SAMPLING STRATEGY FOR AN OCCUPATIONAL EXPOSURE SURVEY

William Karl Sieber, Jr., National Institute for Occupational Safety and Health

I. Introduction

The development of a sampling strategy for an occupational hazard/ exposure survey involves consideration of many factors. The range of industries as well as the types of results expected from the survey are important. In performing such a survey the industrial base is so large that a suitable sampling frame should first be established. Similarly, given financial and temporal constraints, the size (defined as the number of employees) of individual facilities is important in determining an overall number of facilities to be interviewed. Results needed from such a survey also may not simply be numbers of workers employed, or numbers of facilities in various industries, but also numbers of employees in specific subgroups such as unions, those undergoing medical tests, or employees potentially exposed to chemical or physical hazards. This paper reviews some considerations involved in designing an occupational exposure survey. Different sampling strategies are compared using results from the National Occupational Hazard Survey (NOHS) and National Occupational Exposure Survey (NOES) conducted by the National Institute for Occupational Safety and Health (NIOSH) between 1972-1974 and 1981-1983, respectively.

II. The NOHS and NOES

Both NOHS and NOES had similar objectives: [U.S. Department of Health, Education, and Welfare, 1977, 1985]

1. For selected industrial sectors, to develop estimates of the number of workers potentially exposed to chemical, physical, and biological agents.
2. To develop data that describe the nature and extent of these potential exposures and the degree to which businesses have implemented programs to reduce occupational health problems.

The sampling frame for each survey was defined as all employees working in facilities or job sites located in the United States reporting eight or more employees and with a primary activity or line of business on a list of selected Standard Industrial Classification (SIC) codes [Office of Management and Budget, 1972]. The number of employees in a facility was determined by comparison to the County Business Patterns publication of the Bureau of the Census [Bureau of the Census, 1982] (for NOHS) or Dun's Marketing Index [Dun and Bradstreet, 1980] (for NOES). The choice of which SIC codes to include was based on those industries whose employees were thought to have the greatest exposure to different agents.

In both NOHS and NOES, primary sampling units (PSUs) were defined from which individual facilities were selected. A two-stage selection procedure beginning at the PSU level was used. Each PSU defined a geographic cluster of business and industrial facilities. 247 PSUs were defined in NOHS and 604 in NOES. In the first stage, all PSUs were stratified by employee concentration by SIC and geography, and those PSUs from which facilities were to be selected were determined. Stratification was imposed at this stage to best handle data from the many industries and employee size groups included in the survey. One PSU was selected from each stratum, resulting in 67 PSUs in NOHS and 98 PSUs in NOES being selected in the first stage sample. With the exception of samples of very large establishments in NOES drawn irrespective of geographic location (because of the few facilities involved and proximity to selected PSUs), the interviewed sample was confined to the selected PSUs.

The second stage selection involved selection of facilities from the selected PSUs. Both NOHS and NOES used a systematic procedure, resulting in 5,745 and 4,894 facilities being selected in NOHS and NOES, respectively. Figure I compares the sampling plans in the NOHS and NOES.

III. Sampling Strategy

A. The Sampling Frame

Two-digit SIC codes included in both the NOHS and the NOES and the number of facilities interviewed in each SIC category are shown in Table I. Not all SIC codes were represented in either NOHS or NOES. Little exposure to hazards of any type occur in some industries, e.g., banking (SIC 60-67); others, such as mining (SIC 14), are so large and heterogeneous that they warrant surveys of their own; while others, such as private households (SIC 88), are difficult to survey accurately. In both the NOHS and the NOES the following industries were excluded:

1. agricultural production.
2. mining, except oil and gas extraction.
3. railroad transportation.
4. private households.
5. federal, state, and municipal government.

Financial, insurance, and real estate facilities were completely excluded from the NOES.

In both NOHS and NOES survey establishments were also limited to those employing eight or more employees. This had practical as well as financial motivation: interviews, particularly for smaller establishments, were difficult to obtain since smaller operations tend to appear and disappear quicker than larger ones. On the other hand, the few large facilities took more time to interview and collect exposure data. To be considered eligible for the survey, the facility must

also have been located in metropolitan or urbanized areas of the 50 states, and have been a worksite operating in one of the survey SICs during the period of the NOHS or NOES.

B. The Sampling Method

A two-stage selection procedure was used in both NOHS and NOES. The first stage involved selection of PSUs from which facilities were selected in the second stage. This two stage approach was used so that a final sample as representative as possible of all national industries considered in the survey might be obtained.

