

SECTION 4

SAMPLE DESIGN, ESTIMATION OF SAMPLE WEIGHTS AND VARIANCES

The national probability sample for the Underground Storage Tank survey was drawn in a three-stage sample design that involved sampling establishments from establishment frame lists within 34 survey sites which had been sampled to represent six survey regions. Data were collected from the sampled establishments using several data collection techniques. All sampled establishments were first screened to determine survey eligibility, that is, whether they had an underground motor fuel storage tank. At eligible establishments, the owner or operator of the tank was interviewed in person and instructed in the completion of 30-day inventory records. A sub-sample of the eligible establishments was selected for physical tank tightness testing.

This section reviews the target universe of the survey and then describes the three stages of sampling: Primary Sampling Units (PSUs) or survey sites; establishments for questionnaire and inventory data collection; and the sub-sample of eligible establishments for physical tightness tests. In brief, the sample consisted of 34 PSUs in which 2,218 establishments were sampled. Of these, 896 establishments were eligible for the survey, (i.e., had underground motor fuel storage tanks that were not abandoned) and 890 cooperated with the interview phase. Two hundred eighteen were selected for physical tank testing, which was accomplished at 202 establishments. The section concludes by describing the methods used to calculate the final weights used in making national estimates from the survey and to estimate the sampling error of those estimates. Appendix A gives more details

on these subjects, and Appendix G gives a detailed account of the farm sample.

I. SCOPE OF THE SURVEY

The scope of the Underground Storage Tank survey was limited, for practical and regulatory reasons, to underground tanks used to store and dispense motor fuel for business, commercial and government use. This limitation excludes materials other than motor fuels that may be stored in underground storage tanks, such as chemicals, waste water, hazardous waste, heating oil, and used or waste oil. Also excluded by definition are motor fuel storage tanks that are at private residences, above-ground or partially buried tanks, and all motor fuel tanks at bulk storage facilities that do not dispense fuel to end-users. Tanks that are abandoned or empty were also excluded from consideration. Included within the scope of the survey are tanks that are owned and operated by private businesses, public and government institutions, military facilities, and farms.

As a result, our sample includes such establishments as gasoline stations, airports, marinas, rental car agencies, fleets of trucks or company cars, bus companies, and many other establishments. For practical reasons for list building and screening costs, small establishments (with fewer than 20 employees) in industries not judged to be fuel-related were not included in the survey. Table 4-1 is a list of the industries that were judged to be fuel-related. For these industries all establishments were included in the listing process.

Table 4-1. Selected SIC codes for fuel tank establishments frame

<u>SIC code</u>	<u>Description</u>
4010	Railroads, switching and terminal companies
4110+	Local and suburban passenger transportation companies (includes airport transportation, ambulance and limousine services)
4121+	Taxicab companies
4131+	Intercity highway transportation services
4140+	Passenger transportation charter services (includes bus charter, rentals and tours)
4151	School bus companies
4170	Passenger transportation terminal and service facilities
4210+	Trucking companies
4231+	Motor freight terminals
4469A	Marinas
4511	Air transportation, certificated carriers
4521+	Aircraft charter, rental and leasing -- non-certificated carriers
4582A	Airports
4582B+	Aircraft maintenance services
4583	Airport terminal services
5511+	Auto and truck dealers (new and used)
5521+	Used car dealers
5541+	Gasoline service stations
7512+	Passenger car rental and leasing agencies
7513+	Truck rental and leasing agencies
7519+	Utility and house trailer rental agencies
7992+	Public golf courses
7997B+	Golf and country clubs

II. SAMPLE DESIGN AND SITE SELECTION

A. Sample Design

The contiguous U.S. (forty-eight states plus the District of Columbia) was divided into six survey regions based on broad soil and climatic characteristics. Table 4-2 lists the states which comprise each region. Six Primary Sampling Units (PSUs) were drawn from each region, except for the Mountain region, where four were drawn. The PSUs consist of counties or groups of counties with a minimum count of gas stations and the fuel-related establishments (see Table 4-1). They were sampled within the survey regions on the basis of probability proportional to this count.

