



Independent Statistics & Analysis

U.S. Energy Information
Administration

Supporting Statement for Petroleum Marketing Program

Part B: Collections of Information Employing Statistical Methods

Form EIA-14, Refiners' Monthly Cost Report

Form EIA-182, Domestic Crude Oil First Purchase Report

Form EIA-782A, Refiners'/Gas Plant Operators' Monthly Petroleum Product Sales Report

Form EIA-782C, Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption

Form EIA-821, Annual Fuel Oil and Kerosene Sales Report

Form EIA-856, Monthly Foreign Crude Oil Acquisition Report

Form EIA-863, Petroleum Product Sales Identification Survey

Form EIA-877, Winter Heating Fuels Telephone Survey

Form EIA-878, Motor Gasoline Price Survey

Form EIA-888, On-Highway Diesel Fuel Price Survey

OMB No. 1905-0174

July 2017



Independent Statistics & Analysis

www.eia.gov

U.S. Department of Energy

Washington, DC 20585

Table of Contents

Part B: Collections of Information Employing Statistical Methods.....	i
B.1. Respondent Universe.....	1
B.2. Statistical Methods.....	4
B.3. Maximizing Response Rates.....	16
B.4. Test Procedures and Form Consultations.....	16
B.5. Statistical Consultations.....	17

B.1. Respondent Universe

The Petroleum Marketing Program collects data from two types of entities in general. In terms of the market chain, the data system begins by collecting data from firms, generally parent companies that have complex structures with multiple offices, locations, subsidiaries, etc. The target population for the surveys is the firms, which are defined in terms of their oil market activities. This includes firms which purchase domestic or foreign crude oil (Forms EIA-182 and EIA-856); firms which refine crude oil into finished petroleum products (Form EIA-14); and firms which supply and/or sell finished petroleum products to customers (Forms EIA-782A, EIA-782C, and EIA-821). The remaining surveys focus on individual outlets selling the products to consumers (Forms EIA-877, EIA-878, and EIA-888) and collect price data on a weekly basis. The Petroleum Product Sales Identification Survey (Form EIA-863) is used to build and maintain the frame for several of these surveys. The frames for the monthly surveys are kept current using information from other surveys as well as information from industry journals and other sources.

B.1.1 Petroleum Marketing Frame

Frame building for Form EIA-863, “Petroleum Product Sales Identification Survey,” begins with the frame constructed from the previous cycle, relevant data records from an internal frame system maintained by the Office of Petroleum and Biofuels Statistics (PBS), and outside source lists of petroleum marketers from State Energy Offices and Petroleum Marketing Associations. The target population for Form EIA-863 is estimated over 20,000 companies.

B.1.2 Monthly Crude Oil Surveys Frames and Target Population

Form EIA-14, “Refiners' Monthly Cost Report”

The target population for this survey is all refiners of crude oil. The frame for Form EIA-14 was constructed from a list of 206 refiners obtained from the Oil and Gas Journal in 1983. The frame is updated periodically via information derived from Form EIA-782A, “Refiners’/Gas Plant Operators’ Monthly Petroleum Product Sales Report,” and Form EIA-810, “Monthly Refinery Report.” There are currently 69 active respondents filing Form EIA-14.

Form EIA-182, “Domestic Crude Oil First Purchase Report”

The target population for this survey is all firms that buy domestic crude oil at the lease boundary, acquiring ownership of the crude in a first purchase transaction. The frame for Form EIA-182 was initially

compiled from the 1974 Federal Energy Administration (FEA) Oil and Gas Survey of Producers and Operators. Collection of data from first purchasers began in February 1976. By 1978, the frame consisted of 340 respondents. Of these, 198 purchased more than 150,000 barrels per year and together represented 99.9 percent of the total reported volume. Following decontrol in January 1981, many small firms went out of business or were absorbed by larger companies. By January 1986 the frame had been reduced to 170 respondents. Over the years, adjustments to the frame have mostly been deaths, with relatively few births. The size of the frame has declined from 340 firms in 1978 to 106 firms in 2016.

Form EIA-856, “Monthly Foreign Crude Oil Acquisition Report”

All companies acquiring more than 500,000 barrels of foreign crude oil in the report month for importation into the United States are required to submit this form monthly. The frame for Form EIA-856 is composed of all companies reporting data in June 1982 on the Transfer Pricing Report (Form ERA-51) regardless of the total volumes of crude oil that were imported and all companies acquiring more than 500,000 barrels of oil in the reference month for importation into the United States. Recently it was determined data from the latter is only required. Currently the frame consists of 42 respondents.

