USING BUSINESS ESTABLISHMENT SIZE IN AN ENVIRONMENTAL SURVEY

Robert P. Clickner, Westat, Inc.
Jim Craig, U.S. Environmental Protection Agency
Robert P. Clickner, 1650 Research Blvd., Rockville, MD 20850

1. INTRODUCTION AND BACKGROUND

Size measures are often used to design surveys when the key response variables are expected to be associated with sample unit size. This paper reports on the successful application of a measure of business establishment size -- the number of employees -- to an environmental survey. The major conclusion presented here is that the use of a design variable that, a priori, does not appear to be very highly associated with the key response variables can be quite fruitful, resulting in considerable savings in sample size and variance.

The survey, sponsored by the United States Environmental Protection Agency (EPA), was conducted to obtain information on industrial solid waste management practices in selected industries. We begin with a little background to show why the survey was conducted, and why we took the approach we did. The sample design is then presented, followed by empirical survey results. We conclude with a discussion of the empirical efficiency of the design.

Under the Resource Conservation and Recovery Act of 1975 (RCRA), EPA promulgated regulations defining hazardous waste and setting standards for treating, storing and disposing it. Reporting requirements were established for generators and managers of hazardous wastes. Under Subtitle D of RCRA, wastes not regulated as hazardous by EPA were left to the States to regulate. No Federal reporting requirements were set up for "non-hazardous" waste.

In 1984, Congress passed a series of amendments to RCRA, one of which required EPA to study and report on the management of non-hazardous solid wastes to determine if the regulatory definition of hazardous waste needed to be modified. This present study of industrial non-hazardous solid waste ("Subtitle D waste") is a part of this Congressionally mandated effort.

In particular, EPA focused attention on industrial establishments that generate Subtitle D waste and manage it onsite. Because the management of Subtitle D wastes was left to the States, EPA had very little knowledge about how much waste is generated, what are the prevalent practices, or even which establishments generate the waste. It was therefore necessary to conduct the study in two phases: (1) conduct a screening survey of all industrial establishments to ascertain which ones generate industrial Subtitle D waste; and (2) conduct a follow-up survey of the eligible establishments. This paper is devoted entirely to the screening survey.

In order to make the most efficient use of limited resources, EPA focused attention on the 17 industrial groups that, according to prior available information, generate the greatest amounts of Subtitle D waste. (EPA had previously conducted a study [2] of secondary data sources on the generation and management of industrial Subtitle D waste. That study was the basis for the selection of the 17 industries. It also revealed large gaps in the available information on Subtitle D waste.) These 17 groups include electric power generation, water treatment, petroleum refining, five categories of chemical products manufacturing, and nine categories of other manufacturing: primary iron and steel, stone/glass/clay/ concrete, pulp and paper, primary nonferrous metals, food and kindred products, rubber products, transportation equipment, textiles, and leather products.

The specific objectives of the study were therefore to:

- Develop a list of industrial establishments that generate industrial Subtitle D waste, and estimate the number of these establishments, nationwide and by industry.
- Develop a list of industrial establishments that generate Subtitle D solid waste and manage it onsite using certain types of waste management facilities, i.e., landfills, surface impoundments (typically wastewater treatment ponds), land application units (waste treatment farms), and waste piles, and estimate the number of these establishments, nationwide and by industry.
- Estimate the numbers of industrial landfills, surface impoundments, land application units, and waste piles, nationwide and by industry.
- Estimate the amounts of waste generated and managed in these waste management facilities, nationwide and by industry.

Because the survey was primarily a screening survey to identify industrial waste generators, we decided to conduct it as a telephone survey using computer assisted telephone interviewing (CATI) techniques. This would collect the data quickly and efficiently, compared with mail or onsite data collection. In the development of the study, a few questions were added concerning the characteristics of the waste generated and the waste management practices.

A list of all establishments in the 17 target industries was purchased from Duns Marketing, Inc., and used as the frame. The data provided by Duns Marketing included establishment name, locational and mailing addresses and phone number, up to six SIC codes identifying the establishment's industry(s), number of employees onsite, corporate revenue and other financial data. There were nearly 150,000 establishments on the list.

