group 2 Establishments Ratio factors based on published counts and estimates of number of CBP establishments. Ratio cells determined by iteration.				
	Group 3 Establishments Ratio factors based on averages of ratios computed for corresponding size classes in Groups 1 and 2. Iteration not involved.								

August 29 1983

Table 8

Order for combining 2-digit SIC summaries to define second stage ratio estimate cells by reported establishment size and SIC

GROUP 1: Establishments reporting 10-999 employees in the following SIC's:

The following chaining order of establishments by SIC is used:

	Order	SIC		Order	SIC	
D	1	15		21	36	
	2	16		22	37	
	3	17		23	38	
	4	20		24	39	
	5	21		25	41	= 411+412+415+417
	6	22		26	44	
	7	23		27	45	
	8	24	D	28	46	
	9	25		29	48	
	10	26		30	49	
	11	27	D,L	31	50	= 501+503+505+5093
	12	28	D,L	32	51	= 516+517
	13	29		33	55	= 552+553+554
	14	13	D	34	72	= all of 72 less 7218
	15	30	D,L	35	73	= 733+734+7391+7395+ 7397+7399
	16	31	D	36	75	= all of 75 less 752
	17	32		37	76	
	18	33		38	84	
	19	34				
	20	35				

D CBP employee counts for one or more size classes will show "Disclosure".

L CBP count of large establishments (more than 1000 employees) cannot be determined for the 2-digit SIC; Group 3 ratio procedure used for large firms.

(more)

Table 8 (Continued)

GROUP 2: Establishments reporting 1000 or more employees in SIC's listed in Group 1:

The Group 1 chaining order is also used in Group 2. Establishments reporting 1000 or more employees in the following SIC's are assigned to Group 3:

50 = 501+503+505+5093
51 = 516+517
73 = 733+734+7391+7395+7397+7399

GROUP 3: Establishments reporting 8 or 9 employees in any of the SIC's enumerated in Groups 1 and 2 plus all establishments reporting 8 or more employees in the following SIC's:

0723	422
0724	423
0742	4742
0782	478
0783	8062
4013	807
4212	809
4214	

Exhibit 10

THIS STUDY IS AUTHORIZED BY THE OCCUPATIONAL SAFETY AND HEALTH ACT OF 1970 [27 U.S.C. 669(a)]

OMB # 68S80036
EXPIRES: September 30, 1981

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH

NATIONAL OCCUPATIONAL HAZARD SURVEY II

CASE ID # |__|__|__| - |__|__|__|__| - |__|__| - |__|__|

Prepared by:

WESTAT
An Employee-Owned Research Corporation
11600 Nebel Street • Rockville Maryland 20852 • 301 881-5310

Hello. My name is _____ with Westat, a national research company near Washington, D.C. We are conducting a study for the National Institute for Occupational Safety and Health.

1. First, am I correct that this is (NAME OF ESTABLISHMENT)?

YES 1 (Q4)
NO 2 (Q2)

2. What is the name of this establishment?

NAME: _____

3. Are you in any way affiliated with (NAME OF ESTABLISHMENT)?

YES 1 (Q4)
NO. . . . . 2 → Thank you very much but I seem to have reached the wrong establishment.

4. Who would be the best person in your organization to speak to about the nature of your business and the number of employees you have? (RECORD NAME, TITLE, AND PHONE NUMBER OF R BELOW. IF R IS NOT AVAILABLE, ALSO RECORD NAME, PHONE NUMBER, AND BEST TIME TO CALL BACK ON THE SCREENING CALL RECORD.)

1st REFERRAL
NAME: _____
TITLE: _____
PHONE #: ( ) _____

2nd REFERRAL
NAME: _____
TITLE: _____
PHONE #: ( ) _____

3rd REFERRAL
NAME: _____
TITLE: _____
PHONE #: ( ) _____

BOX A

INTERVIEWER: REVIEW Q4 AND CIRCLE ONE:

ORIGINAL R IS BEST RESPONDENT . . . . . 1 (INTRODUCTION #1)
NEW R IS BEST RESPONDENT. . . . . 2 (INTRODUCTION #2)

INTRODUCTION #1 - ORIGINAL RESPONDENT

Before we begin, I would like to give you a little more information about the study. The study is authorized by the Occupational Safety and Health Act of 1970.

The main study, called the National Occupational Hazard Survey, is a special research project designed to obtain basic information about health and safety practices within workplaces.

The purpose of this initial part of the study, however, is to verify some information about your establishment so we can be sure that establishments like yours are adequately represented in this study. Your responses are voluntary, however your cooperation is needed to make the results of this study comprehensive, accurate and timely. The questions will take about five minutes to answer. (Is this a convenient time to talk?) (IF YES, BEGIN INTERVIEW WITH Q5. IF NOT A CONVENIENT TIME, RECORD RESPONDENT'S NAME, PHONE NUMBER AND BEST TIME TO CALL BACK ON THE SCREENING CALL RECORD.)

(SKIP TO Q5)

INTRODUCTION #2 - NEW RESPONDENT

Hello. My name is _____ with Westat, a national research company near Washington, D.C. We are conducting a study for the National Institute for Occupational Safety and Health. The study is authorized by the Occupational Safety and Health Act of 1970.

The main study, called the National Occupational Hazard Survey, is a special research project designed to obtain basic information about health and safety practices within workplaces.

The purpose of this initial part of the study, however, is to verify some information about your establishment so we can be sure that establishments like yours are adequately represented in this study. Your responses are voluntary, however your cooperation is needed to make the results of this study comprehensive, accurate and timely. The questions should take about five minutes to answer. (Is this a convenient time to talk?) (IF YES, BEGIN INTERVIEW WITH Q5. IF NOT A CONVENIENT TIME, RECORD RESPONDENT'S NAME, PHONE NUMBER AND BEST TIME TO CALL BACK ON THE SCREENING CALL RECORD.)

5. Does your company have any other legal name(s)? (PROBE: Are you a subsidiary of a larger corporation?)

YES. 1 (Q6)
NO 2 (Q7)

6. What (is/are) the other name(s)?

_____

_____

_____

7. In what county and state is this establishment located? (RECORD VERBATIM AND CIRCLE CODE BELOW)

_____
COUNTY / STATE

IN TARGET PSU 1 (Q8)
NOT IN TARGET PSU. 2 (Q9)

8. Does your organization have any other establishments located in (TARGET PSU)?

YES. 1 (Q10)
NO 2 (Q11)

9. Does your organization have any establishments located in (TARGET PSU)?

YES 1 (Q10)
NO. . . . . 2 → Thank you very much but at this time we are only talking to organizations with establishments in (TARGET PSU). (TERMINATE)

10. How many establishments are there (including this one)?

OF ESTABLISHMENTS: _____

INTERVIEWER: ASK Qs. 11-19 FOR EACH ESTABLISHMENT	ESTABLISHMENT #1
11. What is the full name and local address of (your establishment/ ESTABLISHMENT #)?	NAME: _____ ADDRESS: _____ STREET # _____ CITY STATE ZIP
12. Approximately how many employees are employed at (this/that) establishment?	# OF EMPLOYEES: _____
13. What is the principal activity of (this/that) establishment? Would you say it is: (READ CARD A AND RECORD APPROPRIATE CATEGORY #)	CATEGORY #: _____
14. INTERVIEWER: SELECT THE CARD # WHICH CORRESPONDS TO THE CATEGORY IN Q13. I am going to read a second list. Please tell me which of the following categories corresponds to (your principal merchandise or product line/ the principal service rendered/the type of construction performed). (READ CARD)	CATEGORY #: _____

BOX B

INTERVIEWER: REVIEW Q10

Q10 = BLANK (ONLY ONE ESTABLISHMENT IN TARGET PSU) . . . 1 (Q16)
Q10 = NOT BLANK (MORE THAN ONE EST. IN TARGET PSU). . . . 2 (Q15)

15. Are all of your establishments in the same type of business?	YES. . . . . 1 (CODE Qs 13 AND 14 FOR EACH ESTABLISHMENT WITHOUT ASKING) NO . . . . . 2 (REASK Qs 13 AND 14 FOR EACH ESTABLISHMENT AS YOU GET TO IT)
16. Are there any unions at (your establishment/ESTABLISHMENT #)?	YES. . . . . 1 (Q17) NO . . . . . 2 (Q18) DON'T KNOW . . . 3 (Q18)
17. What is the exact name of (each of) the union(s)?	UNION #1 _____ UNION #2 _____ UNION #3 _____
18. Who would be the most knowledgeable person to contact at (your/that) establishment if additional information is necessary concerning unions?	CONTACT PERSON: _____ TITLE: _____ PHONE #: ( ) _____
19. (Am I correct that you would be/Who would be) the most knowledgeable person to contact at (your/that) establishment if additional information is necessary concerning your business?	CONTACT PERSON: _____ TITLE: _____ PHONE #: ( ) _____

ESTABLISHMENT #2	ESTABLISHMENT #3
NAME: _____ ADDRESS: _____ STREET # _____ CITY STATE ZIP	NAME: _____ ADDRESS: _____ STREET # _____ CITY STATE ZIP
# OF EMPLOYEES: _____	# OF EMPLOYEES: _____
CATEGORY #: _____	CATEGORY #: _____
CATEGORY #: _____	CATEGORY #: _____
YES. . . . . 1 (Q17) NO . . . . . 2 (Q18) DON'T KNOW . . . 3 (Q18)	YES. . . . . 1 (Q17) NO . . . . . 2 (Q18) DON'T KNOW . . . 3 (Q18)
UNION #1 _____ UNION #2 _____ UNION #3 _____	UNION #1 _____ UNION #2 _____ UNION #3 _____
CONTACT PERSON: _____ TITLE: _____ PHONE #: ( ) _____	CONTACT PERSON: _____ TITLE: _____ PHONE #: ( ) _____
CONTACT PERSON: _____ TITLE: _____ PHONE #: ( ) _____	CONTACT PERSON: _____ TITLE: _____ PHONE #: ( ) _____

Exhibit 11

September 2 1983

MEMORANDUM TO: D Ward

FROM : R Hanson

SUBJECT : Variances Among Plants within Size Class

This is to record the items for which variances are to be computed to check the assumption that relvariances among plants within size classes are fairly uniform.

These computations are to be made using the half sample replications defined for the purpose of computing within PSU variances. Second stage ratio factors of 1.0000 are to be used; that is, the weights used for establishments in the replicate are to be equal to the product of the field inflation weight and the first stage ratio factor only. Complete these computations by the end of September 1983.