B.1.3 Monthly and Annual Petroleum Product Frames and Target Populations

The target population for Form EIA-863 is all firms which sell petroleum products. The firms surveyed on Form EIA-863, along with their associated volumetric data and tracking information, serve solely or partially as the sampling frame for Forms EIA-821, “Annual Fuel Oil and Kerosene Sales Report,” EIA-877, “Winter Heating Fuels Telephone Survey,” EIA-878, “Weekly Motor Gasoline Price Survey,” EIA-888, “Weekly On-Highway Diesel Price Survey,” and in the past, other ad hoc surveys, such as the National Petroleum Council Surveys.

Form EIA-821, “Annual Fuel Oil and Kerosene Sales Report”

The frame for this survey is constructed from Form EIA-863 survey results, supplemented by retailers/resellers and importers of residual fuel oil who were not identified by Form EIA-863. Currently, the sampling frame consists of over 20,000 companies.

Form EIA-782A, “Refiners’ Gas Plant Operators’ Monthly Petroleum Product Sales Report”

The target population for this survey includes the universe of refiners and gas plant operators. The original frame was derived from a consolidated list of refiners known to have reported on several EIA surveys; and the frame of gas plant operators from Form EIA-64, “Natural Gas Liquids Operations Report.” In 2016 the frame consisted of 100 companies.

Form EIA-782C, “Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption”

The target population for this survey includes all suppliers who make the first sale of any of the products listed on Form EIA-782C, and deliver that product into a state for consumption in that state. The product slate includes: motor gasoline, No. 1 distillate, kerosene, fuel oil, diesel fuel, aviation gasoline, jet fuel, No. 4 fuel, residual fuel oil, and propane. The original frame was constructed from the respondent frame of the former Form EIA-25, “Prime Supplier’s Monthly Report.” Currently the frame consists of 200 prime suppliers and is updated as needed due to births, deaths and mergers.

B.1.4. Weekly Petroleum Product Frames and Target Populations

Form EIA-877, “Winter Heating Fuels Telephone Survey”

The company/state level frame for this survey is constructed from company state units identified in Form EIA-863. The current sampling frame for units which sell either residential heating oil or propane consists of 16,000 Company State Units (CSUs) in 38 states. For residential propane sellers, an outlet level frame is used for the second phase selection of individual outlets to represent the companies selected in the first phase. There are approximately 6,500 active propane outlets in the 38 states.

Form EIA-878, “Motor Gasoline Price Survey”

The sample for Form EIA-878 was drawn from a frame of approximately 130,000 retail gasoline outlets. The gasoline outlet frame was constructed by combining outlet level name and address information purchased from the Oil Price Information Service (OPIS) with information from various internet sources including the Yellow Pages. The individual frame outlets were mapped to zip codes and counties. The outlets were then assigned to the sampling regions. These regions are non-overlapping and one or more sampling regions will comprise a publication region – reformulated gasoline areas as defined by the U.S. Environmental Protection Agency (EPA) program area, or for conventional gasoline areas.

Form EIA-888, “On-Highway Diesel Fuel Price Survey”

The outlet sampling frame was constructed using commercially available lists from several sources in order to provide comprehensive coverage of truck stops and service stations that sell on-highway diesel fuel in the United States. The frame includes about 62,000 service stations and 4,000 truck stops.

B.2. Statistical Methods

Six of the ten petroleum marketing surveys— Forms EIA-14, EIA-182, EIA-782A, EIA-782C, EIA-856, and EIA-863—are census surveys so no weighting is required. The target population for Form EIA-856 is all companies with over 500,000 barrels of foreign crude oil in the report month for importation into the United States. One respondent with less than 500,000 barrels is required to report because they were originally on the frame in 1982. The volume weighted price is used for estimation in four of these surveys. The total cost or revenue (price times volume) is divided by a corresponding total volume to arrive at a volume weighted average cost.

For the remaining surveys, sample weights are typically calculated as the inverse of the probability of selection of the sampled company or outlet. The price surveys create volume-weights for use in estimation, which are the product of the sampling weight times a sales volume measure for the outlet.

