



Independent Statistics & Analysis
U.S. Energy Information
Administration

Supporting Statement for Survey Clearance

Part B: Collections of Information Employing Statistical Methods

*Form EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas
Production Report*

OMB No. 1905-0205

September 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.....	1
B.3. Maximizing Response Rates.....	7
B.4. Test Procedures and Form Consultations.....	7
B.5. Statistical Consultations.....	8

B.1. Respondent Universe

The universe for the EIA-914 sample consists of all operators of wells in the lower 48 United States that are producing natural gas, crude oil and/or lease condensate, including federal and state offshore wells. A well operator producing natural gas, crude oil and/or lease condensate forms the responding unit. The files used to construct the frame includes a file by [Drilling Info](#) (DI) that contains well operator and monthly production volumes by state. The Form EIA-23L *Annual Survey of Domestic Oil and Gas Reserves* contains names, addresses, and annual production numbers for the previous calendar year.

A monthly cut-off sample of well operators is designed to give the targeted relative standard errors in all published geographic areas with the minimum number of operators while reducing sample turnover from month to month. Currently, EIA met this quality goal with about 360 respondent companies. This number fluctuates from month to month.

The EIA-914 program office uses a cut-off sample each month. The sample consists of operators whose combined production equals 85% of the total state production. When an operator is selected, they must report for all the states they produce crude/ lease condensate and/or natural gas in. This method allows the EIA-914 program office to obtain approximately 90% production coverage at the lower-48 level for natural gas, crude oil and lease condensate. This ensures adequate production coverage for each geographic area where production data are collected.

The cut-off sample operators are selected based on their natural gas and oil production obtained from a current commercially available data source, "These sources provide readily consolidated data that is recent and accurate. These two data sources constitute the frame file for the EIA-914. Currently, the EIA-914 frame consists of about 12,000 active operators of crude oil and gas wells.

B.2. Statistical Methods

B.2.1 Sampling

The DI database supplemented with Form EIA-23 data is used for both the sampling and estimation processes. DI acquires well or lease level oil and natural gas information from state agencies, places it in their own database format, performs some data standardization and validation. A new DI database is transmitted to EIA at the end of every month.

DI data for the most recent reporting months are usually incomplete. The months used for sampling are determined by how much of a lag there is in the DI data before they become complete. A different lag time is used for every state/area and for oil and gas. The process used to determine the lag time to complete data is described in further detail in the [EIA-914 Monthly Crude Oil, Lease Condensate, and Natural Gas Production Report Methodology](#). Table 1 shows lag times by state and commodity.

Table 1: State lags, September 2017

State	Oil Lags (number of months)	Gas Lags (number of months)
Arkansas	6	4
Alabama	2	2
California	4	5
Colorado	13	13
Federal Gulf of Mexico	6	5
Federal Offshore Pacific	4	5
Kansas	5	5
Louisiana	5	6
Michigan	4	3
Mississippi	3	3
Montana	5	6
New Mexico	6	4
North Dakota	6	6
Ohio	19	19
Oklahoma	18	18
Pennsylvania	7	7
Texas	17	9
Utah	4	4
Virginia	4	3
West Virginia	7	7
Wyoming	5	4

For operators not in the sample, the most recent four months of DI production is examined for each of them. If a non-sampled operator is observed to produce above the state cutoff for either crude oil or natural gas for four consecutive months, that company is added to the sample for all the states in which

they produce. The production cutoff varies by each state because state production levels differ. If a sampled company reports production on the Form EIA-914 below the cutoff in every area in which it produces, for six consecutive months, it is dropped from the sample. These procedures maintain high sample coverage of the population, and minimize frequent adding and dropping of operators.

B.2.2 Rationale for 85 percent sample coverage target

The 85 percent target was selected to be consistent with the current sampling rate for the EIA-914 survey and to maintain consistency of data series, and to guard against inaccurate or biased model estimates. In general, relative standard errors are within acceptable ranges at an 85 percent sample volume coverage. Of more practical concern is that in some specific cases, large operators can behave very differently from small operators. This occurred in Louisiana during 2009 when the Haynesville Shale was developed. When the sampled companies do not represent the non-sampled companies the model estimates will not accurately reflect production activity in the state. By using a high sample coverage rate, data quality issues arising from model failure is mitigated.