1. For each of the 11 size classes determined by using the employees reported at interview and the total for all size classes combined, show:

For the total over all target SIC's:

Unweighted number of establishments in sample by the 11 size classes,
Estimated (weighted number of) establishments at time of interview and relvariance of the estimate within each of the 11 size classes,
Estimated employees at time of interview and relvariance of the estimate by 11 size classes.

Repeat the above for five of the more significant 2 digit target SIC's
(Count of uweighted number of establishments not required by 2 digit SIC)

Repeat for the total of all manufacturing SIC's in the target universe (1 digit SIC's 2 and 3)

2. The relvariances requested are WITHIN PSU relvariances. These relvariances should be computed by following the instructions in Memorandum III section III, paragraphs B, C, D, and E but omit the portion of paragraph E that asks for the difference between the within PSU variance and the total variance. For weights use

W * R1 * 1.0 (i.e. all values of R2 = 1.0).

3. The following table illustrates the results needed; the output format of VARNOES will be suitable but be sure the items are adequately identified. :

Estimates and relva	Size class at interview 11	1	2 9	10	Total all classes
Unweighted count of all establishments	x	x	x x	x	x
Total all target SIC's combined					
Estimated estabs	x	x	x x	x	x
Relvariance	x	x	x x	x	x
Estimated employees	x	x	x x	x	x
Relvariance	x	x	x x	x	x
2 DIGIT SIC A*					
Estimated estabs	x	x	x x	x	x
Relvariance	x	x	x x	x	x
Estimated employees	x	x	x x	x	x
Relvariance	x	x	x x	x	x
2 DIGIT SIC B*					
Estimated estabs	x	x	x x	x	x
Relvariance	x	x	x x	x	x
Estimated employees	x	x	x x	x	x
Relvariance	x	x	x x	x	x
:					
:					
2 DIGIT SIC E*					
Estimated estabs	x	x	x x	x	x
Relvariance	x	x	x x	x	x
Estimated employees	x	x	x x	x	x
Relvariance	x	x	x x	x	x
Total all manufacturing SIC's					
Estimated estabs	x	x	x x	x	x
Relvariance	x	x	x x	x	x
Estimated employees	x	x	x x	x	x
Relvariance	x	x	x x	x	x

* Use the results of tabulations requested in the September 1 memo to choose five of the more significant SIC's.

cc:
Waksberg
Edmonds
Escatell

Exhibit 12
September 3 1983

MEMORANDUM TO : D. Ward

FROM : R. Hanson

SUBJECT : Variances for NOES Estmates

This memorandum records the variances to be computed based on completed NOES questionnaires using the NOESVAR package. These variance computations are to provide a means for evaluating the accuracy of our estimation and variance computation packages and to provide the source for example variances to be included in the final NOES report.

All computer print outs produced are to be kept as permanent records; the results will be of use in evaluating the NOES and other similar designs with related objectives.

1. Computation of variances for unbiased estimates

The following computations are to have first priority as they will provide our first look at weighted estimates of NOES data. We will examine the results for evidence that the establishment weights we have assigned will produce reasonable results.

Compute total variances and standard errors for items that follow the format of the two tables listed below; see for example, Table 1 as described for the NOHS I (in Volume III page 42 of the NOHS I publication). The first table asks for estimates and variances for number of plants and total employees for three size classes of plants and for the total of all plant sizes; use employees reported at interview to determine size of plants:

Small (8 to 249 employees),
Medium (250 to 499 employees), and
Large (over 500 employees).

Tabulate at the US level for the total of all target SIC's and for the sums of target SIC's summarized to each 2-digit target SIC. Use the half sample replications associated with the total variance.

Prepare a second tabulation for the plants identified in the headings of NOHS I Table 36. I have modified the request to ask for injuries at all plants and separately for plants with unions to conform to the information available to us on the present Part I data from NIOSH.

These two tables relate to:

NOHS I Table	Description
1	National universe (about 38,300,000 employees in NOHS I)
36	Number of recorded occupational injuries in all plants and for the subset of plants which have one or more unions operating at the facility.

2. Variances of selected NOES publication estimates

The following computations are to have second priority. They call for the use of the final publication weight so that it will be necessary to complete the first and second stages of ratio estimation before preparing these tabulations. Use the half sample replications associated with the total variance.

The estimates and variances produced by these tabulations are to be hand posted into five tables (Tables 7-1 through 7-5), retyped, and included in the final report for the NOES as illustrations for variances expected for similar items by other users of the NOES data. The following tables describe the variance items desired:

NOES final report table	NOHS I table	Description
7-1	1	National universe
7-2	2	Plants receiving industrial hygiene services
7-3	10	Plants requiring replacement physical examinations
7-4	34	Total injuries in plants with safety engineers and in plants with both safety engineer and industrial hygiene services
7-5	36	Number of recorded occupational injuries in all plants and in plants which have one or more unions operating at the facility

3. Within PSU Variances

Define the half sample replications associated with within PSU variance computations, recompute the weights including both ratio factors for each of these replicates and compute variances for the items in NOHS I table 36 mentioned in paragraph 2 preceding; perform the same tabulations for additional tables discussed in paragraph 2 if there is time. The computations will not need to be posted to tables as these variances will not appear in the final report. VARNOES output from these runs will be sufficient; however, be sure that the items are adquately identified. These tabulations must be completed before September 30, 1983.

cc:
Waksberg
Edmonds
Escatell

Exhibit 13. Example of VARNOES (Continued)

```
 STMT NO. MESSAGE
-
    6      IEF653I SUBSTITUTION JCL - SYSOUT=A,HOLD=NO
    7      IEF653I SUBSTITUTION JCL - SYSOUT=A,HOLD=NO
   10      IEF653I SUBSTITUTION JCL - PGM=SAS,PARM='',REGION=384K
   18      IEF653I SUBSTITUTION JCL - SYSOUT=A,HOLD=NO,
   19      IEF653I SUBSTITUTION JCL - SYSOUT=A,HOLD=NO,
ICH70001I BIT      LAST ACCESS AT 15:42:04 ON FRIDAY, NOVEMBER 11, 1983
IEF236I ALLOC. FOR BIT663GO SASGEN GO
IEF237I 2C7  ALLOCATED TO STEPLIB
IEF237I 805  ALLOCATED TO SYS00402
IEF237I JES2 ALLOCATED TO SYSOUT
IEF237I JES2 ALLOCATED TO SYSDBOUT
IEF237I 880  ALLOCATED TO SAS1
IEF237I 2C5  ALLOCATED TO SPEC1
IEF237I 2C5  ALLOCATED TO SYS00404
IEF142I BIT663GO SASGEN GO - STEP WAS EXECUTED - COND CODE 0000
IEF285I   OSH.BIT.VARNOES.LOAD                          KEPT
IEF285I   VOL SER NOS= USER02.
IEF285I   SYSCTLG.VL1B007                               KEPT
IEF285I   VOL SER NOS= L1B007.
IEF285I   JES2.JOB00316.S00101                          SYSOUT
IEF285I   JES2.JOB00316.S00102                          SYSOUT
IEF285I   SYS83315.T154703.RU000.BIT663GO.SAS1          PASSED
IEF285I   VOL SER NOS= SCR880.
IEF285I   $BIT.OSH.WTJCLS.CNTL                          KEPT
IEF285I   VOL SER NOS= USRTS2.
IEF285I   SYSCTLG.VUSRTS2                               KEPT
IEF285I   VOL SER NOS= USRTS2.
IEF373I STEP /SASGEN  / START 83315.1547
IEF374I STEP /SASGEN  / STOP  83315.1547 CPU    0MIN 00.05SEC SRB    0MIN 00.00SEC VIRT    52K SYS   168K
********************************* PARKLAWN COMPUTER CENTER - SYSTEM  DV     OS/VS2 SP 1.3.0 REL 3.8 ***********************************
*                                                                                                                                       *
*     JOB  BIT663GO      DATE IN 83/315      STEP CPU TIME       .05 SEC.      REGION= USER     52   SYS    168   TOTAL    220     *
*                                                                                                                                       *
*     STEP SASGEN        TIME IN 15/47       CUMULATIVE CPU      .05 SEC.      COMPLETION - SYS 000  USER 0000  CONDITION 0000     *
*                                                                                                                                       *
*     I/O COUNTS BY UNIT    (2C7) =      1       (805) =      0       (880) =      2       (9C5) =      4       (9C5) =      0     *
*                                                                                                                                       *
*****************************************************************************************************************************************
IEF236I ALLOC. FOR BIT663GO SASGO GO
IEF237I 781  ALLOCATED TO LIBRARY
IEF237I 2CC  ALLOCATED TO STEPLIB
IEF237I 2C6  ALLOCATED TO
IEF237I 781  ALLOCATED TO
IEF237I 2C7  ALLOCATED TO
IEF237I 805  ALLOCATED TO SYS00408
IEF237I 2C4  ALLOCATED TO SYS00406
IEF237I 2CC  ALLOCATED TO SASHELP
IEF237I 880  ALLOCATED TO WORK
IEF237I JES2 ALLOCATED TO FT11F001
IEF237I JES2 ALLOCATED TO FT12F001
IEF237I JES2 ALLOCATED TO FT13F001
IEF237I DMY  ALLOCATED TO FT14F001
IEF237I 781  ALLOCATED TO FT15F001
IEF237I 2CC  ALLOCATED TO MAPS
IEF237I 2CD  ALLOCATED TO ZIP
IEF237I 880  ALLOCATED TO SYSIN
IEF237I 3E3  ALLOCATED TO DATA1
IEF237I JES2 ALLOCATED TO SYSOUT
IEF237I 781  ALLOCATED TO SASWORK3
```