B.2.1 Statistical Methods for Monthly Crude Oil Surveys

- 1) Form EIA-14 is a census. A volume weighted price at the national and regional levels is calculated by dividing the total cost (price times volume) by the corresponding total volume.
- 2) Form EIA-182 is a census. To obtain a national weighted average price, the total cost (amount paid times volume purchased) is divided by corresponding total volume. Subsequently, the data are sorted by crude stream within each state. These data are aggregated across all companies reporting purchases from a given state. Weighted average prices for crude oil are then derived for each producing state and for the Outer Continental Shelf regions, Alaska North Slope and Alaska Other.
- 3) Form EIA-856 is a census of those companies importing over 500,000 barrels of foreign crude oil into the United States in the report month.

B.2.2 Statistical Methods for Monthly Petroleum Product Surveys

There are two monthly petroleum product surveys and both are a census since the target population for each of these surveys is small. The volume weighted price is used for estimation in one of these surveys. The total revenue (price times volume) is divided by a corresponding total volume to arrive at a volume weighted average selling price.

- 1) Form EIA-782A Aggregation: Data from this survey is used to estimate the national average price for each product and market level - e.g., U.S., Petroleum Administration for Defense Districts

(PADD), and state. The price and volume data for each company are multiplied and then aggregated across all companies for each product and market level to obtain a total revenue figure. This revenue is then divided by the corresponding total volume to arrive at a volume weighted average selling price.

- 2) Form EIA-782C Aggregation: Because Form EIA-782C is a census survey and only volume totals are published, the only estimation procedures used are for summing across companies.

B.2.3 Statistical Methods for Weekly Petroleum Product Surveys

All three of the following surveys collect weekly prices for petroleum products – motor gasoline, on-highway diesel, heating fuel, and propane. Table B1 contains a summary of the sample design for the weekly surveys (Forms EIA-877, EIA-878, and EIA-888) in the Petroleum Marketing Program.

Table B1: Summary of Sample Design			
Survey Name	Sample Design	Selection Procedure	Sample Size
EIA-877, “Winter Heating Fuels Telephone Survey”	Heating oil: Stratified random sample with certainty for 21 states and DC in the East Coast and Midwest Regions.	See section 1a	527
EIA-877, “Winter Heating Fuels Telephone Survey”	Propane: Two-stage sample for 38 states, which includes the states in the East Coast, Midwest, Gulf Coast, and Rocky Mountain Regions. Stratified Probability Proportional to Size (PPS) Sequential Sample with Probability Minimum Replacement of Company State Units in the first stage and simple random sample of outlets in the second stage.	See section 1b	1,140
EIA-878, “Motor Gasoline Price Survey”	Stratified systematic simple random sample.	See section 2	1,000
EIA-888, “On-Highway Diesel Fuel Price Survey”	Stratified Probability Proportional to Size sample of retail outlets from 48 continental states and DC.	See section 3	403

- 1) **EIA-877 Sample and Estimation:** Separate sampling and estimation procedures are used for residential No. 2 heating oil and propane. Descriptions of these procedures follow.

1a) EIA-877 Sample and Estimation Procedures for No. 2 Heating Oil Prices: The No. 2 heating oil price data are reported by a statistical sample. According to the requirement of the State Heating Oil and Propane Program (SHOPP) program, 21 states and the District of Columbia (DC) in the East Coast and Midwest regions participate in the No. 2 heating oil price survey. The sampling frame used was a list of all Company State Units (CSUs) in

those 21 states and DC that reported residential No. 2 heating oil sales on the 2006 Form EIA-863, "Petroleum Product Sales Identification Survey."

CSUs that sold at least five percent (5%) of the residential No. 2 heating oil in a state, as reported on the frame survey EIA-863, were automatically included in the sample and are referred to as certainty units. The remaining CSUs, referred to as non-certainty units, were stratified into three groups by their residential No. 2 heating oil sales volumes in each state, as reported on the EIA-863. Strata boundaries were determined using the Dalenius-Hodges procedure. The sample allocations were designed generally to yield average price coefficients of variation (CV) of one percent (1%), but, due to budget constraints, individual state sample sizes were capped at 35 even if the target CV was not met. In addition, a minimum size of fifteen was required for each of the 21 states. The sample of CSUs within each stratum was a simple random sample. The residential No. 2 heating oil sample size inclusive of certainty and non-certainty units is 527 CSUs.

The residential No. 2 heating oil prices (excluding taxes) for a given state are based on the results of a telephone survey of marketers and refiners, one for each of the two products. Data are collected by State Energy Offices under the U.S. Energy Information Administration (EIA) State Heating Oil and Propane Program (SHOPP).

