

## EIA's Frames: How Do We Know If They Are Sufficient?

Presenters include the following staff from Statistics and Methods Group:

Grace Sutherland – [grace.sutherland@eia.doe.gov](mailto:grace.sutherland@eia.doe.gov)

Howard Bradsher-Fredrick – [howard.bradsher-fredrick@eia.doe.gov](mailto:howard.bradsher-fredrick@eia.doe.gov)

Shawna Waugh – [Shawna.waugh@eia.doe.gov](mailto:Shawna.waugh@eia.doe.gov)

Alethea Jennings – [Alethea.jennings@eia.doe.gov](mailto:Alethea.jennings@eia.doe.gov)

### Introduction

The quality of EIA's data has been made a priority and has been made part of its Strategic Plan. Goal 1 of the EIA Strategic Plan states "...EIA's information products will retain or improve their high quality..." One of the performance measures and targets for this goal involves evaluating the EIA frames. The measure is the percent of EIA survey frames with sufficient industry coverage to produce reliable supply, demand and price statistics.

We began by preparing a list of EIA survey frames and update procedures. A table of EIA frames is enclosed (Attachment 1). The table also includes EIA's consumption surveys, which are very large samples. The table shows the survey's name, number of respondents, and frequency of frames updates activities.

We are now in the process of gathering existing information on the quality of our frames. This presentation will provide examples of activities to assess frame quality and ask for the Committee's guidance on other activities that could be used to assess our frames. We are also interested in the Committee's thoughts on how we can develop criteria to define sufficient coverage.

Activities to assess frame quality include:

1. Checking the frame against alternative lists at the respondent level (suggestion taken from the Statistical Policy Working Paper 15). There are some situations where we know we have good coverage. For example, all nuclear plants are known through the Nuclear Regulatory Commission. In addition, the natural gas marketer survey uses state-licensed marketers. In other situations a more detailed examination is necessary. Following are three examples of recent activities.
  - A. Renewable Electric Plant Information System (REPiS) database was developed and is maintained by National Renewable Energy Laboratory with funding from the Department of Energy's office of Energy Efficiency and Renewable Energy (EE), whose data come from publicly available sources, such as federal and state government publications and reports; trade associations; trade press literature, such as weekly newsletters; and personal communications with industry and government officials. No surveys are conducted to collect data. The data represent "best efforts" at compiling and verifying an inventory of all known grid-connected renewable electric facilities in the United States. EIA utility and

non-utility electric power plant data from survey Form EIA-860 were compared to data from REPiS. The eventual outcomes of the comparison were: 1) the data matched exactly across data sets; 2) a unit was in the EIA data but not in REPiS; 3) in REPiS but not in EIA; 4) information for one or more of the data elements of interest was inconsistent across databases. (All project worksheets are available.) If the data differed, staff familiar with the EIA-860 survey was contacted for more information, e.g., updated data, or any changes to the data that might explain the difference in the comparison. As a result of the comparisons 214 potential missing plants amounting to 2418 MW of nameplate capacity were identified. We are still in the process of reviewing these plants prior to updating the frame. Attachment 2 is the paper in its entirety.

B. We have drafted a proposal for the Census Bureau (attachment 3) to compare several EIA surveys with either MECS or the Economic Census. The surveys that we are interested in obtaining an evaluation of coverage of their 2002 frames include EIA's "Annual Electric Generator Report" (Combined Heat and Power Plants, NAICS 31-33), "Quarterly Coal Consumption and Quality Report, Manufacturing Plants", "Annual Solar Thermal Collector Manufactures Survey", "Annual Photovoltaic Mod/Cell Manufacturers Survey", and "Quarterly Coal Consumption and Quality Report, Coke Plants."

C. EIA cross-referenced it's respondent i.d. lists between the Quarterly Coal Consumption and Quality Report, Manufacturing Plants," and the "Power Plant Report" In order to reduce duplication of respondents for both surveys. As a result we added 1 respondent to the frame for the Quarterly Coal Consumption and Quality Report.

2. Data Comparisons at an aggregate level (certain EIA data compared with similar data outside of EIA). The Petroleum Marketing division compared data series with others that should match in volumes and price such as the Bureau of Labor Statistics (BLS) Office of Consumer Price Index (CPI) data for retail prices of motor gasoline, diesel fuel, and residential No. 2 fuel oil and Form EIA-878, "Motor Gasoline Price Survey," for retail prices of gasoline. Differences across data sources could indicate differences in survey methodology and conceptual differences, but can also identify coverage and reporting errors. There was a large difference in estimates of volumes of residual fuel oil prior to 1993. As a result improvements were made when exclusionary lists were added for the respondents' use. This change improved coverage and eliminated double counting of volumes sold. A data correction was also made separating imports of residual fuel oil and unfinished oil. This reduced the gap between data series for residual fuel volumes. See attachment 4 for details.
3. Examining Supply/Disposition Balances (supply should equal disposition, but because the data that comprise the supply/disposition balances are from different surveys, a balancing item is needed. If the balancing item becomes large, it could be an indication of a frame or other data quality problem). The Natural Gas Annual Table 1 shows a narrowing of its balancing items. In 1998, there was a 634,809 million cubic feet (mcf) difference between supply and disposition. By 2002, the gap had narrowed to -39,942 mcf. Two exercises that the natural gas division engaged in may have been the factors for the improvement. 1) The source from where the data on deliveries to the electric power sector was obtained was changed. And 2) The frames for the form EIA-176, "Annual Report of Natural and Supplemental Gas Supply Disposition" was evaluated and the respondent list

was adjusted to include or drop respondents as appropriate. See the attachment 5 for more details on the balancing item. Other balancing items have been examined in other parts of EIA. See the following URLs.