Exhibit 13. Example of VARNOES

```
1     //BIT663GO JOB (OSH,BIT99),'ESC-SASPRCGO',                          JOB  316
      //              CLASS=S,MSGCLASS=A,NOTIFY=$BIT,TIME=(,30)          00000020
      ***PASS                                                             0000003
      ***ROUTE  PRINT RMT189                                              0000004
      ***TSO    OSH.WTJCLS.CNTL(SASPRCGO)                                00000050
      ***                                                                00000060
2     //GO       EXEC BITVAR                                             00000070
      ***                                                                00000010
3     XXBITVAR  PROC ENTRY=SAS,OPTIONS=,SYSOUT=A,HOLD=NO                 00000020
      ***                                                                00000030
4     XXSASGEN  EXEC PGM=SASGEN                                          00000040
5     XXSTEPLIB   DD  DSN=OSH.BIT.VARNOES.LOAD,DISP=SHR                  00000050
6     XXSYSOUT    DD  SYSOUT=&SYSOUT,HOLD=&HOLD                          00000060
7     XXSYSDBOUT  DD  SYSOUT=&SYSOUT,HOLD=&HOLD                          00000070
8     XXSAS1      DD  DSN=&&SAS1,DISP=(NEW,PASS),                        00000080
      XX              UNIT=SYSDA,SPACE=(80,(500,100)),                   00000090
      XX              DCB=(RECFM=FB,LRECL=80,BLKSIZE=3120)               00000100
      ***                                                                00000110
9     //SASGEN.SPEC1 DD  DSN=$BIT.OSH.WTJCLS.CNTL(SASPEC),DISP=SHR       00000080
10    XXSASGO   EXEC PGM=&ENTRY,PARM='&OPTIONS',REGION=384K              00000120
11    XXLIBRARY   DD  DSN=&LIBRARY,DISP=(MOD,PASS),                      00000130
      XX              UNIT=SYSDA,SPACE=(CYL,(1,,20))                     00000140
12    XXSTEPLIB   DD  DSN=PCC.SAS823.LIBRARY,DISP=SHR                    00000150
13    XX            DD  DSN=SYS1.PLIBASE,DISP=SHR                        00000160
14    XX            DD  DSN=*.LIBRARY,VOL=REF=*.LIBRARY,DISP=(OLD,PASS)  00000170
15    XX            DD  DSN=OSH.BIT.VARNOES.LOAD,DISP=SHR,DCB=BLKSIZE=19069 00000180
16    XXSASHELP   DD  DSN=PCC.SAS823.SASHELP,DISP=SHR                    00000190
17    XXWORK      DD  UNIT=SYSDA,SPACE=(6160,(500,200),,,ROUND)          00000200
18    XXFT11F001  DD  SYSOUT=&SYSOUT,HOLD=&HOLD,                         00000210
      XX              DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)              00000220
19    XXFT12F001  DD  SYSOUT=&SYSOUT,HOLD=&HOLD,                         00000230
      XX              DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)              00000240
20    XXFT13F001  DD  SYSOUT=B,DCB=(RECFM=F,BLKSIZE=80)                  00000250
21    XXFT14F001  DD  DUMMY                                              00000260
22    XXFT15F001  DD  UNIT=SYSDA,SPACE=(400,(100,300)),                  00000270
      XX              DCB=(RECFM=FB,LRECL=80,BLKSIZE=400,BUFNO=1)        00000280
23    XXMAPS      DD  DSN=PCC.SAS823.MAPS,DISP=SHR                       00000290
24    XXZIP       DD  DSN=PCC.SAS.ZIPCODES.COUNTIES,DISP=SHR             00000300
25    XXSYSIN     DD  DSN=&&SAS1,DISP=(OLD,DELETE)                       00000310
      ***                                                                00000320
26    //SASGO.DATA1  DD  DSN=OSH.BIT.WTSNDATA.DATA,DISP=SHR              00000090
```

Exhibit 13. Example of VARNOES (Continued)

```
IEF237I 781  ALLOCATED TO SASSWK02
IEF237I 880  ALLOCATED TO SASSWK01
IEF142I BIT663GO SASGO GO - STEP WAS EXECUTED - COND CODE 0000
IEF285I   SYS83315.T154703.RD000.BIT663GO.LIBRARY      PASSED
IEF285I   VOL SER NOS= SCR781.
IEF285I   PCC.SAS823.LIBRARY                           KEPT
IEF285I   VOL SER NOS= LIB004.
IEF285I   SYS1.PLIBASE                                 KEPT
IEF285I   VOL SER NOS= LIB003.
IEF285I   SYS83315.T154703.RD000.BIT663GO.LIBRARY      PASSED
IEF285I   VOL SER NOS= SCR781.
IEF285I   OSH.BIT.VARNOES.LOAD                         KEPT
IEF285I   VOL SER NOS= USER02.
IEF285I   SYSCTLG.VLIB007                              KEPT
IEF285I   VOL SER NOS= LIB007.
IEF285I   SYSCTLG.VLIB001                              KEPT
IEF285I   VOL SER NOS= LIB001.
IEF285I   PCC.SAS823.SASHELP                           KEPT
IEF285I   VOL SER NOS= LIB004.
IEF285I   SYS83315.T154703.RD000.BIT663GO.R0000001     DELETED
IEF285I   VOL SER NOS= SCR880.
IEF285I   JES2.JOB00316.S00103                         SYSOUT
IEF285I   JES2.JOB00316.S00104                         SYSOUT
IEF285I   JES2.JOB00316.S00105                         SYSOUT
IEF285I   SYS83315.T154703.RD000.BIT663GO.R0000002     DELETED
IEF285I   VOL SER NOS= SCR781.
IEF285I   PCC.SAS823.MAPS                              KEPT
IEF285I   VOL SER NOS= LIB004.
IEF285I   PCC.SAS.ZIPCODES.COUNTIES                    KEPT
IEF285I   VOL SER NOS= LIB002.
IEF285I   SYS83315.T154703.RD000.BIT663GO.SAS1         DELETED
IEF285I   VOL SER NOS= SCR880.
IEF285I   OSH.BIT.WTSNDATA.DATA                        KEPT
IEF285I   VOL SER NOS= USER03.
IEF285I   JES2.JOB00316.S00106                         SYSOUT
IEF285I   SYS83315.T154745.RA000.BIT663GO.R0000000     DELETED
IEF285I   VOL SER NOS= SCR781.
IEF285I   SYS83315.T154745.RA000.BIT663GO.R0000001     DELETED
IEF285I   VOL SER NOS= SCR781.
IEF285I   SYS83315.T154745.RA000.BIT663GO.R0000002     DELETED
IEF285I   VOL SER NOS= SCR880.
IEF373I STEP /SASGO   / START 83315.1547
IEF374I STEP /SASGO   / STOP  83315.1548 CPU    0MIN 17.89SEC SRB    0MIN 00.26SEC VIRT   384K SYS   220K
********************************* PARKLAWN COMPUTER CENTER - SYSTEM  DV     OS/VS2 SP 1.3.0 REL 3.8 ***********************************
*                                                                                                                                       *
*     JOB  BIT663GO      DATE IN 83/315      STEP CPU TIME     17.89 SEC.      REGION- USER    384   SYS    220   TOTAL    604     *
*                                                                                                                                       *
*     STEP SASGO         TIME IN 15/47      CUMULATIVE CPU    17.94 SEC.      COMPLETION - SYS 000  USER 0000  CONDITION 0000     *
*                                                                                                                                       *
*     I/O COUNTS BY UNIT    (781) =      0       (9CC) =     44       (9C6) =    27       (781) =      0       (2C7) =      4     *
*      (F05) =      0       (2C4) =      0       (9CC) =      0       (880) =   385       (781) =      0       (9CC) =      0     *
*      (9CD) =      0       (880) =      3       (EE3) =    206       (781) =    19       (781) =      0       (880) =     23     *
*                                                                                                                                       *
*****************************************************************************************************************************************
IEF237I 781  ALLOCATED TO SYS00004
IEF285I   SYS83315.T154800.RA000.BIT663GO.R0000004     KEPT
IEF285I   VOL SER NOS= SCR781.
IEF285I   SYS83315.T154703.RD000.BIT663GO.LIBRARY      DELETED
IEF285I   VOL SER NOS= SCR781.
IEF285I   SYS83315.T154703.RD000.BIT663GO.LIBRARY      KEPT
IEF285I   VOL SER NOS= SCR781.
IEF375I  JOB /BIT663GO/ START 83315.1547
IEF376I  JOB /BIT663GO/ STOP  83315.1548 CPU    0MIN 17.94SEC SRB    0MIN 00.26SEC
```

Exhibit 13. Example of VARNOES (Continued)

```
1       S A S   L O G    OS SAS 82.3        VS2/MVS JOB BIT663GO STEP SASGO    PROC GO                15:47 FRIDAY, NOVEMBER 11, 1983

NOTE: THE JOB BIT663GO HAS BEEN RUN UNDER RELEASE 82.3 OF SAS AT PARKLAWN COMPUTER CENTER (01233001).

NOTE: NO OPTIONS SPECIFIED.

1             OPTIONS S=72;
2             DATA   DATA1;
3             INFILE DATA1;
4             INPUT
5             @1   FACNUM      $CHAR6.                                          00000010
6             @8   WESTID      $CHAR14.                                         00000020
7             @23  W0           7.3                                             00000030
8             @31  (W1-W32)    (7.3)                                            00000040
9             @256 SIC2DGT      2.0                                             00000050
10            @259 EMPS         5.0                                             00000060
11            @265 INJWL        3.0                                             00000070
12            @269 INJNWL       3.0                                             00000080
13            @279 HYGSERV      1.0                                             00000090
14            @281 SAFSERV      1.0                                             00000100
15            ;
16            NUMPSU=99;
17            TOTINJ = INJWL + INJNWL ;        /* TOTAL INJURED */             00000110
18            IF   EMPS >= 50 ;                                                 00000120
19            IF   SIC2DGT < 15 THEN DELETE ;                                   00000130
20            IF   EMPS < 500 THEN SIZE = 1 ;                                   00000140
21            ELSE IF   EMPS < 2500 THEN SIZE = 2 ;                             00000150
22                 ELSE SIZE = 3 ;                                              00000160
23            IF   HYGSERV = 1 OR SAFSERV = 1                                   00000170
24            THEN DO ;                                                         00000180
25                 HYGSAFTY = 1 ;                                               00000190
26                 HYGSFEMP = EMPS ;                                            00000200
27                 HYGSFINJ = TOTINJ ;                                          00000210
28                 END ;                                                        00000220
29            ELSE DO ;                                                         00000230
30                 HYGSAFTY = 0 ;                                               00000240
31                 HYGSFEMP = 0 ;                                               00000250
32                 HYGSFINJ = 0 ;                                               00000260
33                 END ;                                                        00000270
34            PLANTS = 1 ;                     /* COUNT FACILITIES */          00000280
35            LABEL PLANTS = U.S. FIRMS                                         00000290
36                  EMPS = EMPLOYEES                                            00000300
37                  SIC2DGT = TWO DIGIT SIC CODE                                00000310
38                  HYGSFEMP = EMPS AT PLANTS WITH HYGIENE SERVICE              00000320
39                  HYGSFINJ = INJURED AT HYGIENE PLANTS                        00000330
40                  HYGSAFTY = PLANTS WITH HYGIENE OR SAFETY SERVICE            00000340
41                  TOTINJ = TOTAL INJURED                                      00000350
42                  W0 = PUBLICATION WEIGHT ;                                   00000360
43                                                                              00000370


NOTE: INFILE DATA1 IS:
      DSNAME=USH.BIT.WISNDATA.DATA,
      UNIT=DISK,VOL=SER=USER03,DISP=SHR,
      DCB=(BLKSIZE=6182,LRECL=281,RECFM=FB)

NOTE: MISSING VALUES WERE GENERATED AS A RESULT OF PERFORMING
      AN OPERATION ON MISSING VALUES.
      EAC     ACE     GIVE     : (     :R (     MES     (L:     (CO     ).
```