To estimate the average residential No. 2 heating oil price data for a state, the sample and volume weights are applied to the reported price, summed and divided by the sum of the weighted volume:

$$\frac{\sum_{j=1}^s \sum_{i=1}^{n_j} w_{ij} v_{ij} p_{ij}}{\sum_{j=1}^s \sum_{i=1}^{n_j} w_{ij} v_{ij}}$$

where w_{ij} = sample weight

v_{ij} - volume weight

p_{ij} = price,

i = respondent

n_j = sample size of stratum j , and

s = number of strata

The sample weights W_{ij} were calculated as ratios of population number of CSUs to the sampled number of CSUs in each stratum. Volume weights V_{ij} were assigned using the data reported in the frame survey. State level residential No. 2 heating oil average prices are then aggregated into regional and overall averages with state level total residential No. 2 heating oil volumes as weights.

These fixed volume weights indicate the relative importance of the individual companies according to the size of their sales at the time of the frame. Therefore, changes in the average price across time reflect only the change in the price being offered by the company, and not the change in the amounts sold. Price indexes constructed using fixed volumes, such as these annual sales, are known as Laspeyres Indexes. One alternative method of weighting, used in Paasche Indexes, uses current weights. This method would require each company to report the number of gallons sold at the reported price each pricing period and would be more burdensome on the companies. Both methods of weighting are correct but provide different averages particularly when volumes are changing. It has been argued in the literature that during periods of change, the Laspeyres method has a tendency to overestimate price changes, while the Paasche method tends to underestimate price changes.

In this survey, it is expected that the weekly change in volumes during the heating season is small. Residential sales are not bulk in nature and do not tend to reflect discounts on price for large volume purchases. Absolute changes in volume with a year's time would more likely reflect demand and be consistent across companies within a geographic area. Therefore, even though the volume weights used in the calculation of average prices in the SHOPP tend to lag behind the actual volumes sold in the reference period, fixed volumes are used to reduce company burden and enable timely release of average prices. EIA seeks to collect information on heating oil volumes in order to periodically calibrate the sampling weights for the volumes used to calculate the sample weights. Moreover, this collection will be consistent with how EIA uses the reported propane volumes in the

same survey form, as described in the next subsection. Outlet level heating oil volumes will be collected to inform sample design.

- 1b) EIA-877 Sample and Estimation Procedures for Propane Prices:** Propane price data are collected from a sample of companies from a total of 38 states. The sampling frame used was a list of all Company State Units (CSUs) in those 38 states that reported residential propane sales on the 2006 and 2010 Form EIA-863, "Petroleum Product Sales Identification Survey." The population of the survey was first stratified by state, which is the publication cell. Due to high residential propane price variation and budget constraints, sample sizes of all strata were limited to 30 even though the target CV of one percent (1%) was not met in many states. To select the sample, the CSUs in each state were ordered by zip code in order to control for the geographic location of the companies. A Probability Proportional to Size (PPS) Sequential Sample with Probability Minimum Replacement, using the propane volumes reported on the sample frame of CSUs from the EIA-863 as a measure of sampling unit size, was then selected. With the targeted maximum sample size of 30 CSUs in each state, any CSU that sold more than 3.3 percent (1/30) of the residential propane in a state was selected at least once.

Within each sampled CSU, a simple random sample of residential propane outlets was drawn by using an outlet address listing EIA developed with information provided by the industry and state energy officials. The number of outlets selected from each CSU was the same as the number of times that CSU was selected in the first stage PPS sample. The resulting total number of outlets selected to report on the propane sample was 1,140.

In cases where there were fewer outlets in a sampled CSU than the number of times that CSU was sampled in the PPS sample, all outlets for that CSU were selected and their weights were adjusted to n'_i/n_i , where n'_i is the number of times that CSU was sampled and n_i is the number of outlets sampled. This was also the practice if a CSU preferred to report on the survey by providing the average of all its outlets in a given state. Therefore, the actual number responding each week may deviate from the 1,140 outlets sampled.

To estimate the average residential propane price data for a state, a simple average of the prices from each sampled outlet yields a valid estimate as a result of the cancellation of sample weights of the PPS sample design and volume weights in the estimate:

$$\frac{1}{n} \sum_{i=1}^k \frac{n_i'}{n_i} p_i$$

where p_i = price,

i = outlet respondent,

k = the number of outlets sampled in a state ($k \leq n$)

n = sample size, and

n_i'/n_i = weight adjustment as described in the previous

paragraph

Regional and overall averages are calculated as averages of the state level residential propane prices with state level total residential propane volumes as weights. Outlet level propane volumes can be collected to inform a better sample design.