The two-stage approach of NOHS and NOES was a compromise in obtaining statistics on number of employees in all in-scope industries and by industry. Stratification (in the first stage) and selection of PSUs with probability proportional to number of employees gave a sample of PSUs representative of the range of numbers of employees and types of exposures by PSU, while systematic selection of facilities in the second stage gave a sample of facilities proportional to the total number of facilities in a given industry.

Stratification of both NOHS and NOES PSUs was imposed to best handle data from the many industries and varying number of employees in each industry included in the survey. Groups of PSUs with significant concentrations of employees in certain industries were also likely to have serious and common health hazards. Grouping these PSUs before selection reduced the variance expected in results from the survey. Strata were defined of approximately equal number of employees such that the PSUs within the strata were as homogeneous as possible, with respect to industries in the PSU. Homogeneity within strata reduced variance between PSUs within each stratum so that when a single PSU was selected from that stratum in the first stage sample, that PSU would be characteristic of all PSUs in the stratum.

Not all PSUs could be grouped into strata, however. Some PSUs were much larger than the desired average stratum size. In the NOHS and NOES, the largest of these PSUs were defined as separate strata (self-representing PSUs) and the remaining PSUs (non self- representing) were grouped into strata of approximately equal size. Each self-representing PSU was included in the first stage sample, whereas a random sample of the non self-representing PSUs was selected for inclusion in the first stage sample.

The 247 Standard Metropolitan Statistical Areas (SMSAs) defined by the Bureau of the Census at the time of the NOHS in 1972 were taken as PSUs for NOHS. PSUs were grouped into 67 strata. Self- representing strata accounted for 31 of these PSUs, while the remaining 211 PSUs were grouped into 36 strata. Since one PSU was selected from each stratum, the first stage NOHS sample consisted of 67 PSUs.

PSUs in NOES were defined nationally as combinations of contiguous counties. This process resulted in 604 PSUs for NOES. These PSUs were grouped into 98 strata: 26 of these were the self-representing strata discussed above, while the remaining 578 non self- representing PSUs were grouped into 72 strata and 1 PSU was selected from each of the 72 strata. The first stage sample for NOES consisted of 98 PSUs, with 1 PSU from each stratum.

Second stage selection of facilities from within PSUs was done using systematic selection in both NOHS and NOES. Samples of facilities were selected independently across size classes, where size was defined as the number of employees at that worksite. The size classes, average number of employees, and number of facilities used in both NOHS and NOES are shown in Table II. In the systematic selection procedure, all facilities in the sample PSU were first ordered and every k^{th} facility on the list was selected for interview. Systematic selection was chosen for both NOHS and NOES because of the need to include facilities operating in many SIC ranges and with varying numbers of employees in the survey. Although both NOHS and NOES used systematic selection, the definition of k (the sampling interval) and method by which the sampling was performed was different in each survey.

In the NOHS, all facilities were selected independently from cells of SIC code by size class. 3,402 cells were formed from 63 SICs included in the survey and 9 size ranges. The sampling interval was based on the proportion of the total number of employees in the SIC found in that SIC-size range cell. Second stage selection resulted in 5,745 facilities being selected for field interviews in NOHS.

In the NOES, facilities were stratified by size class. Facilities in size classes 1 through 8 (8-2499 employees) were then arranged by increasing 4-digit SIC code within each size class. Since facilities in size classes 9 and 10 (2500 employees or more) were not necessarily located within the 98 sample PSUs chosen in the first stage sample, all facilities in these size classes were arranged by zip code and 4-digit SIC code. Systematic selection in each size class was done using a sampling interval based on the proportion of facilities in each size class. This process resulted in 4,894 facilities being selected for interview in the NOES.

C. Estimation

Estimation of numbers of employees and numbers of facilities from interview results in NOHS or NOES involved weighting each survey facility according to the probability of including a facility like it in the sample. The numerical value of the weight was determined by the sampling scheme chosen. In the NOHS, estimates of the number of employees and number of facilities in the sample frame were calculated as the sum of weighted sample values. Variances for these estimates were found in each SIC-size class cell, then summed over all

cells. NOES estimates improved upon the NOHS estimates by using ratio estimation. The ratio factor was determined using listings from an outside source such as the Bureau of the Census publication County Business Patterns (CBP), or the Dun Master Inventory (DMI). The NOES used two ratio factors: one for number of employees, and one for number of facilities. Variances in the NOES were found using balanced repeated replications within 32 replicates [Hansen et al, 1953; McCarthy, 1966].