Once the 34 survey sites (consisting of 76 counties) were drawn, establishment lists for sampling were constructed for each county. In order to construct the lists, the target universe was divided into three sectors:

1. Fuel-related establishments -- Establishments which by the nature of their business are likely to have underground motor fuel storage tanks. The industries in this category are listed in Table 4-1 and include gas stations, trucking companies, airports, marinas, and others. Government and military establishments with underground motor fuel storage tanks were also part of this sector.
2. Large establishments (20 or more employees) in other industries -- Although the nature of their business would not suggest the presence of underground motor fuel storage tanks; by virtue of their size, these large establishments may have such tanks.
3. Farms were listed and sampled separately -- (See Appendix G for detailed discussion of the farm sample.)

Table 4-2. Six regions for the National Survey of Underground Fuel Storage Tanks

1 -- Northeast

Maine
New Hampshire
Vermont
Connecticut
Massachusetts
Rhode Island
New York
New Jersey
Pennsylvania
Maryland
Delaware
Virginia
West Virginia
Washington, D. C.

2 -- Southeast

Kentucky
Tennessee
Arkansas
Louisiana
Mississippi
Alabama
Georgia
North Carolina
South Carolina
Florida

3 -- Midwest

Wisconsin
Minnesota
Iowa
Missouri
Illinois
Indiana
Ohio
Michigan

4 -- Central

North Dakota
South Dakota
Nebraska
Kansas
Oklahoma
Texas

5 -- Mountain

Montana
Wyoming
Idaho
Nevada
Utah
Colorado
Arizona
New Mexico

6 -- Pacific

Washington
Oregon
California

Separate samples were drawn from the three frames thus established, since they were expected to yield widely varying eligibility rates.

The fuel establishment sample was drawn by region, with equal probability of selection within each region. To reach the target of 800 survey-eligible establishments in this sector, 1,618 establishments were sampled for screening. The large establishment and farm samples were drawn on an equal probability basis nationwide. Six hundred establishments were drawn from each of these frames and the survey-eligible establishments kept.

After the eligible establishments were determined, a sub-sample was drawn for tank tightness testing. All eligible farms were selected for this testing since so few (20) farms were eligible. For the fuel-related and large establishments, the two samples were combined and an equal probability sample drawn from each region. All tanks at sub-sampled establishments were to be tested.

B. PSU (Site) Selection

Once the six survey regions were defined (Table 4-2), a master list of PSUs was developed. For each of the 3,111 counties in the contiguous U.S., several counts were developed. The 1981 County Business Patterns (CBP) data base supplied figures for the number of gas stations, other fuel-related establishments, and establishments in other industries with 20 or more employees. A report prepared by Versar for the EPA (Leaking Underground Storage Tanks Containing Engine Fuels, draft, March 1984) supplied estimates for the number of gas stations on a state-wide basis, based on figures from Petroleum Marketing News (PMN). These counts included all retail outlets for branded

gasoline, i.e., convenience stores and other outlets as well as gas stations. The CBP county totals for gas stations were adjusted upwards to sum to the PMN totals. These adjusted counts were added to the CBP other fuel-related establishment totals to get a fuel establishment count for each county. Minimum PSU counts were established by region (so that a sampled PSU would be sure to have enough establishments to list and sample). Counties with fewer fuel establishments than these minima were grouped together to form multi-county PSUs. The 3,111 counties yielded 1,362 PSUs.

Within each survey region the PSUs were sorted by urban versus rural, then by state and finally by PSU measure of size (count of fuel establishments). Six PSUs were selected from each region (four in Region 5 -- Mountain) with probability proportional to measure of size. The resulting 34 sampled PSUs are made up of 76 counties. Twenty-three PSUs are urban and eleven rural, and together they form a probability sample representing the entire contiguous United States.