2) Form EIA-878 Sample Design: The sample for the Motor Gasoline Price Survey was drawn from a frame of approximately 130,000 retail gasoline outlets. The gasoline outlet frame was constructed by combining information purchased from a private commercial source with information available on the Internet. Outlet names and zip codes were obtained from the same private commercial data source. Additional information was obtained directly from companies selling retail gasoline to supplement information on the frame. The individual frame outlets were mapped to zip codes and counties. The outlets were then assigned to the sampling geographic regions. These regions are non-overlapping and one or more sampling region will comprise a publication region – reformulated gas areas as defined by the EPA program area, or for conventional gasoline areas.

The sampling regions will be further stratified by type of outlet – hypermarkets and non-hypermarkets. This results in 76 (=38x2) strata for sampling. Since hypermarkets sell about 2.5 times more volume of gasoline than non-hypermarkets, we will allocate a sample size to hypermarkets at 2.5 times their number as compared to the other outlets. Also, the estimation

formula for volume weighted price shown below needs a slight modification – there will be two (adjusted) sampling weights per sampling region, one for hypermarkets and another for non-hypermarkets.

Sampling within each region and type of outlet will be done by first listing the counties alphabetically and then the zip code within the counties listed in numerically increasing order. Outlets within a given zip code will be listed alphabetically and serialized from one to the total number of outlets in the sampling region. The sampling will be done using systematic random sampling without replacement.

After the initial sample of outlets has been selected, available sources of information on new outlets since the sampling frame was constructed will be analyzed each year until the sample is redesigned, with the goal of selecting an independent birth sample annually to augment the initial sample. In designing these annual birth samples, sampling geographic regions will be oversampled where there are relatively higher rates of sample attrition because of outlets that were selected in the initial sample and subsequently went out of business.

Form EIA-878 Estimation: In the first phase of this survey, a measure of size – annual volumetric sales by grade of gasoline- will be collected from about 1,000 selected outlet owners. These volumes will be applied each week to the reported outlet gasoline prices to get volume weighted average prices and measures of uncertainty. Since we need volumetric data for all the sampled outlets, imputation for nonresponse for these volumes will be done at the outlet level, by substituting the average reported volume by imputation cell - PADD/SubPADD and grade (regular, midgrade, and premium). Also, we will do nonresponse imputation separately for hypermarkets and non-hypermarkets. Small area estimation will be used for imputation by collapsing over adjacent PADD/subPADD if two or less responses fall into an imputation cell.

For the estimation of average prices and variances, the following notation is used:

- j sampling region
- i outlet
- N population size (number of outlets)
- n_j sample size in sampling region j

m_j	number of respondents in sampling region j
P	price
V	volume
W	sampling weights = inverses of probability of selection
W'	adjusted sampling weight due to nonresponse

By definition,

$$W_{ji} = \frac{N_j}{n_j}, \quad W'_{ji} = \frac{n_j}{m_j} W_{ji}$$

Define $x_{ji} = v'_{ji} = W'_{ji} V_{ji}$

$$y_{ji} = R'_{ji} = V'_{ji} P_{ji}$$

$$x_j = \sum_i x_{ji}$$

$$y_j = \sum_i y_{ji}$$

Then the volume weighted average price for a region is as follows:

$$(1) \hat{P}_j = \frac{y_j}{x_j}$$

The relvariance of \hat{P}_j is as follows:

$$(2) V^2_{P_j} = V^2_{x_j} + V^2_{y_j} - 2 V_{x_j y_j}, \text{ where}$$

$$V^2_{x_j} = \frac{\sigma^2_{x_j}}{x_j^2} \text{ and } V_{x_j y_j} = \frac{\sigma_{x_j y_j}}{(x_j)(y_j)}$$