Table 1 displays the 17 in-scope industries, their SIC codes, and the number of establishments provided by Duns Marketing. Table 2 shows the prior estimated occurrence rate of establishments with one or more of the four types of target waste management facilities. These prior rates were the prior expected eligibility rates for the screening study. The occurrence rate data were developed from data in two prior EPA studies ([2], [4]). The prior rates were of uncertain accuracy because they were based upon numerous sources of varying age and statistical validity, as well as unverified assumptions.

2. SAMPLE DESIGN

The prior occurrence rates in Table 2 show a need for a sample design that would enhance the eligibility rate in the sample above the nationwide average of seven percent. We

estimated that 2,500 to 3,000 eligible establishments would be needed for the second phase follow-up survey. This would have required screening 30,000 to 40,000 establishments, if we used simple random sampling in each industry. The budget and schedule did not allow for a screening effort this large.

Our solution was to stratify by establishment size, for two basic reasons. First, it was believed that the larger establishments would be more likely to manage their waste onsite, rather than shipping it offsite. Second, larger factories generate more waste than smaller ones. In fact, we expected the distribution to be highly skewed. A prior study of hazardous waste generators and managers [3] had found that 5 percent of the facilities generated 95 percent of the waste. Thus both the expected eligibility rate in the sample and the variance of the estimated amount of waste would be improved by a size-based design.

We selected the number of employees as the measure of size because it was the only one available at the required, i.e., establishment level. Other size measures that were believed to associate better with amount of industrial waste generated, e.g., production volume or revenue, were available only at the corporate level or not available at all. Nevertheless, for several reasons, we expected a low, but positive, association between the size measure and the key response variables: administrative offices can have many employees, but generate no industrial waste; two factories, one automated and one not, could produce the same volume of product and hence the same volume of waste, but have greatly differing numbers of employees.

For these reasons, and because we were also interested in estimating proportions, we did not want a design highly dependent on establishment size. Therefore, as a compromise between simple random sampling and PPS sampling (say), three size strata were developed independently within each of the 17 industries; small, medium and large.

The three strata were constructed using optimality criteria (Dalenius-Hodges bounds), adjusted to minimize the potential adverse impacts of (1) a design variable that was expected to have a low association with the response variables and (2) potential errors in the frame data. The medium-large breakpoints ranged from 40 to 650 employees, except for one industry at 2,100 employees (transportation equipment manufacturers). The small-medium breakpoints ranged from 13 to 34 employees.

The large establishments were sampled with certainty in order to maximize the eligibility rate in the sample and control the variance of the estimated totals. As indicated above, we anticipated a highly skewed distribution of waste quantities generated. The sampling sizes and rates for the medium and small establishments varied with industry. The sample size for the small establishments was set to achieve plus or minus five percent sampling error at 95 percent confidence for estimating a proportion (e.g., small establishments with onsite facilities) assumed to be 0.05. The sample size was adjusted upwards, if necessary, to sample at least four percent of the frame, to keep the stratum sampling weights from exceeding 25. (Because of some later adjustments, a few industries have weights slightly above 25.) The resultant small establishment sample sizes ranged from 112 to 880 and the rate varied from 3.9 percent to 14.7 percent.

A similar analysis was conducted for the medium establishments with the assumed proportion equal to 0.20 and the maximum weight at 10. The resultant medium establishment sample sizes ranged from 78 to 1,559 and the sampling rate varied from 13.5 percent to 100 percent.

The resulting total sample size was 18,051. Table 1 displays the sample sizes by industry. On a size basis, the frame counts and sample sizes were as follows.

Size Stratum	Frame Count	Sample Size
Small	106,750	3,887
Medium	37,315	9,078
Large	5,086	5,086
Total	149,151	18,051

The 18,051 establishments were selected independently within each of the 17 industries and three size strata. Approximately 12.3 percent of the sampled establishments were found to be out of scope (went out of business, moved, duplicate entries, etc.) or unlocatable (wrong address or phone number and no directory assistance record), leaving 15,832 establishments in the sample. Only 36 establishments (0.2 percent) refused to respond to the survey.

3. RESULTS OF THE STUDY

The discussion of the results will emphasize the empirical results on the nature and strength of the relationship between the design and response variables. That is, rather than reporting the estimated number of industrial surface impoundments, we discuss how the design variable improved the efficiency of the estimates.