IV. Results

The different sampling schemes used in NOHS and NOES, although following a similar two-stage sampling strategy, produced different results. Tables III and IV compare the coverage of facilities and employees by SIC category in NOHS and NOES, with respect to tabulated listings shown in the CBP (for NOHS) or DMI (for NOES). Coverage greater than 100% indicates over-representation of that SIC group in the sample, while coverage less than 100% indicates under-representation.

From Table III, 100% coverage of in-scope facilities was obtained from the NOHS sample. Projected coverage was near or greater than 100% in all SICs except oil and gas extraction, indicating over-representation of facilities in most SIC categories in the NOHS. Facilities in oil and gas extraction were under-represented in the NOHS sample. Overall projected coverage in NOES (95%) was lower than that in NOHS. Under-representation of most SIC ranges was found in NOES.

Projected coverage of number of employees from Table IV was 100% from the NOES sample. Coverage is near 100% for contract construction and manufacturing, two SIC ranges on which emphasis was placed in defining SICs to be included in NOES. Compare these results with those from NOHS. NOHS coverage of employees in the surveyed industries was 97%, and was higher in each SIC, except services, than that obtained from NOES. In many cases over-representation of numbers of employees was also found in the NOHS.

Different results in Tables III and IV indicate the effect of different sampling schemes in NOHS and NOES. In the NOHS, the number of facilities was an important factor in sample selection both in defining primary sampling units for first stage selection and in defining cells for second stage selection. Accordingly, coverage of all facilities is 100% in the NOHS and close to 100% for each industrial breakdown. The NOES, on the other hand, used employee counts at every stage of selection. Percent coverage of number of employees was 100% for all industries, and coverage for most individual industrial breakdowns was close to 100%.

The different sampling schemes are also reflected in magnitude of the standard error. Table V shows standard errors for number of facilities and number of employees calculated from NOHS and NOES. Over all industries, standard errors are smaller for estimates made using NOES data compared to those from NOHS data. Standard errors by SIC range are smaller for the NOHS, however, in part because of using larger inflated estimates, and in part because variances were determined over much more homogeneous groups in defined cells than over the larger replicates of the NOES. Note also that a midsurvey reduction in the number of plants to be surveyed with over 500 employees occurred in the NOHS.

V. Conclusions

In a large multi-purpose survey such as NOHS or NOES, where a range of activities is to be surveyed over a large area, a two-stage approach with selection of primary sampling units followed by selection of individual units within those primary units is a useful approach. Some variation in the sampling scheme can occur in each stage.

Important characteristics to be considered in developing the sample scheme are the number of facilities and number of employees within a given SIC. The NOHS emphasized number of facilities in its sampling scheme. Projected coverage of facilities in in-scope industries in the NOHS was 100%, compared to 97% coverage of numbers of employees. The sampling scheme in the NOES was based on number of employees. Projected coverage of employees in NOES is 100%, compared to a coverage of 95% of facilities in in-scope industries. Numbers of employees were more accurately estimated using information from the NOES, while numbers of facilities were better obtained from NOHS. Note, however, that to obtain the information on facilities in the NOHS, almost 1,000 more facilities were selected for interview than were needed for results on numbers of employees in the NOES. Standard errors for estimates of both facilities and employees over all industries were also lower for NOES than for NOHS, in spite of the fact that the NOHS was truncated in facilities with over 500 employees.

REFERENCES

1. Dun and Bradstreet, Inc. (1980): Dun's Marketing Index.
2. Executive Office of the President - Office of Management and Budget (1972): Standard Industrial Classification Manual-1972. U.S. G.P.O. 041-001-00066-6.
3. Hansen, M. H., W. N. Hurwitz, and W. G. Madow (1953): Sample Survey Methods and Theory: Volume II Theory. John Wiley and Sons, New York, p. 218.
4. Hanson, R., D. Ward, J. Edmonds, and J. Escatell (1980): National Occupational Exposure Survey (NOES). Final Report, Contract No. 210-80-0057. Westat, Inc., Rockville, Maryland.
5. Kish, L. (1965): Survey Sampling. John Wiley and Sons, New York.
6. Kish, L. and M. R. Frankel (1970): Balanced Repeated Replications for Standard Errors.

Journal of the American Statistical Association. (65), 331.

7. Levy, P. S., and S. Lemeshow (1980): Sampling for Health Professionals Lifetime Learning Publications, Belmont, California.