DI data for four of the smaller producing states are missing or inadequate. For Illinois, Indiana, Kentucky, and Tennessee, annual production data from the Form EIA-23 (*Annual Oil and Gas Reserves Report*) survey are used to supplement the DI database. Hereafter, references to DI data include supplemental data from the Form EIA-23 survey for these four states. These data are used to determine a different cutoff sample for each state/area designed to achieve a roughly 85 percent sample coverage of that area's oil and natural gas production volumes. The cutoff production rates will be re-determined, annually, to maintain the target of 85 percent coverage of state level production. Although a target of 85 percent sample coverage is used to design the sample, in practice this typically results in coverage closer to 90 percent because when an operator is selected to report their production for one state, they also are required to report their production for all states in which they operate.

The problem of inaccurate or biased model estimates can normally be solved by combining a certainty stratum with a probability-proportional-to-size sampling or other random sampling scheme of the smaller companies. Because the sampling frame is in a constant state of flux due to operator acquisitions and mergers, sample weights calculated on a monthly basis are prone to error. Using the classical ratio estimator, this sort of behavior causes much less fluctuation in estimates.

Another way of solving the problem of inaccurate or biased model estimates is through stratification of variables based on well characteristics, such as formation type (shale, conventional, etc.). While the frame has such information, it is often reported inconsistently and sometimes incorrectly. Also, to be effective, the respondent would have to provide this information in their survey response since an individual operator may produce large volumes from multiple wells in multiple formations. Not only would this increase burden on the respondent, in EIA's experience with the EIA-23, respondents do not provide this information reliably enough to clean the data within the necessary turnaround time.

States such as Kansas and Arkansas have higher numbers of small oil producers than other major states. States composed primarily of small operators and without potential for dramatic growth are generally much less likely to suffer from inaccurate or biased model estimates. This is because the sampled

companies accurately reflect the behavior of the non-sampled companies. As a result, the below 85 percent sample coverage shown in Table 1 is acceptable.

B.2.3 Rationale for 500 barrel per day minimum threshold

In addition to the 85 percent sample coverage criteria used to determine the oil and gas production cutoff values, a minimum crude oil cutoff rate of 500 barrels per day (bpd) is used as an additional constraint to limit the number of smaller operators sampled. This is needed because some liquids producing states, such as Kansas, are composed of an inordinate number of small operators, and achieving an 85 percent sample coverage would create a much larger sample in a region that EIA does not expect model failure to be an issue. Application of the 500 bpd minimum oil cutoff rate will yield less than 85 percent coverage in some states with small production volumes and also minimizes the approximate sample size to 500 operators. A sample size of 500 accommodates fluctuations in operators that are selected to report due to external economic triggers. The triggers can include significant price movements or merger and acquisitions activity. The application of this lower bound on the oil cutoff production rate creates an acceptable balance between sample coverage and reduced burden on smaller operators.

EIA considered what impact various minimum cutoffs would have on the sample and resulting data quality. Sample coverage declines at increasing cutoff levels (see Table 1). With no minimum cutoff, coverage across all states meets the EIA goal of 85 percent coverage in each state, but sample size increases considerably, thereby increasing burden. Raising the minimum threshold to 500 bpd, 1,000 bpd, and finally 5,000 bpd succeeds in lowering the sample size, but generates progressively worse coverage and less reliable estimates. The minimum cutoff rate of 500 b/d was selected in part because this was used effectively on Form EIA-23 to estimate annual crude oil production.

State	Minimum Cutoff Applied (bpd)			
	0	500	1,000	5,000
Alabama	87%	87%	83%	78%
Arkansas	85%	44%	26%	20%
California	90%	90%	88%	88%
Colorado	92%	92%	91%	91%
Federal Gulf of Mexico	92%	92%	92%	90%
Federal Pacific	86%	86%	86%	62%
Kansas	85%	57%	48%	29%
Louisiana	86%	79%	71%	39%
Michigan	85%	78%	61%	8%
Mississippi	87%	86%	81%	50%
Montana	88%	88%	87%	70%
North Dakota	87%	87%	87%	87%
New Mexico	86%	86%	86%	86%
Ohio	86%	85%	85%	72%
Oklahoma	85%	85%	79%	68%
Other States	86%	67%	60%	31%
Pennsylvania	86%	70%	70%	55%
Texas	86%	86%	86%	86%
Utah	87%	87%	87%	76%
Virginia	91%	0%	0%	0%
West Virginia	87%	87%	87%	56%
Wyoming	85%	85%	85%	64%
Lower 48	87%	87%	86%	83%
Sample Size	643	380	297	254

Note: This table is the result of a simulation based on some simplifying assumptions that are more conservative than reality. It illustrates that there is a significant reduction in sample size with the application of minimum cutoffs. Actual coverages are slightly higher.