Exhibit 13. Example of VARNOES (Continued)

```
2      S A S   L O G    OS SAS 82.3        VS2/MVS JOB BIT663GO STEP SASGO    PROC GO              15:47 FRIDAY, NOVEMBER 11, 1983

      1289 AT 17:12

NOTE: 4490 LINES WERE READ FROM INFILE DATA1.
NOTE: DATA SET WORK.DATA1 HAS 2580 OBSERVATIONS AND 48 VARIABLES. 118 OBS/TRK.
NOTE: THE DATA STATEMENT USED 4.89 SECONDS AND 284K.

43           PROC NASSVAR DATA=DATA1 BEST TOTAL OUTPUT OUTDATA=REPS ;                00000370
44                VARIABLES EMPS HYGSFEMP HYGSFINJ HYGSAFTY PLANTS ;                00000380
45                WEIGHT    W0 W1-W32 ;                                             00000390
46                TITLE     'HYGIENE AND SAFETY; U.S. TOTALS' ;                     00000400
47                COMPVAR   HYGSAFTY PLANTS ;                                       00000410
48                OUTVAR    HSRATIO ;                                               00000420
49                HSRATIO = HYGSAFTY / PLANTS ;                                     00000430
50                                                                                  00000440

NOTE: NASSVAR IS AN UNSUPPORTED, EXPERIMENTAL PROCEDURE.
NOTE: DATA SET WORK.REPS HAS 33 OBSERVATIONS AND 7 VARIABLES. 782 OBS/TRK.

WESTAT INC
1650 RESEARCH BLVD
ROCKVILLE,MD 20852
(301) 251-1500

NOTE: THE PROCEDURE NASSVAR USED 11.07 SECONDS AND 284K AND PRINTED PAGE 1.

50           PROC PRINT DATA=REPS; TITLE REPLICATE LEVEL ESTIMATES ;                00000440
51                                                                                  00000450

NOTE: THE PROCEDURE PRINT USED 0.13 SECONDS AND 316K AND PRINTED PAGE 2.

51           PROC SORT DATA=DATA1 ;  BY SIZE ;                                      00000450
52                                                                                  00000460

NOTE: 3 CYLINDERS DYNAMICALLY ALLOCATED ON SYSDA FOR EACH OF 3 SORT WORK DATA SETS.
NOTE: DATA SET WORK.DATA1 HAS 2580 OBSERVATIONS AND 48 VARIABLES. 118 OBS/TRK.
NOTE: THE PROCEDURE SORT USED 0.98 SECONDS AND 368K.

52           PROC NASSTIM DATA=DATA1 BEST ;                                         00000460
53                VARIABLES EMPS PLANTS ;                                           00000470
54                BY        SIZE ;                                                  00000480
55                WEIGHT    W0 ;                                                    00000490
56                TITLE     TOTAL EMPLOYEES & PLANTS BY PLANT SIZE ;                00000500

NOTE: NASSTIM IS AN UNSUPPORTED, EXPERIMENTAL PROCEDURE.
NOTE:  OUTVAR/COMPVAR STATEMENT NOT USED OR INVALID.
NOTE: ABOVE MESSAGE FOR BY GROUP:
      SIZE=1

WESTAT INC
1650 RESEARCH BLVD
ROCKVILLE,MD 20850
(301) 251-1500

NOTE: THE PROCEDURE NASSTIM USED 0.61 SECONDS AND 324K AND PRINTED PAGES 3 TO 5.
```

Exhibit 13. Example of VARNOES (Continued)

HYGIENE AND SAFETY: U.S. TOTALS 15:47 FRIDAY, NOVEMBER 11, 1983 1
NASS SAMPLING ERRORS PROCEDURE

OPTIONS USED:
TOTAL
OUTPUT
2580 OBSERVATIONS PROCESSED
110777 WEIGHTED OBSERVATIONS PROCESSED
32 REPLICATES IN SAMPLE DESIGN

EMPS EMPLOYEES

ESTIMATE	MISSING	WEIGHTED MISSING	VARIANCE	REL-VARIANCE	CV(%)	STANDARD ERROR	LOWER 95% CONF INTVL	HIGHER 95% CONF INTVL
25703528	0	0	2.424E+11	.000366964	1.91563	492385	24738454	26668602

HYGSFEMP EMPS AT PLANTS WITH HYGIENE SERVICE

ESTIMATE	MISSING	WEIGHTED MISSING	VARIANCE	REL-VARIANCE	CV(%)	STANDARD ERROR	LOWER 95% CONF INTVL	HIGHER 95% CONF INTVL
21500784	0	0	3.962E+11	.000857137	2.92769	629476	20267010	22734558

HYGSFINJ INJURED AT HYGIENE PLANTS

ESTIMATE	MISSING	WEIGHTED MISSING	VARIANCE	REL-VARIANCE	CV(%)	STANDARD ERROR	LOWER 95% CONF INTVL	HIGHER 95% CONF INTVL
2105244	202	12196	6887006218	0.00155391	3.94197	82988	1942587	2267900

HYGSAFTY PLANTS WITH HYGIENE OR SAFETY SERVICE

ESTIMATE	MISSING	WEIGHTED MISSING	VARIANCE	REL-VARIANCE	CV(%)	STANDARD ERROR	LOWER 95% CONF INTVL	HIGHER 95% CONF INTVL
80040.1	0	0	4882292	.000762094	2.76061	2209.59	75709.3	84370.9

PLANTS U.S. FIRMS

ESTIMATE	MISSING	WEIGHTED MISSING	VARIANCE	REL-VARIANCE	CV(%)	STANDARD ERROR	LOWER 95% CONF INTVL	HIGHER 95% CONF INTVL
110777	0	0	1704373	.000138888	1.17851	1305.52	108218	113336

HSRATIO COMPUTED ESTIMATE

ESTIMATE	MISSING	WEIGHTED MISSING	VARIANCE	REL-VARIANCE	CV(%)	STANDARD ERROR	LOWER 95% CONF INTVL	HIGHER 95% CONF INTVL
0.722534	.	.	.000262379	.000502589	2.24185	0.0161981	0.690785	0.754282

Exhibit 13. Example of VARNOES (Continued)

REPLICATE LEVEL ESTIMATES 15:47 FRIDAY, NOVEMBER 11, 1983 2

OBS	REPL_ID	EMPS	HYGSFEMP	HYGSFINJ	HYGSAFTY	PLANTS	HSRATIO
1	0	25703528	21500784	2105244	80040.1	110777	0.722534
2	1	25999823	21818049	2148270	80913.2	109075	0.741815
3	2	24766319	20512640	2139621	79683.4	109132	0.730154
4	3	26225399	22071736	2294209	81474.4	111487	0.730795
5	4	26058574	21926919	2095054	76505.6	107407	0.712298
6	5	26179318	22522373	2268604	82635.7	112007	0.737771
7	6	25733954	21837573	2183430	82325.5	111819	0.736240
8	7	26065675	22088311	2102924	86286.8	113105	0.762895
9	8	25396010	21087541	2191742	80248.2	111744	0.718147
10	9	25768914	21469404	2049670	76132.9	109909	0.692690
11	10	26402056	22274300	2073061	82030.7	111140	0.738083
12	11	25636949	21703891	2131162	80292.1	111570	0.719660
13	12	26031640	21637285	2107410	78008.7	111604	0.698975
14	13	26274174	22124088	2184239	81208.8	111958	0.725353
15	14	25852316	21498848	2120536	82316.6	110491	0.745008
16	15	25562576	21331742	2098722	80558.7	112326	0.717185
17	16	26817975	22733604	2123998	82769.7	111714	0.740906
18	17	25560882	21285314	2125907	78290.6	109371	0.715825
19	18	25834250	21990181	2206905	80984.1	108767	0.744562
20	19	25971051	21958392	2208492	81025.5	109994	0.736635
21	20	26139872	22518263	2232832	83195.2	111936	0.743238
22	21	26192607	22430995	2247548	82554.9	111031	0.743532
23	22	25853584	22199357	2251844	82060.4	111292	0.737341
24	23	25633267	21233287	2104971	80427.8	109299	0.735854
25	24	26209973	22224162	2086309	80865.0	112103	0.721343
26	25	26794408	22524818	2128418	81971.6	111955	0.732183
27	26	25682380	21909043	2179792	79964.0	111004	0.720370
28	27	26371500	22421857	2150208	81335.6	110437	0.736491
29	28	25949844	21520785	1967583	80540.9	109836	0.733282
30	29	26127583	22007827	2097859	81544.5	111345	0.732362
31	30	26376055	21884406	2077571	79594.0	111179	0.715910
32	31	26154352	22431102	2220015	83830.8	112258	0.746772
33	32	26312975	21616086	2011853	78177.5	108825	0.718377

Exhibit 13. Example of VARNOES (Continued)

SIZE=1
NASS ESTIMATION PROCEDURE

OPTIONS USED:
BEST

1914 OBSERVATIONS PROCESSED
101398 WEIGHTED OBSERVATIONS PROCESSED

NAME	LABEL	ESTIMATE	MISSING	WEIGHTED MISSING
EMPS	EMPLOYEES	13274209	0	0
PLANTS	U.S. FIRMS	101398	0	0

Exhibit 13. Example of VARNOES (Continued)

SIZE=2
NASS ESTIMATION PROCEDURE

OPTIONS USED:
BEST

488 OBSERVATIONS PROCESSED
8432 WEIGHTED OBSERVATIONS PROCESSED

NAME	LABEL	ESTIMATE	MISSING	WEIGHTED MISSING
EMPS	EMPLOYEES	8016071	0	0
PLANTS	U.S. FIRMS	8432.06	0	0

Exhibit 13. Example of VARNOES (Continued)