The estimated variance of x_j is as follows (The variance of y_j is defined similarly):

$$(3) \hat{\sigma}^2_{x_j} = \frac{m_j}{m_j - 1} \sum_i (x_{ji} - \bar{x}_j)^2,$$

$$\bar{x}_j = \frac{\sum_i x_{ji}}{m_j}$$

The estimated covariance is as follows:

$$(4) \hat{\sigma}_{x_j y_j} = \frac{m_j}{m_j - 1} \sum_i (x_{ji} - \bar{x}_j)(y_{ji} - \bar{y}_j),$$

$$\bar{y}_j = \frac{\sum_i y_{ji}}{m_j}$$

The volume weighted average price for a publication region, which is made up of one or more sampling regions j (and if the sampling regions are further stratified by hypermarket and non-hypermarket then j is over sampling regions with hypermarkets and sampling regions w/o hypermarkets – thus doubling the range of j), is as follows:

$$(5) \quad \hat{P}_{\square} = \frac{\sum_j y_j}{\sum_j x_j}$$

The estimated relvariance of p is given as follows:

$$(6) \quad \hat{V}_p^2 = \frac{\sum_j \hat{\sigma}_{x_j}^2}{\left(\sum_j x_j\right)^2} + \frac{\sum_j \hat{\sigma}_{y_j}^2}{\left(\sum_j y_j\right)^2} + \frac{2 \sum_j \hat{\sigma}_{x_j y_j}^2}{\left(\sum_j y_j\right)\left(\sum_j x_j\right)}$$

$$(7) \quad \text{RSE}(P) = (\hat{V}_p^2)^{(1/2)}$$

A sample size of approximately 1,000 outlets was selected because it is only a modest increase in sample size from the earlier collection and we expect that a larger sample size is needed because of the higher variances using annual volumes as weights versus number of pumps. Once volume data are collected, the variance formulae will determine whether this sample size is sufficient. Parallel testing for three to four weeks will be conducted during sample implementation.

The target coefficient of variation was set at 0.4 for the United States, 0.55 for PADDs and U.S. formulations, 0.70 for sub-PADDs and the PADD formulations, 0.85 for cities and states, and 1.0 for the remaining cells – i.e., state and sub-PADD formulations. The sample size is approximately 1,000 outlets. The survey is conducted every Monday (Tuesday on Federal holidays), and more frequently during emergency situations. Data are released on [EIA's website](#) around 5 p.m. each Monday (Tuesday on Federal holidays). Data are made available through email notification to those customers who sign up for that service. The U.S., PADD, sub-PADD, state, and city levels regular gasoline average prices are made available on EIA's prerecorded telephone hotline at 202-586-6966 and in the publications [Weekly Petroleum Status Report](#) and [This Week in Petroleum \(TWIP\)](#).

3) **Form EIA-888 Sample Design:** The respondents reporting to the weekly diesel price survey represent a stratified probability proportional to size (PPS) sample selected from a frame list of retail outlets. The outlet sampling frame was constructed using commercially available lists from several sources in order to provide comprehensive coverage of truck stops and service stations that sell on-highway diesel fuel in the United States. The frame includes about 62,000 service stations and 4,000 truck stops. Due to statistical and operational considerations, outlets in the States of Alaska and Hawaii are excluded from the target population.

The primary publication cells of the survey include PADDs 2-4, three sub-PADDs within PADD 1, and the two subparts of PADD 5 (the State of California and the West Coast region excluding California). The U.S., the East Coast (PADD 1), and the West Coast (PADD 5) are considered secondary publication cells since their prices are aggregated based on the prices from their primary publication cell components. To select the sample, allocations were first assigned to all primary publication cells through a simulation of coefficients of variation of average prices using historical price data. The target coefficient of variation for each primary publication cell was capped at one percent (1%). Allocations were further assigned to the states covered by each primary publication cell. The distribution of allocations was proportional to the annual state total volume of retail on-highway diesel fuel sales. This allocation procedure yielded a total target sample size of 403 retail outlets. The states were treated as sampling strata in the sample design.

It is believed that on-highway diesel fuel sold through all service stations combined only accounts for a small portion of the retail on-highway diesel fuel market. The truck stops on the frame were also classified into two categories, dependent on whether they belong to the nation's four largest on-highway diesel sellers. Based on information from other survey data and industry sources the proportions of total diesel volumes sold by outlets in the three categories - service stations, mid-sized truck stops, and top four - was assumed to be 20, 55, and 25 percent, respectively. These volume proportions, along with the outlet counts for the three categories on the frame, were used to calculate relative size measures for the outlets in each of the three categories. Pareto Sampling, which is a PPS procedure, and the size measures for each outlet were then used to select sampling units from each state.

The survey is conducted every Monday (Tuesday on Federal holidays), and data are released on [EIA's website](#) around 5 p.m. each Monday (Tuesday on Federal holidays). Data are made available

through email notification to those customers who sign up for that service. The U.S., PADD, sub-PADD, and the State of California levels retail on-highway diesel average prices are made available on EIA's prerecorded telephone hotline at 202-586-6966 and in the publications [Weekly Petroleum Status Report](#) and [This Week in Petroleum \(TWIP\)](#).