8. McCarthy, Philip J. (1966): Replication: An Approach to the Analysis of Data from Complex Surveys. Vital and Health Statistics. PHS No. 1000 - Series 2 No. 14.

9. National Institute for Occupational Safety and Health: National Occupational Exposure Survey: Sampling Methodology. In preparation.

10. U.S. Bureau of the Census (1982): County Business Patterns, 1980, United States. U.S. Department of Commerce G.P.O. No. 003-024-05774-3.

11. U.S. Department of Health, Education, and Welfare (1977): National Occupational Hazard Survey (NOHS): Vol. II, Data Editing and Data Base Development. DHEW (NIOSH) Publication No. 77-213.

12. U.S. Department of Health, Education, and Welfare (1977): National Occupational Hazard Survey (NOHS): Vol. III, Survey Analysis and Supplemental Tables. DHEW (NIOSH) Publication No. 78-114.

FIGURE I. NOHS AND NOES SAMPLING PLANS

NOHS	NOES
247 PSU	604 PSU
stratification into 67 strata	stratification into 98 strata
selection of 67 PSU from which establishments chosen	selection of 98 PSU from which establishments chosen
systematic selection of establishments from 3,402 cells sampling interval based on employment	systematic selection of establishments from 10 size classes sampling interval based on employment
5,745 establishments selected	4,894 establishments selected

CATEGORY		NUMBER OF FACILITIES	
	SIC Range	NOHS	NOES*
Agricultural Services	07-09	47	22
Oil and Gas Extraction	13	32	61
Contract Construction	15-17	503	573
Manufacturing	20-39	2,751	2,656
Utilities	41-49	308	365
Wholesale and Retail Trade	50-59	506	237
Specialized Services	70-89	345	566
Finance, Insurance, and Real Estate	60-67	144	0
TOTAL		4,636	4,480

* Only SIC 07 was sampled in agricultural services in NOES.

TABLE II. AVERAGE NUMBER OF FACILITIES AND NUMBER OF EMPLOYEES
BY EMPLOYEE SIZE CLASS, NOHS AND NOES

Employee Size Class (Number of Employees)	NOHS		NOES	
	Number of Facilities	Average Number of Employees	Number of Facilities	Average Number of Employees
8-19	2023	12.4	1190	12.8
20-49	1268	32.4	914	31.8
50-99	842	71.7	675	70.0
100-249	567	158.1	838	155.5
250-499	376	349.9	512	341.3
500-999	279	690.1	343	705.7
1000-1499*	194	1200	123	1186.1
1500-2499*		1900	108	1880.6
2500-4999	126	3500	94	3470.8
5000+	70	9250	97	9051.7
TOTAL	5745		4894	

* These classes not distinguished in NOHS.

TABLE III. PROJECTED COVERAGE OF SURVEY FACILITIES, NOHS AND NOES

Standard Industrial Classification	Percent Coverage	
	NOHS	NOES
All Industries	100%	95%
Agricultural Services	103%	133%
Oil and Gas Extraction	70%	80%
Contract Construction	107%	88%
Manufacturing	109%	87%
Utilities	108%	126%
Wholesale and Retail	97%	107%
Services	94%	99%
Finance	104%	--

TABLE IV. PROJECTED COVERAGE OF PROJECTED POPULATION, NOHS AND NOES

Standard Industrial Classification	Percent Coverage	
	NOHS	NOES
All Industries	97%	100%
Agricultural Services	111%	108%
Oil and Gas Extraction	87%	67%
Contract Construction	113%	98%
Manufacturing	114%	97%
Utilities	123%	118%
Wholesale and Retail	91%	83%
Services	74%	104%
Finance	70%	--

TABLE V. STANDARD ERRORS FOR NUMBER OF FACILITIES AND EMPLOYEES, NOHS AND NOES

	Standard Error of Facilities		Standard Error of Employees	
	NOHS*	NOES	NOHS*	NOES
All Industries	17,195	6,648	2,296,649	487,521
Agricultural Services	445	1,556	8,617	29,333
Oil and Gas Extraction	150	2,132	8,781	101,840
Contract Construction	4,529	2,409	105,753	132,323
Manufacturing	3,586	3,677	565,239	557,816
Utilities	1,511	3,528	275,772	289,123
Wholesale and Retail	13,448	3,970	378,889	142,606
Services	3,770	3,678	269,949	380,840
Finance	7,941	---	106,882	---

* Affected by midsurvey sample reduction - truncation of surveys in larger
  facilities with over 500 employees.