TOTAL EMPLOYEES & PLANTS BY PLANT SIZE 15:47 FRIDAY, NOVEMBER 11, 1983 5
SIZE=3
NASS ESTIMATION PROCEDURE

OPTIONS USED:
BEST

178 OBSERVATIONS PROCESSED
947 WEIGHTED OBSERVATIONS PROCESSED

NAME	LABEL	ESTIMATE	MISSING	WEIGHTED MISSING
EMPS	EMPLOYEES	4413248	0	0
PLANTS	U.S. FIRMS	947.197	0	0

Exhibit 14

NIOSH NOHS II
STANDARD INDUSTRIAL CLASSIFICATION
VERIFICATION MANUAL

CARD 2

CATEGORY #	CATEGORY NAME	SIC RANGE
1	RETAIL TRADE	5200 - 5999
2	WHOLESALE TRADE	5000 - 5199
3	MANUFACTURING	2000 - 3999
4	AGRICULTURE, FORESTRY, FISHING	0000 - 0999
5	OIL AND GAS EXTRACTION	1300 - 1389
6	CONSTRUCTION, OR SPECIAL TRADE CONTRACTOR	1500 - 1999
7	TRANSPORTATION, COMMUNICATIONS, ELECTRIC, GAS OR SANITARY SERVICES	4000 - 4999
8	FINANCE, INSURANCE, REAL ESTATE OR BANKING	6000 - 6799
9	SPECIALIZED SERVICES	7000 - 8999
10	OTHER BUSINESS/INDUSTRY ACTIVITY	9999

CARD 1

RETAIL TRADE (5200 - 5999)

CATEGORY =	CATEGORY NAME	SIC RANGE
1	LUMBER AND OTHER BUILDING MATERIALS (lumber, paint, hardware, garden supplies, mobile home dealer, etc.)	5210 - 5299*
2	GENERAL MERCHANDISE STORES (department stores, variety stores, general stores, catalog stores, etc.)	5310 - 5399*
3	FOOD STORES (grocery, bakery, fruit stores, etc.)	5410 - 5499*
4	AUTOMOTIVE DEALERS AND SERVICE STATIONS (motor cycles, recreation vehicles, etc.)	5510 - 5599

CATEGORY	SIC
Motor Vehicle Dealers (Used Only).	5521
Auto and Home Supply Stores.	5531
Gasoline and Service Stations.	5541

5	APPAREL AND ACCESSORY STORES (men's, women's, and children's clothing, shoe stores, etc.)	5610 - 5699*
6	FURNITURE, HOME FURNISHINGS AND APPLIANCE STORES (furniture, radio and TV, floor coverings, etc.)	5710 - 5733*
7	EATING AND DRINKING PLACES (vending, carry-outs, etc.)	5810 - 5813*
8	OTHER RETAIL (drug stores, liquor stores, jewelry stores, stationery stores, etc.)	5910 - 5999*

*These SIC codes are ineligible. Code case "I" and retire.

CARD 2

WHOLESALE TRADE (5000 - 5199)

CATEGORY #	CATEGORY NAME	SIC RANGE
1	WHOLESALE TRADE, DURABLE GOODS (motor vehicles and parts, construction materials and supplies, furniture, electric appliances, etc.	5010 - 5099

CATEGORY	SIC
Automobiles and other motor vehicles.	5012
Automotive parts and supplies	5013
Tires and tubes	5014
Lumber, plywood and millwork.	5031
Construction materials, not elsewhere classified. .	5039
Metal service centers and offices	5051
Coal and other minerals and ores.	5052
Scrap and waste materials	5093

2	WHOLESALE TRADE, NONDURABLE GOODS (paper products, food and beverage products, drugs, apparel, etc.)	5110 - 5199

CATEGORY	SIC
Chemicals and allied products	5161
Petroleum bulk stations and terminals	5171
Petroleum and petroleum products wholesalers, except bulk stations and terminals.	5172

MANUFACTURING (2000 - 3999

CATEGORY #	CATEGORY NAME	SIC RANGE
1	FOOD AND KINDRED PRODUCTS (meat, fruit, grain mill products, etc.)	2010 - 2099

CATEGORY	SIC
Meat packing plants	2011
Sausages and other prepared meat products	2013
Poultry and egg processing	2017
Creamery butter	2021
Cheese, natural and processed	2022
Condensed and evaporated milk	2023
Icecream and frozen desserts	2024
Fluid milk	2026
Canned specialties	2032
Canned fruits, vegetables, preserves, jams and jellies	2033
Dried and dehydrated fruits, vegetables, and soup mixes	2034
Pickled fruits and vegetables, vegetable sauces and seasonings, and salad dressings	2035
Frozen fruits, fruit juices and vegetables	2037
Frozen specialties	2038
Flour and other grain mill products	2041
Cereal breakfast foods	2043
Rice milling	2044
Blended and prepared flour	2045
Wet corn milling	2046
Dog, cat and other pet food (2D42)	2047
Prepared feeds and feed ingredients for animals and fowls, not elsewhere classified (2D42)	2048
Bread and other bakery products, except cookies and crackers	2051
Cookies and crackers	2052
Cane sugar, except refining only	2061
Cane sugar refining	2062
Beet sugar	2063
Candy and other confectionery products	2065
Chocolate and cocoa products	2066
Chewing gum	2067
Cottonseed oil mills	2074
Soybean oil mills	2075
Vegetable oil mills, except corn, cottonseed, and soybean	2076
Animal and marine fats and oils	2077
Shortening, table oils, margarine and other edible fats and oils, not elsewhere classified	2079

CARD 3 (cont.

MANUFACTURING (2000 - 3999.

CATEGORY =	CATEGORY NAME	SIC RANGE
1	FOOD AND KINDRED PRODUCTS (meat, fruit, grain mill products, etc.) cont.	2010 - 2099

CATEGORY	SIC
Malt Beverages	2082
MALT	2083
Wines, brandy, and brandy spirits	2084
Distilled, rectified, and blended liquors	2085
Bottled and canned soft drinks and carbonated waters	2086
Flavoring extracts and flavoring syrups, not elsewhere classified	2087
Canned and cured fish and seafoods	2091
Fresh or frozen packaged fish and seafoods	2092
Roasted coffee	2095
Manufactured ice	2097
Macaroni, spaghetti, vermicelli, and noodles	2098
Food preparations, not elsewhere classified	2099

2	TOBACCO MANUFACTURERS (cigarettes, cigars, etc.)	2110 - 2141

CATEGORY	SIC
Cigarettes	2111
Cigars	2121
Tobacco (chewing and smoking) and snuff	2131
Tobacco stemming and redrying	2141

MANUFACTURING (2000 - 3999)

CATEGORY #	CATEGORY NAME	SIC RANGE
3	TEXTILE MILL PRODUCTS (weaving mills, knitting mills, yarn mills, carpets and rugs, etc.)	2211 - 2299

CATEGORY	SIC
Broad woven fabric mills, cotton	2211
Broad woven fabric mills, man-made fiber and silk	2221
Broad woven fabric mills, wool (including dyeing and finishing)	2231
Narrow fabrics and other smallwares mills: cotton, wool, silk, and man-made fiber	2241
Women's full length and knee-length hosiery	2251
Hosiery, except women's full length and knee-length hosiery	2252
Knit outerwear mills	2253
Knit underwear mills	2254
Circular knit fabric mills	2257
Warp knit fabric mills	2258
Knitting mills, not elsewhere classified	2259
Finishers of broad woven fabrics of cotton	2261
Finishers of broad woven fabrics of man-made fiber and silk	2262
Finishers of textiles, not elsewhere classified	2269
Woven carpets and rugs	2271
Tufted carpets and rugs	2272
Carpets and rugs, not elsewhere classified	2279
Yarn spinning mills: cotton, man-made fibers and silk	2281
Yarn texturizing, throwing, twisting and winding mills cotton, man-made fibers and silk	2282
Yarn mills, wool, including carpet and rug yarn	2283
Thread mills	2284
Felt goods, except woven felts and hats	2291
Lace goods	2292
Paddings and upholstery filling	2293
Processed waste and recovered fibers and flock	2294
Coated fabrics, not rubberized	2295
Tire cord and fabric	2296
Nonwoven fabrics	2297
Cordage and twine	2298
Textile goods, not elsewhere classified	2299

CARD 3 (cont.

MANUFACTURING (2000 - 3999)

CATEGORY #	CATEGORY NAME	SIC RANGE
4	APPAREL AND OTHER TEXTILE PRODUCTS (men's, women's outerwear, and accessories, home furnishers, etc.)	2310 - 2399

CATEGORY	SIC
Men's, youths', and boys' suits, coats and overcoats	2311
Men's, youths', and boys' shirts (except work shirts) and nightwear	2321
Men's, youths', and boys' underwear	2322
Men's, youths', and boys' neckwear	2323
Men's, youths', and boys' separate trousers	2327
Men's, youths', and boys' work clothing	2328
Men's, youths', and boys' clothing, not elsewhere classified	2329
Women's, misses', and juniors' blouses, waists, and shirts	2331
Women's, misses', and juniors' dresses	2335
Women's, misses', and juniors' suits, skirts, and coats	2337
Women's, misses', and juniors' outerwear, not elsewhere classified	2339
Women's misses', children's and infants' underwear and nightwear	2341
Brassieres, girdles, and allied garments	2342
Millinery	2351
Hats and caps, except millinery	2352
Girls', children's, and infants' dresses, blouses, waists, and shirts	2361
Girls', children's, and infants' coats and suits	2363
Girls', children's, and infants' outerwear, not elsewhere classified	2369
Fur goods	2371
Dress and work gloves, except knit and all leather	2381
Robes and dressing gowns	2384
Raincoats and other waterproof outer garments	2385
Leather and sheep lined clothing	2386
Apparel belts	2387
Apparel and accessories, not elsewhere classified	2389
Curtains and draperies	2391
House furnishings, except curtains and draperies	2392
Textile bags	2393
Canvas and related products	2394
Pleating, decorative and novelty stitching, and tucking for the trade	2395
Automotive trimmings, apparel findings, and related products	2396
Schiffli machine embroideries	2397
Fabricated textile products, not elsewhere classified	2399

CARD 3 (cont.