Form EIA-914 will utilize 44 cutoff samples, one for each of the 22 states/areas for crude oil and lease condensate production, and another 22 cutoff samples for each of the states/areas for natural gas production. If an operator is sampled for one state/area for either crude oil or natural gas, then it must report natural gas, crude oil and lease condensate for all states/areas where it has production, regardless of volume in those areas.

Form EIA-914 utilizes samples for the 22 states/areas for which crude oil and lease condensate, and natural gas production respondents must provide information separately for each state. The remaining 14 states (the "Other States" group) that have crude oil and lease condensate, or natural gas production will also be sampled, as a single target variable, collectively rather than individually by state.

B.2.4 Estimating

The target population to report on Form EIA-914 is all active crude oil and lease condensate, and natural gas producers. Members of the target population not sampled must be estimated for.

Estimation will be performed separately for natural gas and liquids production for each of the 22 states/areas using the Classical Ratio Estimator (CRE), which multiplies the current survey reported production by a ratio. The CRE ratio is an area's total production in an earlier 6-month period as found in the DI database, to the area's production from the current sample of operators in the same period, also found in the DI database. The six month period is dependent on the reporting lag time for each state. A six-month ratio is used to avoid potential volatility from using a ratio of a single month's production. The DI data are typically incomplete for more recent months. The months used in calculating the ratio are determined by how much of a lag there is in the DI data before they become sufficiently complete. A different lag is used for each state/area and for liquids and natural gas. Estimates for the survey month are computed using the following formula for each area:

$$\widehat{TP}_m = S_m \left(\frac{\sum_{i=L}^{L+5} TP_{m-i}}{\sum_{i=L}^{L+5} SP_{m-i}} \right)$$

where:

$\widehat{TP}_m = \hat{i}$ Estimated total production for the state/area in month 'm'

$S_m = \hat{i}$ Total survey reported production for the state/area in month 'm'

L = The DI lag for the particular state/area

$TP_i = \hat{i}$ Total production for the area in month 'i' from the DI database

$SP_i = \hat{i}$ Production of the current sample of operators in month 'i' from the DI database

Errors can occur in the estimation process when producing properties are transferred between operators, and this transfer is not reflected in both the DI database and the Form EIA-914. This can be caused by mergers and acquisitions. EIA monitors these events by following trade journals and news releases and by following up with respondents whose reported production is substantially different from what is expected given their production history. When transfers and other changes occur, EIA adjusts either the DI data or the survey data, as appropriate, to synchronize reporting until the two are in sync again.

Some responses can be far outside the anticipated range if the company is experiencing extreme growth or decline, or some of the production reported on the survey is not represented in the DI data (this can occur for various reasons). In such cases, the company is treated as an outlier/add-on. This means that the company production is excluded from the sums in the ratio equation above, and its reported production is then added to the total estimate.

All estimates will be based on the CRE method with the exception of "Other States," which is currently determined by applying the ratio of total gross natural gas, and crude oil and condensate production

published in the Natural Gas Annual and Petroleum Supply Annual (NGA and PSA) to the EIA-914 surveyed production for the most recent available year. The numbers from the NGA and PSA come from state agencies. This is used instead of the CRE because the frame for the “Other States” group is less complete. If the frame coverage in the other states improves it could be estimated using the CRE method as well, but this is unlikely. The “Other States” group is roughly 1 percent or less for both oil and gas production in the lower 48 states.

B.2.5 Imputation for Item Non-Response

In the event that item level non-responses cannot be resolved within a survey cycle, EIA uses the average of the previous three months’ data to impute current month data for a non-respondent. If three months of survey data are unavailable, then DI data is used in the imputation formula. EIA will replace the imputed data with actual data, once the respondent files the report with EIA. If any of the three previous months’ survey data are deemed not analogous to the current missing reporting month’s data (i.e., being influenced by maintenance, weather event, power outage, etc.) that previous month’s data would not be used in the imputation process.