III. ESTABLISHMENT FRAME CONSTRUCTION AND SAMPLE

A. Frame Construction

Since lists of establishments with underground motor fuel storage tanks do not exist, it was necessary to create establishment frame lists for each of the 34 PSUs. As described above, the target universe of all establishments with underground motor fuel storage tanks was divided into three segments. The first segment consisted of establishments which, by the nature of their business, were considered fairly likely to have such tanks. This segment, called the "fuel-related establishments" segment,

contained gas stations, trucking companies, bus services, auto dealers, marinas, golf courses, airports and other industry groups that use large amounts of motor fuel or dispense it to the public. (See Table 4-2 for a list of these industries.) Also included in this segment were government and military establishments with underground motor fuel storage tanks.

The second sample segment, the large establishment segment, consisted of establishments in all nonfuel-related industries (i.e., those industries excluded from the first segment) that have 20 or more employees. This segment was designed to provide estimates of the number of large, nonfuel-related establishments that have underground motor fuel storage tanks to service company vehicles and private fleets.

The third sample segment consisted of farms. Recent census of agriculture statistics indicate that about half of the more than two million farms in the United States have on-farm motor fuel storage, but no information existed on how much of this fuel storage was in underground tanks. This segment was designed to provide estimates of the number of farms that have underground motor fuel storage tanks to service farm equipment.

To construct the first list of fuel-related establishments, several sources and methods were used. A listing of all establishments with a primary or secondary Standard Industrial Classification code appearing on the list in Table 4-1 was purchased from National Business Lists. By specifying firms with a fuel-related SIC code as the secondary code, we included such establishments as convenience stores which also sell gasoline. This list was supplemented by adding any establishments with such a code as their primary or secondary code appearing on the large establishments list, purchased from another source. To complete the fuel establishments sampling frame, lists of government

(Federal, State and local) and military establishments with eligible tanks were needed. The Department of Defense provided lists of military tank locations in the sampled PSUs to the EPA. The civilian government list was constructed using a telephone contact and network approach for the government officials serving the sampled PSUs at the local, State and Federal levels. The frames for all 34 PSUs had about 34,000 entries.

The large establishment list was purchased from Dun and Bradstreet. All establishments in the 76 counties with 20 or more employees were purchased from the Dun's Market Indicators list, a very complete business listing. As noted above, all establishments on the purchased list with a fuel-related primary or secondary SIC code were removed from the large establishments frame and clerically compared with the existing fuel establishments frame. If they were not already on that frame they were added to it. About four percent of the final fuel establishments frame came from the Dun and Bradstreet list. All establishments with an agricultural SIC code were also removed from the large establishments frame and added to the farm frame if not already there. In this way, the particular establishments were on the correct frame and duplication between frames was ruled out. The final count for the large establishments frame in the 34 PSUs was about 68,000 establishments.

The farm frame was provided to the EPA by the U.S. Department of Agriculture (USDA). As noted, it was supplemented by the (very few) farm establishments found on the purchased Dun and Bradstreet list of large establishments. About 31,000 farm owners and operators were listed on the farm frame for the 34 PSUs.

B. Establishment Sample Draw

The fuel establishment sample was drawn by survey region. The total sample size of 800 eligibles was allocated among the six regions in proportion to their count of fuel-related establishments in the 1981 County Business Patterns data. Based on initial field results of 50 percent eligibility, the target sample sizes were approximately doubled. Within each region an equal probability sample was drawn. Table 4-3 gives the counts of sampled cases by region.

The large establishments and farms were both sampled on an equal probability basis nationwide. Six hundred of each were sampled, with only the eligibles remaining in the survey. Because so little was known regarding incidence of underground motor fuel storage tanks in these sectors, the initial sample size was fixed rather than the final number of eligibles being fixed.

Table 4-4 shows the results of screening the initial sample. Eight hundred fuel establishments and 76 large establishments were eligible for the survey (in business and operating an underground motor fuel storage tank). Of these, 871 provided questionnaire data. In addition, 20 of the 600 sampled farms had underground storage tanks. As indicated in Appendix G, about half of all farms report motor fuel storage, but only about 10 percent have more than 1,000 gallons of storage capacity. For these small amounts of fuel, above ground storage is often a reasonable alternative.