MANUFACTURING (2000 - 3999

CATEGORY #	CATEGORY NAME	SIC RANGE
5	LUMBER AND WOOD PRODUCTS (sawmills, millwork, wood containers, etc.)	2410 - 2499

CATEGORY	SIC
Logging camps and logging contractors	2411
Sawmills and planning mills, general	2421
Hardwood dimension and flooring mills	2426
Special product sawmills, not elsewhere classified	2429
Millwork	2431
Wood kitchen cabinets	2434
Hardwood veneer and plywood	2435
Softwood veneer and plywood	2436
Structural wood members, not elsewhere classified	2439
Nailed and lock corner wood boxes and shook	2441
Wood pallets and skids	2448
Wood containers, not elsewhere classified	2449
Mobile homes	2451
Prefabricated wood buildings and components (2433)	2452
Wood preserving	2491
Particleboard	2492
Wood products, not elsewhere classified	2499

6	FURNITURE AND FIXTURES (household, office furniture, partitions, etc.)	2510 - 2599

CATEGORY	SIC
Wood household furniture, except upholstered	2511
Wood household furniture, upholstered	2512
Metal household furniture	2514
Mattresses and bedsprings	2515
Wood television, radio, phonograph, and sewing machine cabinets	2517
Household furniture, not elsewhere classified	2519
Wood office furniture	2521
Metal office furniture	2522
Public building and related furniture	2531
Wood partitions, shelving, lockers, and office and store fixtures	2541
Drapery hardware and window blinds and shades	2591
Furniture and fixtures, not elsewhere classified	2599

CARD 3 (cont.

MANUFACTURING (2000 - 3999

CATEGORY #	CATEGORY NAME	SIC RANGE
7	PAPER AND ALLIED PRODUCTS (paper, pulp mills, paperboard boxes, etc.)	2610 - 2661

CATEGORY	SIC
Pulp mills	2611
Paper mills, except building paper mills	2621
Paperboard mills	2631
Paper coating and glazing	2641
Envelopes	2642
Bags, except textile bags	2643
Die-cut paper and paperboard and cardboard	2645
Pressed and molded pulp goods	2646
Sanitary paper products	2647
Stationery, tablets and related products	2648
Converted paper and paperboard products, not elsewhere classified	2649
Folding paperboard boxes	2651
Set-up paperboard boxes	2652
Corrugated and solid fiber boxes	2653
Sanitary food containers	2654
Fiber cans, tubes, drums, and similar products	2655
Building paper and building board mills	2661

8	PRINTING AND PUBLISHING (newspapers, books, periodicals, greeting cards, office forms, etc.)	2710 - 2795

CATEGORY	SIC
Newspapers: publishing, publishing and printing	2711
Periodicals: publishing, publishing and printing	2721
Book printing	2732
Miscellaneous publishing	2741
Commercial printing, letterpress and screen	2751
Commercial printing, lithograph	2752
Engraving and plate printing	2753
Commercial printing, gravure	2754
Manifold business forms	2761
Greeting card publishing	2771
Blankbooks, looseleaf binders and devices	2782
Bookbinding and related work	2789
Typesetting	2791
Photoengraving	2793
Electrotyping and stereotyping	2794
Lithographic platemaking and related services	2795

CARD 3 (cont.)

MANUFACTURING (2000 - 3999)

CATEGORY =	CATEGORY NAME	SIC RANGE
9	CHEMICALS AND ALLIED PRODUCTS (drugs, toiletries, paints, plastics and synthetics, etc.)	2810 - 2899

CATEGORY	SIC
Alkalies and chlorine	2812
Industrial gases	2813
Inorganic pigments	2816
Industrial inorganic chemicals, not elsewhere classified	2819
Plastics materials, synthetic resins, and nonvulcanizable elastomers	2821
Synthetic rubber (vulcanizable elastomers)	2822
Cellulosic man-made fibers	2823
Synthetic organic fibers, except cellulosic	2824
Biologic products	2831
Medicinal chemicals and botanical products	2833
Pharmaceutical preparations	2834
Soap and other detergents, except specialty cleaners	2841
Specialty cleaning, polishing, and sanitation preparations	2842
Surface active agents, finishing agents, sulfonated oils and assistants	2843
Perfumes cosmetics, and other toilet preparations	2844
Paints, varnishes, lacquers, enamels, and allied products	2851
Gum and wood chemicals	2861
Cyclic (coal tar) crudes, and cyclic intermediates, dyes, and organic pigments (lakes and toners)	2865
Industrial organic chemicals, not elsewhere classified (2818)	2869
Nitrogenous fertilizers (2871)	2873
Phosphatic fertilizers	2874
Fertilizers, mixing only	2875
Pesticides and agricultural chemicals, not elsewhere classified	2879
Adhesives and sealants	2891
Explosives	2892
Printing ink	2893
Carbon black	2895
Chemicals and chemical preparations, not elsewhere classified	2899

MANUFACTURING (2000 - 3999

CATEGORY #	CATEGORY NAME	SIC RANGE
10	PETROLEUM AND COAL PRODUCTS (petroleum refining, lubricating oils, etc.)	2910 - 2999

CATEGORY	SIC
Petroleum refining	2911
Paving mixtures and blocks	2951
Asphalt felts and coatings	2952
Lubricating oils and greases	2992
Products of petroleum and coal, not elsewhere classified	2999

CATEGORY #	CATEGORY NAME	SIC RANGE
11	RUBBER AND MISCELLANEOUS PRODUCTS (tires, rubber base products, etc.)	3010 - 3079

CATEGORY	SIC
Tires and inner tubes	3011
Rubber and plastics footwear	3021
Reclaimed rubber	3031
Rubber and plastics hose and belting	3041
Fabricated rubber products, not elsewhere classified	3069
Miscellaneous plastics products	3079

CATEGORY #	CATEGORY NAME	SIC RANGE
12	LEATHER AND LEATHER PRODUCTS (leather footwear, luggage, etc.)	3110 - 3199

CATEGORY	SIC
Leather tanning and finishing	3111
Boot and shoe cut stock and findings	3131
House slippers	3142
Men's footwear, except athletic (3141)	3143
Women's footwear, except athletic (3141)	3144
Footwear, except rubber, not elsewhere classified (3141)	3149
Leather gloves and mittens	3151
Luggage	3161
Women's handbags and purses	3171
Personal leather goods, except women's handbags and purses	3172
Leather goods, not elsewhere classified	3199

MANUFACTURING (2000 - 3999)

CATEGORY =	CATEGORY NAME	SIC RANGE
13	STONE, CLAY AND GLASS PRODUCTS (glass, cement, pottery, abrasives, etc.)	3210 - 3299

CATEGORY	SIC
Flat glass	3211
Glass containers	3221
Pressed and blown glass and glassware, not elsewhere classified	3229
Glass products, made of purchased glass	3231
Cement, hydraulic	3241
Brick and structural clay tile	3251
Ceramic wall and floor tile	3253
Clay refractories	3255
Structural clay products, not elsewhere classified	3259
Vitreous china plumbing fixtures and china and earthernware fittings and bathroom accessories	3261
Vitreous china table and kitchen articles	3262
Fine earthenware (whiteware) table and kitchen articles	3263
Porcelain electrical supplies	3264
Pottery products, not elsewhere classified	3269
Concrete block and brick	3271
Concrete products, except block and brick	3272
Ready-mixed concrete	3273
Lime	3274
Gypsum products	3275
Cut stone and stone products	3281
Abrasive products	3291
Asbestos products	3292
Gaskets, packing, and sealing devices	3293
Minerals and earths, ground or otherwise treated	3295
Mineral wool	3296
Nonclay refractories	3297
Nonmetallic mineral products, not elsewhere classified	3299

MANUFACTURING (2000 - 3999

CATEGORY #	CATEGORY NAME	SIC RANGE
14	PRIMARY METAL INDUSTRIES (steel mills, foundries, primary nonferrous metals, etc.)	3312 - 3399

CATEGORY	SIC
Blast furnaces (including coke ovens), steel works, and rolling mills	3312
Electrometallurgical products	3313
Steel wire drawing and steel nails and spikes	3315
Cold rolled steel sheet, strip and bars	3316
Steel pipe and tubes	3317
Gray iron foundries	3321
Malleable iron foundries	3322
Steel investment foundries	3324
Steel foundries, not elsewhere classified	3325
Primary smelting and refining of copper	3331
Primary smelting and refining of lead	3332
Primary smelting and refining of zinc	3333
Primary production of aluminum	3334
Primary smelting and refining of nonferrous metals, not elsewhere classified	3339
Secondary smelting and refining of nonferrous metals	3341
Rolling, drawing, and extruding of copper	3351
Aluminum sheet, plate, and foil	3353
Aluminum extruded products	3354
Aluminum rolling and drawing, not elsewhere classified	3355
Rolling, drawing, and extruding of nonferrous metals, except copper and aluminum	3356
Drawing and insulating of nonferrous wire	3357
Aluminum foundries (castings)	3361
Brass, bronze, copper, copper base alloy foundries (castings)	3362
Nonferrous foundries (castings), not elsewhere classified	3369
Metal heat treating	3398
Primary metal products, not elsewhere classified	3399

CARD 3 CONT.

MANUFACTURING (2000 - 3999)

CATEGORY #	CATEGORY NAME	SIC RANGE
15	FABRICATED METAL PRODUCTS (metal cans, cutlery, structural metal work, hardware, etc.)	3410 - 3499

CATEGORY	SIC
Metal cans	3411
Metal shipping barrels, drums, kegs, and pails	3412
Cutlery	3421
Hand and edge tools, except machine tools and hand saws	3423
Hand saws and saw blades	3425
Hardware, not elsewhere classified	3429
Enameled iron and metal sanitary ware	3431
Plumbing fixture fittings and trim (brass goods)	3432
Heating equipment, except electric and warm air furnaces	3433
Fabricated structural metal	3441
Metal doors, sash, frames, molding and trim	3442
Fabricated plate work (boiler shops)	3443
Sheet metal work	3444
Architectural and ornamental metal work	3446
Prefabricated metal buildings and components	3448
Miscellaneous metal work	3449
Screw machine products	3451
Bolts, nuts, screws, rivets, and washers	3452
Iron and steel forgings	3462
Nonferrous forgings	3463
Automotive stampings (3461)	3465
Crowns and closures (3461)	3466
Metal stampings, not elsewhere classified (3461)	3469
Electroplating, plating, polishing, anodizing and coloring	3471
Coating, engraving and allied services, not elsewhere classified	3479
Small arms ammunition	3482
Ammunition, except for small arms, not elsewhere classified	3483
Small arms	3484
Ordnance and accessories, not elsewhere classified	3489
Steel springs, except wire	3493
Valves and pope fittings, except plumbers' brass goods	3494
Wire springs	3495
Miscellaneous fabricated wire products	3496
Metal foil and leaf	3497
Fabricated pipe and fabricated pipe fittings	3498
Fabricated metal products, not elsewhere classified	3499

CARD 3 (cont.)