Form EIA-888 Estimation: The reported and imputed prices each week are aggregated in multiple steps to obtain price estimates for publication cells. First, state average prices are calculated as simple unweighted averages of reported and imputed prices. Volumes of on-highway diesel sold in the states, as published by the Federal Highway Administration (FHWA), are then used to weight the state average prices and obtain average prices for primary publication cells. These weights are updated annually when FHWA data are available. Average prices for secondary publication cells are weighted averages of primary publication cell prices based on the proportion of diesel volumes attributable to their component primary publication cells.

B.2.4 Statistical Methods for Annual Petroleum Product Survey

Form EIA-821 Sample Design: The target population for Form EIA-821, "Annual Fuel Oil and Kerosene Sales Report" survey is the universe of all active companies that sell distillate fuel oil, residual fuel oil, or kerosene in the 50 states and the District of Columbia.

The 2006 Form EIA-863 database provided a base sampling frame for Form EIA-821 survey. Form EIA-863 was mailed to approximately 24,000 companies in January 2007 to collect 2006 state-level sales volume data for No. 2 distillate fuel, residual fuel oil, motor gasoline, and propane. Companies also indicated if they sold kerosene. The No. 2 distillate fuel data were further identified by residential No. 2 fuel oil and by nonresidential retail and wholesale for No. 2 fuel oil and No. 2 diesel fuel; the residual fuel oil data were identified by retail and wholesale. In addition, company/state-level volumes for distillate fuel, residual fuel oil, and kerosene from the 2008 Form EIA-821, 2008 Form EIA-782A, and 2008 Form EIA-782B were also merged with the 2006 Form EIA-863 data. The integrated and comprehensive frame was then used to design and select the 2009 Form EIA-821 sample, which the survey is based.

To select a sample for Form EIA-821, subsidiaries and parents of a company were merged by adding the volumes of parents and subsidiaries in a cluster (i.e., parent-subsidiary combination) to represent the company. The sample was drawn from a multi-attribute frame with four target variables of No. 2

residential fuel oil, No. 2 nonresidential fuel oil, No. 2 nonresidential diesel fuel, and No. 2 wholesale distillate fuel.

A company was classified as a certainty company if it met one of the following criteria:

- The company (or one of its subsidiaries) was a refiner as identified in the 2008 Form EIA-782A survey.
- The company had residual fuel oil sales.
- The company sold any Form EIA-821 product in at least five states.
- The sum of maximum percentage of the four distillate products at the state level across states was five percent or more.
- The company reported over five percent of the total weighted volume in any state for any specified product by end-use category in the 2008 Form EIA-821 survey.

A systematic probability proportional to size design (PPS) was used to sample noncertainty companies. Company State Units (CSUs) were the sampling units. A CSU selected by the sampling procedure was referred as a “basic” CSU. A company was included in the sample if it had at least one “basic” CSU. All non-“basic” CSUs of a sampled company were referred as “volunteer” CSUs.

In each state, the Dalenius-Hodges procedure was used to stratify CSUs, with each of the four target distillate variables, into zero, low, medium, and high volume four strata. Neyman allocation was used to obtain the sample size for each stratum to meet target coefficient of variation of five percent. The population of CSUs was divided into mutually exclusive cells by crossing the four stratifications such that every CSU in a particular cell was in the same stratum for each of the four stratifications. Each CSU was assigned a probability of selection, which was the largest sample proportion across all four stratifications. All CSUs within a cell had the same probability of selection. A systematic PPS sample of CSUs was then drawn for the state.

This design produced a final sample of approximately 4,000 companies. Selected companies were asked to report sales by end-use categories for distillate fuel, residual fuel oil, and kerosene.

Form EIA-821 Estimation: For obtaining total estimates of volume, the adjusted probability estimator is used. This estimator, the sum of the weighted volumes, is defined as follows:

$$V = \sum_h (\sum_i W_{ih} V_{ih}), \text{ where:}$$

V = total estimated volume,

Σ_h = summation over strata,

Σ_i = summation over units within stratum h ,

W_{ih} = weight attached to unit i in stratum h

(the reciprocal of the probability of selection, P_{ih} , for that unit), and

V_{ih} = volume reported or imputed for units i in stratum h .

Survey nonrespondent volumes are also imputed as the mean of their strata.