Table 4-3. Initial sample sizes for fuel establishment, large establishment and farm samples by survey region

Survey region	Fuel establishments	Large establishments	Farms
1 Northeast	449	158	11
2 Southeast	415	116	88
3 Midwest	325	142	324
4 Central	194	68	142
5 Mountain	75	29	33
6 Pacific	160	87	2
Total	1,618¹	600	600

¹Subsequent fieldwork determined that six of the sampled fuel establishments were duplicates.

Table 4-4. Number of eligible cases for fuel establishments, large establishments, and farm samples by survey region

Survey region	Fuel establishments	Large establishments	Farms
1 Northeast	225	21	0
2 Southeast	197	18	3
3 Midwest	161	13	5
4 Central	92	7	5
5 Mountain	42	4	4
6 Pacific	83	13	0
Total	800 ¹	76	20 ¹

¹Five fuel establishments and one farm refused at the interview phase of the survey.

IV. SUB-SAMPLE OF ELIGIBLE ESTABLISHMENTS FOR PHYSICAL TANK TIGHTNESS TESTS

Since so few farms screened had underground motor fuel storage tanks, it was decided to physically test all such tanks at all eligible farms. At the time of sample allocation, it was estimated that there would be at most 50 tanks at eligible farms, so that number was set aside for farm tank tests.

This left a target number of 450 tanks or manifolded tank systems to be tested in the business and government sector (fuel and large establishments). The 450 were allocated to the six survey regions in the same proportions as the original establishment sample allocation, except that a minimum number, 40, were allocated to Region 5, the smallest region, before allocating the remainder to the other five regions. As each region was completed by the interviewers, a list of eligible government, fuel-related establishments and large establishments in the questionnaire sample was constructed, with the number of tanks or manifold tank systems for each establishment listed. At the time of sub-sampling it was assumed that a manifolded tank system (two or more tanks connected by various lines and pipes) would be physically tested as one unit. Therefore the sub-sample was drawn on that basis. During the actual testing, some such systems were isolated, and the individual tanks (and associated lines) were tested separately. Thus, the total number of possible tank tests is more than the number of tanks or tank systems reported here but less than the total number of tanks at these establishments.

The sub-sample of tanks to be tested was drawn on an establishment basis, with all tanks at a given establishment tested. The establishment list for a given region was sorted by number of tanks or tank systems, then PSU, and then fuel-related

and government versus large establishment. The target number of establishments to select was calculated from the list, which included initial sampling weights, and the target tank sample size, using the weighted average number of tanks per establishment. An equal probability sub-sample of establishments was then drawn from the list. Table 4-5 shows the target number of tank tests and the number of establishments sub-sampled with the number of tanks or tank systems at the sub-sampled establishments.

V. CALCULATION OF FINAL SAMPLE WEIGHTS AND VARIANCE ESTIMATION

A. Calculation of Final Sample Weights

1. Questionnaire Weights for Business and Government Establishments

The final questionnaire weights for establishments sampled with fuel-related SICs other than gas stations were based on a ratio adjustment of the initial sample weights for all such screened establishments to 1982 County Business Patterns (CBP) counts of these SICs followed by a nonresponse adjustment among the eligible other fuel-related establishments to account for the few nonrespondents. (By the time final weights were being calculated, the 1982 data were available.) The adjustments were made by survey region. The ratio adjustment served to calibrate the initial sample to CBP estimates of the number of establishments with one of the fuel-related SICs in each region. The sum of the weights of the eligible cases is the survey estimate of the number of such establishments with eligible tanks, by region. The nonresponse adjustment assures that the

Table 4-5. Summary of business and government establishment subsample¹ for tank tightness testing, by region

Region	Target number of tank systems ² to subsample for business and government sectors	Number of business and government establishments subsampled	Number of business and government tank systems ¹ at subsampled establishments
1 Northeast	115	51	112
2 Southeast	110	47	111
3 Midwest	90	38	86
4 Central	50	23	52
5 Mountain	40	17	43
6 Pacific	45	22	46
Total	450	198	450

¹All eligible farm underground motor fuel storage tanks were assigned for tightness testing. There were 20 eligible farms with 35 tanks.