MANUFACTURING 2000 - 3999

CATEGORY #	CATEGORY NAME	SIC RANGE
16	MACHINERY, EXCEPT ELECTRICAL (engines, farm and industrial machinery metal work machinery, etc.)	3510 - 3599

CATEGORY	SIC
Steam, gas, and hydraulic turbines and turbine generator set units	3511
Internal combustion engines, not elsewhere classified	3519
Farm machinery and equipment (3522)	3523
Garden tractors and lawn and garden equipment (3522)	3524
Construction machinery and equipment	3531
Mining machinery and equipment, except oil field machinery and equipment	3532
Oil field machinery and equipment	3533
Elevators and moving stairways	3534
Conveyors and conveying equipment	3535
Hoists, industrial cranes, and monorail systems	3536
Industrial trucks, tractors, trailers, and stackers	3537
Machine tools, metal cutting types	3541
Machine tools, metal forming types	3542
Special dies and tools, die sets, jigs and fixtures, and industrial molds	3544
Machine tool accessories and measuring devices	3545
Power driven hand tools (3548)	3546
Rolling mill machinery and equipment (3548)	3547
Metalworking machinery, not elsewhere classified (3548)	3549
Food products machinery	3551
Textile machinery	3552
Woodworking machinery	3553
Paper industries machinery	3554
Printing trades machinery and equipment	3555
Special industry machinery, not elsewhere classified	3559
Pumps and pumping equipment	3561
Ball and roller bearings	3562
Air and gas compressors	3563
Blowers and exhaust and ventilation fans	3564
Industrial patterns	3565
Speed changer, industrial high speed drives, and gears	3566
Industrial process furnaces and ovens	3567
Mechanical power transmission equipment, not elsewhere classified	3568
	3569
Typewriters	3572
Electronic computing equipment	3573

MANUFACTURING (2000 - 3999)

CATEGORY #	CATEGORY NAME	SIC RANGE

16 MACHINERY, EXCEPT ELECTRICAL (engines, farm and industrial machinery, metal work machinery, etc.) (cont.) — 3510 - 3599

CATEGORY	SIC
Calculating and accounting machines, except electronic computing equipment	3574
Scales and balances, except laboratory	3576
Office machines, not elsewhere classified	3579
Automatic merchandising machines	3581
Commercial laundry, dry cleaning, and pressing machines	3582
Air conditioning and warm air heating equipment and commercial and industrial refrigeration equipment	3585
Measuring and dispensing pumps	3586
Service industry machines, not elsewhere classified	3589
Carburetors, pistons, piston rings, and valves	3592
Machinery, except electrical, not elsewhere classified	3599

17 ELECTRIC AND ELECTRONIC EQUIPMENT (electrical industrial apparatus, household appliances, etc.) — 3610 - 3699

CATEGORY	SIC
Power, distribution, and specialty transformers (3611)	3612
Switchgear and switchboard apparatus	3613
Motors and generators	3621
Industrial controls	3622
Welding apparatus, electric	3623
Carbon and graphite products	3624
Electrical industrial apparatus, not elsewhere classified	3629
Household cooking equipment	3631
Household refrigerators and home and farm freezers	3632
Household laundry equipment	3633
Electric housewares and fans	3634
Household vacuum cleaners	3635
Sewing machines	3636
Household appliances, not elsewhere classified	3639
Electric lamps	3641
Current-carrying wiring devices	3643
Noncurrent-carrying wiring devices	3644
Residential electric lighting fixtures	3645
Commercial, industrial, and institutional electric lighting fixtures	3646
Vehicular lighting equipment	3647

CATEGORY #	CATEGORY NAME	SIC RANGE
17	ELECTRIC AND ELECTRONIC EQUIPMENT (electrical industrial apparatus, household appliances, etc.) (cont.)	3610 - 3699

CATEGORY	SIC
Lighting equipment, not elsewhere classified	3648
Radio and television receiving sets, except communication types	3651
Phonograph records and pre-recorded magnetic tape	3652
Telephone and telegraph apparatus	3661
Radio and television transmitting, signaling, and detection equipment and apparatus	3662
Radio and ~~television~~ receiving type electron tubes except cathode ray	3671
Cathode ray television picture tubes	3672
Transmitting, industrial, and special purpose electron tubes	3673
Semiconductors and related devices	3674
Electronic capacitors	3675
Resistors, for electronic applications	3676
Electronic coils, transformers and other inductors	3677
Connectors, for electronic applications	3678
Electronic components, not elsewhere classified	3679
Storage batteries	3691
Primary batteries, dry and wet	3692
Radiographic X-ray, fluoroscopic X-ray, therapeutic X-ray, and other X-ray apparatus and tubes; electromedical and electrotherapeutic apparatus	3693
Electrical equipment for internal combustion engines	3694
Electrical machinery, equipment, and supplies, not elsewhere classified	3699

CARD 3 (cont.

MANUFACTURING (2000 - 3999

CATEGORY #	CATEGORY NAME	SIC RANGE
18	TRANSPORTATION EQUIPMENT (motor vehicles, aircraft, ships, also parts, etc.)	3710 - 3799

CATEGORY	SIC
Motor vehicles and passenger car bodies	3711
Truck and bus bodies	3713
Motor vehicle parts and accessories	3714
Truck trailers	3715
Aircraft	3721
Aircraft engines and engine parts (3722)	3724
Aircraft parts and auxiliary equipment, not elsewhere classified	3728
Ship building and repairing	3731
Boat building and repairing	3732
Railroad equipment	3743
Motorcycles, bicycles, and parts	3751
Guided missiles and space vehicles	3761
Guided missile and space vehicle propulsion units and propulsion unit parts	3764
Guided missile and space vehicle parts and auxiliary equipment, not elsewhere classified	3769
Travel trailers and campers	3792
Tanks and tank components	3795
Transportation equipment, not elsewhere classified	3799

MANUFACTURING (2000 - 3999)

CATEGORY =	CATEGORY NAME	SIC RANGE
19	INSTRUMENTS AND RELATED PRODUCTS (optical, medical and scientific instruments, watches and clocks, etc.)	3810 - 3873

CATEGORY	SIC
Engineering, laboratory, scientific, and research instruments and associated equipment	3811
Mechanical measuring instruments *(SIC code not listed in manual)	*3821
Automatic controls for regulating residential and commercial environments and appliances	3822
Industrial instruments for measurement, display and control of process variables; and related products	3823
Totalizing fluid meters and counting devices	3824
Instruments for measuring and testing of electricity and electrical signals	3825
Measuring and controlling devices, not elsewhere classified	3829
Optical instruments and lenses	3832
Surgical and medical instruments and apparatus	3841
Orthopedic, prosthetic, and surgical appliances and supplies	3842
Dental equipment and supplies	3843
Opthalmic goods	3851
Photographic equipment and supplies	3861
Watches, clocks, clockwork operated devices, and parts	3873

CARD 3 (cont.

MANUFACTURING (2000 - 3999

CATEGORY =	CATEGORY NAME	SIC RANGE
20	OTHER MANUFACTURING INDUSTRIES (jewelry, musical instruments, pens, etc.)	3910 - 3999

CATEGORY	SIC
Jewelry, precious metal	3911
Silverware, plated ware, and stainless steel ware	3914
Jewelers' findings and materials, and lapidary work	3915
Musical instruments	3931
Dolls	3942
Games, toys and childrens' vehicles; except dolls and bicycles	3944
Sporting and athletic goods, not elsewhere classified	3949
Pens, mechanical pencils, and parts	3951
Lead pencils, crayons, and artists' materials	3952
Marking devices	3953
Carbon paper and inked ribbons	3955
Costume jewelry and costume novelties, except precious metal	3961
Feathers, plumes, and artificial trees and flowers	3962
Buttons	3963
Needles, pins, hooks and eyes, and similar notions	3964
Brooms and brushes	3991
Signs and advertising displays	3993
Burial caskets (3994)	3995
Linoleum, asphalted-felt base, and other hard surface floor coverings, not elsewhere classified	3996
Manufacturing industries, not elsewhere classified	3999

CARD 4

AGRICULTURE, FORESTRY, FISHING (0000 - 0999)

CATEGORY #	CATEGORY NAME	SIC RANGE
1	AGRICULTURAL PRODUCTION (wheat, corn, cotton, tobacco, etc.)	0110 - 0190*
2	AGRICULTURAL PRODUCTION - LIVESTOCK (beef, cattle, hogs, dairy farms, etc.)	0210 - 0291*
3	AGRICULTURAL SERVICES (soil preparation, crop services, animal services, etc.)	0710 - 0783
4	FORESTRY (timber tracts, forest nurseries, etc.)	0810 - 0851*
5	FISHING, HUNTING, TRAPPING (commercial fishing, fish hatcheries, etc.)	0910 - 0971*

CATEGORY	SIC
Crop preparation services for market, except cotton ginning	0723
Cotton ginning	0724
Veterinary services for animal specialities	0742
Lawn and garden services	0782
Ornamental shrub and tree services	0783

CARD 3

OIL AND GAS EXTRACTION (1300 - 1389)

CATEGORY #	CATEGORY NAME	SIC RANGE
1	CRUDE PETROLEUM AND NATURAL GAS	1310 - 1311

CATEGORY	SIC
Crude petroleum and natural gas	1311

CATEGORY #	CATEGORY NAME	SIC RANGE
2	NATURAL GAS LIGUIDS	1320 - 1321

CATEGORY	SIC
Natural gas liguids	1321

CATEGORY #	CATEGORY NAME	SIC RANGE
3	OIL AND GAS FIELD SERVICES	1380 - 1389

CATEGORY	SIC
Drilling oil and gas wells	1381
Oil and gas field exploration services	1382
Oil and gas field services, not elsewhere classified	1389

CARD 6

CONSTRUCTION OR SPECIAL TRADE CONTRACTOR (1500 - 1999

CATEGORY =	CATEGORY NAME	SIC RANGE
1	GENERAL BUILDING CONTRACTING (residential and nonresidential buildings, etc.)	1520 - 1542

CATEGORY	SIC
General contractors - single family homes	1521
General contractors - residential buildings, other than single family	1522
Operative builders	1531
General contractors - industrial buildings and warehouses	1541
General contractors - nonresidential buildings, other than industrial buildings and warehouses	1542

2	HEAVY CONSTRUCTION CONTRACTING (highways, bridges, sewer lines, etc.)	1610 - 1629

CATEGORY	SIC
Highway and street construction, except elevated highways	1611
Bridge, tunnel, and elevated highway construction	1622
Water, sewer, pipeline, communication and power line construction	1623
Heavy construction, not elsewhere classified	1629

3	SPECIAL TRADE CONTRACTING (plumbing, painting, electrical work, etc.)	1710 - 1799

CATEGORY	SIC
Plumbing, heating (except electric), and air conditioning	1711
Painting, paper hanging, and decorating	1721
Electrical work	1731
Masonry, stone setting, and other stone work	1741
Plastering, drywall, acoustical and insulation work	1742
Terrazzo, tile, marble and mosaic work	1743
Carpentering	1751
Floor laying and other floor work, not elsewhere classified	1752
Roofing and sheet metal work	1761
Concrete work	1771
Water well drilling	1781
Structural steel erection	1791
Glass and glazing work	1793
Excavating and foundation work	1794
Wrecking and demolition work	1795
Installation or erection of building equipment, not elsewhere classified	1796
Special trade contractors, not elsewhere classified	1799

CARD 6 (cont.