B.3. Maximizing Response Rates

All the units that are surveyed may not respond (unit non-response) or may not provide all the information requested (item non-response). Alternative modes of data collection and follow-up are employed to encourage maximum response to the surveys in the Petroleum Marketing Program (PMP). Respondents are allowed to report by mail, fax, phone, or electronically using Excel forms or a fillable PDF available from the [survey directory](#) on EIA's website.

The nonresponse strategy for each of the monthly surveys is to generate a follow-up email or phone call within five days of the reporting deadline. Late respondents on both the weekly and monthly surveys are emailed or called and asked to submit data. If a weekly respondent still fails to respond after our initial reminder during data collection, secondary contacts are emailed or called for data. If a firm repeatedly fails to respond, a noncompliance letter requesting submission by a specific date is sent. The strategy for the annual survey is to generate a follow-up email or phone call within two weeks of the reporting deadline and on a monthly basis thereafter. Five months after the reporting deadline, a noncompliance letter requesting submission by a specific date is mailed. Table B2 contains the average annual response rate for the weekly, monthly, and annual surveys from January 2016 to December 2016. The EIA-821 has not finalized data collection for 2016 data.

Table B2: 2016 Average Annual Response Rates for PMP Surveys

Survey	EIA-14	EIA-182	EIA-782A	EIA-782C	EIA-821	EIA-856	EIA-877	EIA-878	EIA-888
--------	--------	---------	----------	----------	---------	---------	---------	---------	---------

Response Rate	100%	100%	100%	100%	*92.7 %	100%	98.6%	99.3%	99.7%
---------------	------	------	------	------	---------	------	-------	-------	-------

*Response rate based on 2015 data.

B.4. Test Procedures and Form Consultations

In preparation for clearance of the 2017 Petroleum Marketing Program by the Office of Management and Budget (OMB), EIA's Survey Development Team (SDT) received approval from OMB to conduct cognitive testing in 2016 on several of the surveys to assess the impact of potential changes for future iterations of the surveys. The evaluation focused principally on: respondent understanding of various definitions, how data is being reported, the ability to report other data items, and to update survey burden estimates. The ultimate goal of this process was to improve data quality.

The first step of the process was for SDT to meet with PBS managers to determine material issues for each survey and formulate the research objectives of the testing. SDT then developed draft cognitive interview protocols for PBS to assess respondent's ability to report new information, clarify areas of confusion or misreporting of data elements, possible deletion or collapsing of data categories, and verifying reporting burden. After the protocols were reviewed and accepted, a generic clearance request for clearance was submitted to OMB and approved for research on Forms EIA-182, 856, 782A, 782C, and 821. SDT contacted potential respondents based on recommendations from PBS. Since this testing was voluntary, several phone calls were made to identify participants for the study. The confidential interviews lasted between 15 to 30 minutes, depending on the survey. The interviews consisted of an introduction of the interviewers, the purpose of the call, background information on reporting practices, survey-specific questions, and a final opportunity for respondents to comment on any matter they wished to mention. Final reports were drafted for each survey summarizing the objectives, methodology, and specific findings and provided to PBS for consideration.

B.5. Statistical Consultations

Publicly available studies and research papers prepared by EIA statisticians and contractors regarding surveys in the PMP are available upon request. This list includes only publicly available reports. In addition, staff worked with contractor Z, Inc. to conduct a quality assessment in FY2011 which serves as a basis for future survey changes and with staff from the Office of Survey Development and Statistical Integration (SDSI) on cognitive testing of the surveys in the PMP.

PMP staff met with numerous internal data users – AEO, IEO, SEDS, and STEO - to consider their needs. In addition, staff also gave presentations at the following conferences to obtain feedback from data users:

- American Statistical Association (ASA) Conference (2009 and 2010)
- State Energy Data Needs Workshop (2009)
- Energy Markets and Financial Initiative (2010)
- EIA's Annual Energy Conference (2011)
- Kauffman Foundation Forum on Establishment Surveys (2011)
- Federal Committee for Statistics and Methodology - FCSM (2012)
- FCSM Statistical Policy Conference (2016)

Contact for the Petroleum Marketing Program: Ms. Tammy Heppner, Supervisory Statistician, Office of Petroleum and Biofuels Statistics (PBS), 202-586-4748.

For information concerning this request for OMB approval, please contact the agency Forms Clearance Officer Lawrence Stroud, at 202-586-6242, or lawrence.stroud@eia.gov.