Nearly 11,000 of the sampled establishments were found to generate industrial Subtitle D waste, of which 2,718 had one or more of the four types of waste management units (landfills, surface impoundments, land application units or waste piles). Table 3 displays the estimated population occurrence rates of establishments with waste management units by industry and size. Fifteen of the 17 industries show the eligibility rate increasing with establishment size. In some industries the increase is quite large. For example, the eligibility rate increases from 9 percent to 50 percent among inorganic chemical manufacturers. The two exceptions are waste treatment plants and leather and leather products manufacturers. Both of these show essentially the same flat pattern: a small increase from small to medium, followed by a small decrease from medium to large. In both industries, the chi-square test of equality of the three percentages failed to reject the null hypothesis.

Table 2 displays the eligibility rate in the sample, i.e., the ratio of the unweighted number of eligibles in the sample to the sample size. A comparison of these rates with the prior rates shows that the design variable considerably improved the eligibility rate from the prior expectation, from 7 percent to 17 percent, nationally. A further comparison of the prior and sample eligibility rates in Table 2 with the total estimated population rates in Table 3 leads to two conclusions. First, the prior rates were not very accurate in many industries. Second, the design variable also realized an improvement over the "posterior" estimated rates, from 11 percent to 17 percent, nationally. ("Posterior" is in quotes because the estimates are not true posterior estimates in the Bayesian sense.)

4. EFFICIENCY

We present here some empirical results on the efficiency of the design with respect to simple random sampling within industries. Table 4 presents the basic data. The first two columns of Table 4 display the actual number of establishments screened and the actual number of eligibles in the sample. The third column displays the expected number of eligibles assuming a simple random sample and the total eligibility rates in Table 3. Nationally, the 15,832 establishments screened would have yielded only 1,775 eligibles -- a 35 percent reduction from 2,718 -- had simple random sampling been employed. The last column of Table 4 presents the converse view. It shows the estimated number of establishments that it would have been necessary to screen using simple random sampling to achieve the number of eligibles actually achieved (shown in the second column). The total number of screeners would have increased to 27,452. Thus, the use of a design variable (number of employees) that was not expected to be highly associated with the presences of waste management units resulted in a savings of over 11,000 in the sample size.

5. ACKNOWLEDGEMENTS

We deeply appreciate the assistance of all the project staff who contributed greatly to the success of the project. In particular, we wish to thank Kirsten Schroeder, Esther Miller, Rotraut Bockstahler, Lisa Puhl, and numerous telephone interviewers, their supervisors, coders, editors and clerical staff.

6. REFERENCES

[1] Census of State and Territorial Subtitle D Non-hazardous Waste Programs. Report by Westat, Inc. under EPA Contract No. 68-01-7047. (1986)

[2] Summary of Data on Industrial Non-hazardous Waste Disposal Practices. Report by Science Applications International Corporation under EPA Contract No. 68-01-7050. (1985)

[3] National Survey of Hazardous Waste Generators and Treatment, Storage and Disposal Facilities Regulated Under RCRA in 1981. Report by Westat, Inc. under EPA Contract No. 68-01-6861, Subcontract No. EPA 33-01. (1984)

[4] Surface Impoundment Assessment National Report. United States Environmental Protection Agency. EPA 570/9-84-002. (1983)

TABLE 1: IN-SCOPE INDUSTRIAL GROUPS, FRAME AND SAMPLE SIZES

INDUSTRY GROUP	SIC CODES:	FRAME COUNT	SAMPLE SIZE
ORGANIC CHEMICALS	2865, 2869	1,413	560
PRIMARY IRON AND STEEL	3312-3321	3,581	975
FERTILIZER & AGRICULTURAL CHEM.	2873-2879	2,159	603
ELECTRIC POWER GENERATION	4911	3,945	1,146
PLASTICS AND RESINS MFR.	2821	1,235	575
INORGANIC CHEMICALS	2812-2819	2,949	692
STONE, CLAY, GLASS & CONCRETE	32	23,109	1,951
PULP AND PAPER INDUSTRY	26	9,212	1,090
PRIMARY NONFERROUS METALS	3331-3399	6,851	861
FOOD AND KINDRED PRODUCTS	20	30,161	2,245
WATER TREATMENT	4941	3,341	615
PETROLEUM REFINING	29	3,712	1,148
RUBBER AND MISC. PRODUCTS	30	19,564	1,241
TRANSPORTATION EQUIPMENT	37	18,048	1,909
SELECTED CHEM. & ALLIED PROD.	2822, 2824, 2851, 2891	3,392	692
TEXTILE MANUFACTURING	22	11,732	1,065
LEATHER AND LEATHER PROD.	31	4,747	683
ALL INDUSTRIES		149,151	18,051