CONSTRUCTION OR SPECIAL TRADE CONTRACTOR (1500 - 1999)

CATEGORY #	CATEGORY NAME	SIC RANGE
4	DEVELOPING OR BUILDING FOR SALE COOPERATIVE OR SPECULATIVE BUILDING, ETC.)	1531*

CARD 7

TRANSPORTATION, COMMUNICATIONS, ELECTRIC, GAS AND SANITARY SERVICES (4000 - 4999)

CATEGORY =	CATEGORY NAME	SIC RANGE
1	RAILROAD TRANSPORTATION (railroads, terminals, etc.)	4010 - 4041

CATEGORY	SIC
Switching and terminal establishments	4013

CATEGORY =	CATEGORY NAME	SIC RANGE
2	LOCAL AND SUBURBAN PASSENGER TRANSPORTATION (bus, rail or subway, etc.)	4110 - 4172

CATEGORY	SIC
Local and suburban transit	4111
Local passenger transportation, not elsewhere classified	4119
Taxicabs	4121
Schoolbuses	4151
Terminal and joint terminal maintenance facilities for motor vehicle passenger transportation	4171
Maintenance and service facilities for motor vehicle passenger transportation	4172

CATEGORY =	CATEGORY NAME	SIC RANGE
3	TRUCKING AND WAREHOUSING (trucking, public warehousing, etc.)	4210 - 4231

CATEGORY	SIC
Local trucking without storage	4212
Local trucking with storage	4214
Farm product warehousing and storage	4221
Refrigerated warehousing	4222
Household goods warehousing and storage	4224
General warehousing and storage	4225
Special warehousing and storage, not elsewhere classified	4226
Terminal and joint terminal maintenance facilities for motor freight transportation	4231

CATEGORY =	CATEGORY NAME	SIC RANGE
4	U.S. POSTAL SERVICE	4311*

TRANSPORTATION, COMMUNICATIONS, ELECTRIC, GAS AND SANITARY SERVICES (4000 - 4999)

CATEGORY #	CATEGORY NAME	SIC RANGE
5	WATER TRANSPORTATION (deep sea, coastal, river and canal, ferries, etc.)	4410 - 4469

CATEGORY	SIC
Deep sea foreign transportation	4411
Transportation to and between noncontiguous territories	4421
Coastwise transportation	4422
Intercoastal transportation	4423
Great lakes - St. Lawrence seaway transportation	4431
Transportation on rivers and canals	4441
Ferries	4452
Lighterage	4453
Towing and tugboat service	4454
Local water transportation, not elsewhere classified	4459
Marine cargo handling	4463
Canal operation	4464
Water transportation services, not elsewhere classified	4469

6 TRANSPORTATION BY AIR (airlines, airports, etc.) — 4510 - 4583

CATEGORY	SIC
Air transportation, certified carriers	4511
Air transportation, noncertified carriers	4521
Airports and flying fields	4582
Airport terminal services	4583

7 PIPELINES, EXCEPT NATURAL GAS (crude petroleum, pipelines, refined petroleum, etc.) — 4610 - 4619

CATEGORY	SIC
Crude petroleum pipe lines	4612
Refined petroleum pipe lines	4613
Pipe lines, not elsewhere classified	4619

TRANSPORTATION, COMMUNICATIONS, ELECTRIC, GAS AND SANITARY SERVICES (4000 - 4999)

CATEGORY #	CATEGORY NAME	SIC RANGE
8	TRANSPORTATION SERVICES (freight forwarding, travel agencies, etc.)	4710 - 4789

CATEGORY	SIC
Rental of railroad cars with care of lading	4742
Inspection and weighing services connected with transportation	4782
Packing and crating	4783
Fixed facilities for handling motor vehicle transportation not elsewhere classified	4784
Services incidental to transportation, not elesewhere classified	4789

9	COMMUNICATION (telephone communication, radio and television broadcasting, etc.)	4810 - 4899

CATEGORY	SIC
Telephone communication (wire or radio)	4811
Telegraph communication (wire or radio)	4821
Radio broadcasting	4832
Television broadcasting	4833
Communication services, not elsewhere classified	4899

10	ELECTRIC, GAS AND SANITARY SERVICES (electrical generation and distribution, gas transmission and distribution, sewage systems, etc.)	4910 - 4971

CATEGORY	SIC
Electric services	4911
Natural gas transmission	4922
Natural gas transmission and distribution	4923
Natural gas distribution	4924
Mixed, manufactured or liquefied petroleum gas production and/or distribution	4925
Electric and other services combined	4931
Gas and other services combined	4932
Combination utilities, not elsewhere classified	4939
Water supply	4941
Sewerage systems	4952
Refuse systems	4953
Sanitary services, not elsewhere classified	4959
Steam supply	4961
Irrigation systems	4971

CARD 8

FINANCE, INSURANCE AND REAL ESTATE (6010 - 6799)

CATEGORY #	CATEGORY NAME	SIC RANGE
1	BANKING (commercial banks, mutual savings banks, etc.)	6010 - 6059*
2	CREDIT AGENCIES, OTHER THAN BANKS (credit unions, savings and loan associations, etc.)	6110 - 6163*
3	SECURITY AND COMMODITY BROKERS AND SERVICES	6210 - 6281*
4	INSURANCE CARRIERS (life, health insurance, etc.)	6310 - 6399*
5	INSURANCE AGENTS, BROKERS AND SERVICE	6410 - 6411*
6	REAL ESTATE (real estate agents, operators, etc.)	6510 - 6553*
7	COMBINATION OF REAL ESTATE, INSURANCE, LOANS AND LAW OFFICES	6610 - 6611*
8	HOLDING AND OTHER INVESTMENT OFFICES (holding offices, trusts, etc.)	6710 - 6799*

CARD 9

SPECIALIZED SERVICES (7000 - 8999)

CATEGORY #	CATEGORY NAME	SIC RANGE
1	HOTELS AND OTHER LODGING PLACES	7010 - 7041*
2	PERSONAL SERVICES (laundry, barber shops, shoe repair, etc.)	7210 - 7299

CATEGORY	SIC
Power laundries, family and commercial	7211
Garment pressing, and agents for laundries and dry cleaners	7212
Linen supply	7213
Diaper service	7214
Coin-operated laundries and dry cleaning	7215
Dry cleaning plants, except rug cleaning	7216
Carpet and upholstery cleaning	7217
Laundry and garment services, not elsewhere classified	7219
Photographic studios, portrait	7221
Beauty shops	7231
Barber shops	7241
Shoe repair shops, shoe shine parlors, and hat cleaning shops	7251
Funeral service and crematories	7261
Miscellaneous personal services	7299

3	BUSINESS SERVICES (advertising, mailing, building maintenance, data processing, etc.)	7310 - 7399

CATEGORY	SIC
Direct mail advertising services	7331
Blueprinting and photocopying services	7332
Commercial photography, art, and graphics	7333
Stenographic services; and reproduction services, not elsewhere classified	7339
Window cleaning	7341
Disinfecting and exterminating services	7342
Cleaning and maintenance services to dwellings and other buildings, not elsewhere classified	7349
Research and development laboratories	7391
Photofinishing laboratories	7395
Commercial testing laboratories	7397
Business services, not elsewhere classified	7399

SPECIALIZED SERVICES (7000 - 8999)

CATEGORY #	CATEGORY NAME	SIC RANGE
4	AUTO REPAIR, SERVICES AND GARAGES (auto rentals, general auto repair, etc.)	7510 - 7549

CATEGORY	SIC
Passenger car rental and leasing, without drivers	7512
Utility trailer and recreational vehicle renting	7519
Top and body repair shop	7531
Tire retreading and repair shops	7534
Paint shops	7535
General automotive repair shops	7538
Automotive repair shops, not elsewhere classified	7539
Car washes	7542
Automotive services, except repair and car washes	7549

CATEGORY #	CATEGORY NAME	SIC RANGE
5	OTHER REPAIR SERVICES (radio and TV repair, electric appliance repair, etc.)	7620 - 7699

CATEGORY	SIC
Radio and television repair shops	7622
Refrigeration and air conditioning service and repair shops	7623
Electrical and electronic repair shops, not elsewhere classified	7629
Watch, clock, and jewelry repair	7631
Reupholstry and furniture repair	7641
Welding repair	7692
Armature rewinding shops	7694
Repair shops and related serivces, not elsewhere classified	7699

6	MOTION PICTURES (motion picture production, distribution, theaters, etc.)	7810 - 7833*
7	AMUSEMENTS AND RECREATION SERVICES (bowling, orchestras, golf courses, etc.)	7910 - 7999*
8	HEALTH SERVICES (offices of doctors, dentists, hospitals, medical laboratories, etc.)	8010 - 8091

CATEGORY	SIC
General medical and surgical hospitals	8062
Medical laboratories	8071
Dental laboratories	8072
Health and allied services, not elsewhere classified	8091

CARD 9 (cont.

SPECIALIZED SERVICES (7000 - 8999)

CATEGORY #	CATEGORY NAME	SIC RANGE
9	LEGAL SERVICES	8110 - 8111*
10	EDUCATIONAL SERVICES (elementary schools, colleges, libraries, etc.)	8210 - 8299*
11	SOCIAL SERVICES (child day care, individual and family services, etc.)	8320 - 8399*
12	MUSEUMS, BOTANICAL, ZOOLOGICAL GARDENS	8410 - 8421
13	MEMBERSHIP ORGANIZATIONS (trade associations, labor organizations, etc.)	8610 - 8699*
14	OTHER SERVICES (accounting, engineering services, etc.)	8910 - 8999*

CATEGORY	SIC
Museums and art galleries	8411
Arboreta, botanical, and zoological gardens	8421