B.2.6 Frame Maintenance

Since the inception of the Form EIA-914 in 2005, EIA has maintained a natural gas operator frame of the nation’s natural gas producers. EIA uses various sources of information to maintain the completeness of the frame, including trade press, other data providers, the EIA-23 frame file, and state regulatory agencies. EIA maintains the natural gas, and crude oil and lease condensate frame by following similar practices.

EIA continuously reviews all available information to adjust the survey frame for the Form EIA-914 data collection for births, deaths, mergers, and company information changes.

B.2.7 Efforts to Reduce Total Survey Error

Frame Coverage Errors

In states with stricter reporting requirements, DI obtains more complete data on a timely monthly basis and frame coverage error is expected to be negligible for those states. These states make up the majority of U.S. production. However, in a minority of generally smaller producing states, data are less timely and less complete, requiring continuing effort on the part of EIA to research companies that may be operating in these areas to ensure the best possible frame coverage.

Reporting Errors and Data Processing Errors

Form EIA-914 questionnaire was carefully developed and tested with potential respondents to estimate reporting burden and respondents’ ability to report the requested information within the 40 day reporting period to minimize the amount on imputation in the published estimates. The survey instrument includes a detailed set of instructions for filing data, subject to a common set of definitions similar to those already used by the industry. EIA continually enhances data validation/editing software (programmed edits checks) to detect probable reporting errors and flag them for resolution by analysts,

either through confirmation of the data by the respondent or submission of amendments to the previously-filed data.

Data processing is automated to minimize the possibility of data entry errors by EIA staff. Data processing errors are further minimized through the use of data quality metrics on the aggregate data and outlier analysis performed on the micro data, which can help identify both reporting and data processing errors.

Estimation Errors by Respondents

EIA uses agency-developed software to detect Form EIA-914 responses that appear out of the ordinary and are candidates for further investigation. Software is used to flag data items that are inconsistent or out of their normal range, which are further investigated by EIA analysts. When necessary, EIA analysts contact the respondents to resolve related questions.

Revisions

Revisions to aggregated data from respondents submitting revised data for prior months or late data for the current month are published according to the Form EIA-914 Revision Policy. This revision policy is [available here](#). Publication revision thresholds have been determined for each state/area and for oil and gas. In addition, crude oil and lease condensate (combined), and natural gas production estimates will be replaced with state reported volumes when state reported volumes are deemed complete.

Respondents are required to resubmit data when their production estimates change by more than 1,000 barrels of crude oil and lease condensate in a state, or by more than 150 million cubic feet of natural gas in a state, and this requirement is detailed in the survey instructions.

B.3. Maximizing Response Rates

EIA will use standard procedures to administer the data collections for Form EIA-914. An introductory letter signed by the responsible EIA official is sent to each company that will be a respondent to Form EIA-914. Follow-up procedures for non-response consist of an email message or a reminder phone call (for those not using email) to all companies that do not return a completed survey form by the due date. This initial re-contact is followed by repeated email messages, letters, and phone calls until an appropriate response is received.

In the event that item level non-responses cannot be resolved within a survey cycle, EIA uses the average of the previous three months' data to impute current month data for a non-respondent. If three months of survey data are unavailable, then DI data is used in the imputation formula. EIA will replace the imputed data with actual data, once the respondent files the report to EIA. If any of the three previous months' survey data are deemed not analogous to the current missing reporting month's data (i.e., being influenced by maintenance, weather event, power outage, etc.) that previous month's data would not be used in the imputation process.

B.4. Test Procedures and Form Consultations

EIA conducted two rounds of cognitive testing of Form EIA-914 with respondents. This research involved verifying the reporting burden and respondents' ability to report crude oil production during the Fall of 2016 and Spring of 2017.

B.5. Statistical Consultations

For additional information concerning this data collection, please contact Jeff Little at (202) 586-6284, or jeffrey.little@eia.gov.

For information concerning this request for OMB approval, please contact the agency Forms Clearance Officer, Patrick Murray, at (202) 287-6435, or patrick.murray@eia.gov.