TABLE 2: PRIOR AND SAMPLE ESTIMATED ELIGIBILITY RATES BY INDUSTRY

INDUSTRY GROUP	PRIOR ELIGIBILITY RATE	SAMPLE ELIGIBILITY RATE
ORGANIC CHEMICALS	45%	16%
PRIMARY IRON AND STEEL	17	31
FERTILIZER & AGRICULTURAL CHEM.	19	21
ELECTRIC POWER GENERATION	21	18
PLASTICS AND RESINS MFR.	26	17
INORGANIC CHEMICALS	12	32
STONE, CLAY, GLASS & CONCRETE	5	31
PULP AND PAPER INDUSTRY	6	20
PRIMARY NONFERROUS METALS	4	16
FOOD AND KINDRED PRODUCTS	11	15
WATER TREATMENT	4	17
PETROLEUM REFINING	26	20
RUBBER AND MISC. PRODUCTS	1	4
TRANSPORTATION EQUIPMENT	2	7
SELECTED CHEM. & ALLIED PROD.	9	9
TEXTILE MANUFACTURING	4	14
LEATHER AND LEATHER PROD.	1	2
ALL INDUSTRIES	7	17

TABLE 3: ESTIMATED POPULATION ELIGIBILITY RATES BY INDUSTRY AND SIZE

INDUSTRY GROUP	SIZE			
	SMALL	MEDIUM	LARGE	TOTAL
ORGANIC CHEMICALS	8%	10%	22%	12%
PRIMARY IRON AND STEEL	9	26	46	21
FERTILIZER & AG. CHEM.	7	15	31	12
ELECTRIC POWER GENERATION	6	11	29	11
PLASTICS AND RESINS MFR.	3	12	30	11
INORGANIC CHEMICALS	9	28	50	18
STONE, CLAY, GLASS & CONCRETE	19	35	38	26
PULP AND PAPER INDUSTRY	3	7	45	7
PRIMARY NONFERROUS METALS	5	10	38	8
FOOD AND KINDRED PRODUCTS	10	12	25	11
WATER TREATMENT	15	18	16	16
PETROLEUM REFINING	10	19	27	14
RUBBER AND MISC. PRODUCTS	1	3	17	2
TRANSPORTATION EQUIPMENT	2	5	16	3
SELECTED CHEM. & ALLIED PROD.	1	6	19	4
TEXTILE MANUFACTURING	3	10	28	7
LEATHER AND LEATHER PROD.	1	3	1	2
ALL INDUSTRIES	8	13	31	11

TABLE 4: DESIGN EFFICIENCY RELATIVE TO SIMPLE RANDOM SAMPLING (SRS)

INDUSTRY GROUP	ACTUAL NUMBER		EXPECTED NUMBER (SRS)	
	SCREENED	SAMPLE ELIGIBLES	SAMPLE ELIGIBLES	SCREENED
ORGANIC CHEMICALS	482	78	58	650
PRIMARY IRON AND STEEL	867	265	182	1,262
FERTILIZER & AG. CHEM.	514	106	62	883
ELECTRIC POWER GENERATION	1,104	199	121	1,809
PLASTICS AND RESINS MFR.	505	88	56	800
INORGANIC CHEMICALS	609	199	110	1,106
STONE/CLAY/GLASS/CONCRETE	1,684	519	438	1,996
PULP AND PAPER INDUSTRY	972	194	68	2,771
PRIMARY NONFERROUS METALS	755	118	60	1,475
FOOD AND KINDRED PRODUCTS	1,968	296	216	2,691
WATER TREATMENT	567	94	91	588
PETROLEUM REFINING	1,020	206	143	1,471
RUBBER AND MISC. PRODUCTS	1,067	44	21	2,200
TRANSPORTATION EQUIPMENT	1,643	117	49	3,900
SELECTED CHEM. & ALLIED PROD.	619	56	25	1,400
TEXTILE MANUFACTURING	911	126	64	1,800
LEATHER AND LEATHER PROD.	545	13	11	650
ALL INDUSTRIES	15,832	2,718	1,775	27,452