²In allocating and drawing the subsample of establishments for tightness testing, a manifold tank system was counted as one unit. Some such systems were separated for physical testing.

weighted results based on questionnaires received equals the estimates based on screening results.

The gas stations were weighted in the same way. First the initial sample was ratio-adjusted by region to CBP totals for gas stations (SIC code 5541). The sum of the weights of eligible cases then estimates the number of gas stations with eligible tanks, by region. A nonresponse adjustment again assures that the weighted results based on questionnaires received will equal the estimates based on screening.

The sample sector of establishments with 20 or more employees in industries not otherwise sampled (the large establishments) was weighted the same way as the gas stations and other fuel-related industries. The CBP totals of establishments of this size in all but the selected fuel-related SICs were used for a region-by-region ratio adjustment of the initial sample. The weighted eligible large establishments then estimate the number of such establishments with eligible tanks in the country, by region. Since all eligible large establishments participated in the interview phase of the survey, no nonresponse adjustment was needed.

No national statistics are currently available to estimate the number of individual government agencies with underground motor fuel storage tanks, which is the universe our frame was built to cover. Therefore no ratio adjustments can be made. Nonresponse adjustments were made to account for the small amount of nonresponse among government establishments.

2. Physical Test Result Weights, Business and Government Establishments

After calculating final questionnaire weights for all responding business and government establishments as described above, the sampling weights for establishments chosen for physical testing were adjusted to sum to the estimated totals for four establishment types (government, gas station, other fuel-related, and other industry) by region. This adjustment was made by an iterative raking procedure, in which the weights were adjusted first to regional totals, then to establishment-type totals, then readjusted to regional totals, and so forth, until no further adjustment was needed. (This took five and a half iterations to achieve.)

A final adjustment was made for tank test result weights. The weight for the individual tank or tank system test would be equal to the establishment physical test weight, except that some tanks were not tested. Thus, a "tank nonresponse" adjustment was made to the tank weights to account for the untested tanks.

3. Farm Questionnaire and Physical Test Weights

Due to the distribution of farms within the survey regions (both overall and in our sample) and the low yield of eligible farms from the screening, the survey regions have been consolidated into three areas for calculating final weights for farms. (See Appendix G for a more detailed discussion.) These are: (1) East (combines survey regions Northeast and Southeast); (2) Midwest; and (3) West (combines survey regions Central, Mountain, and Pacific). Total counts of farms for these areas were obtained from the 1982 Census of Agriculture and used to

form ratio adjustments for eligible farms. Due to one refusal among farms, a nonresponse adjustment was also made.

Since so few farm tanks were tightness tested (21 of 35 -- most not tested were smaller than 1,100 gallons), no weighted estimates will be presented for that data, and hence no final weights were calculated for physical test results for farm tanks.

B. Variance Estimation

National estimates from the survey are based on a sample of cases rather than a complete census of the nation's underground motor fuel storage tanks, so they are subject to variability termed sampling error. This is due to the fact that drawing several samples would result in different sets of establishments being interviewed and different national estimates. Since the sample was drawn on a probability basis, it is possible to use the survey data to estimate the magnitude of this sampling error. Due to the complex nature of the sample design, this variance is not easily expressed as a simple mathematical formula. It has therefore been estimated by a more empirical approach.

The method of variance estimation used in this survey is termed the jackknife approach. Essentially, a series of subsamples of the survey data known as replicates are created. Using the same series of steps given above for the full sample, each replicate is given weights which can be used to create national estimates based on that replicate. The variance of the replicate estimates of the statistic from the full sample estimate estimates the sampling error of the statistic.

In this report the sampling error is generally reported in terms of 95 percent confidence bounds. These are interpreted as

being the numeric range which one can be 95 percent confident includes the true value of the statistic. It is centered on the full sample estimate and its width is determined by the estimated sampling error of the statistic.