[http://www.eia.doe.gov/emeu/aer/pdf/pages/sec7\\_5.pdf](http://www.eia.doe.gov/emeu/aer/pdf/pages/sec7_5.pdf),

[http://www.eia.doe.gov/emeu/aer/pdf/pages/sec8\\_5.pdf](http://www.eia.doe.gov/emeu/aer/pdf/pages/sec8_5.pdf),

[http://www.eia.doe.gov/pub/oil\\_gas/petroleum/data\\_publications/petroleum\\_supply\\_monthly/current/pdf/stable2.pdf](http://www.eia.doe.gov/pub/oil_gas/petroleum/data_publications/petroleum_supply_monthly/current/pdf/stable2.pdf),

[http://www.eia.doe.gov/pub/oil\\_gas/natural\\_gas/data\\_publications/natural\\_gas\\_annual/current/pdf/table\\_001.pdf](http://www.eia.doe.gov/pub/oil_gas/natural_gas/data_publications/natural_gas_annual/current/pdf/table_001.pdf)

4. For price data, a comparison of volumes that the price data represent with total volumes. For example, natural gas commercial price data represent about 78 percent of total commercial volumes in 2002. This is because the price data have been collected from pipelines and local distribution companies which no longer know the price of all of the gas they deliver if it is sold by marketers. EIA introduced Form EIA-910, “Monthly Natural Gas Marketers Survey” to capture the price of natural gas sold by marketers to residential and commercial customers in five states with active customer choice programs. Attachment 6 is our evaluation of the survey at: [www.eia.doe.gov/smg/asa\\_meeting\\_2003/fall/files/natgaseval.doc](http://www.eia.doe.gov/smg/asa_meeting_2003/fall/files/natgaseval.doc)

### **Next Steps and Question for the Committee**

We will continue to gather information on frames assessments throughout EIA and to see where there are areas where no assessments have recently been done. We will also try to obtain information from other agencies on how they determine whether they have sufficient coverage and to begin discussions within EIA on criteria to define sufficient coverage.

Given that we will have information on coverage from different sources (checking alternative lists, comparisons at the aggregate level, examination of balancing items), what are the Committee’s thoughts on how to define sufficient coverage?

## Attachment 3: Proposed Frame Analysis

### A. Purpose of Analysis

1. Evaluate coverage of 2002 frames for EIA-860 (CHP), EIA-3, EIA-5, EIA-63a, and EIA-63b.
  - [EIA-860](#), “Annual Electric Generator Report” (Combined Heat and Power Plants, NAICS 31-33)
  - [EIA-3](#), “Quarterly Coal Consumption and Quality Report, Manufacturing Plants”
  - [EIA-63a](#), “Annual Solar Thermal Collector Manufactures Survey”
  - [EIA-63b](#), “Annual Photovoltaic Mod/Cell Manufacturers Survey”
  - [EIA-5](#), “Quarterly Coal Consumption and Quality Report, Coke Plants”
2. Identify for each frame geographic region and NAICS code where there are coverage problems (if possible).

### B. Analysis of Coverage of EIA Frames

**Energy-intensive industries** are of special importance in evaluating coverage.

MECS/EC      Initially use 2002 MECS to evaluate coverage of EIA-860 and EIA-3.

*Two-step process:*

1. The first step is to match establishments on EIA frame to MECS.  
If unable to match establishments to MECs, then
2. The second step is to match establishment to Census of Manufacturing (CM).

*Rationale for using MECS initially to match establishments*

- MECS is a sample (~15,500) of the Census of Manufacturing (NAICS 31-33)
- MECS identifies establishments with cogeneration technology (~1,300 in 1998)
- MECS is used for these two frames because not all variables selected for evaluating coverage are available on CM.

Census of

Manufacturing      Use CM to evaluate coverage of EIA-63a, EIA-63b, and EIA-5  
Census of Manufacturing (CM) includes NAICS 31-33 and by volume of shipments includes 98% coverage.

Establishments on frame for EIA-63a and EIA-63b are likely in NAICS 334413  
Establishments on frame for EIA-5 (Coke Plants) may be in NAICS 324 or 331.  
EIA does not collect information from establishments on NAICS codes for these frames.

### C. Size of EIA Frames

The number of establishments on the following surveys is for 2002  
Approximately 100 establishments are on both EIA-860 and EIA-3  
EIA will provide crosswalk between EIA-860 and EIA-3

EIA-860 (CHPs, NAICS 31-33)	1,500
EIA-3	562
EIA-63a	29
EIA-63b	22
EIA-5	20

### D. Variables to use to Evaluate Coverage

EIA-860	Total electricity generation (mgwthr) and possible fuel consumption (physical units)
EIA-3	Coal consumption (short tons)
EIA-63a	Total shipments (peak megawatthours) or Value of shipments (\$)
EIA-63b	Total shipments (square feet) or Value of shipments (\$)
EIA-5	Coke and breeze production and coal consumption (all 3 in short tons)

### E. Details on Matching and Evaluating Coverage

Causes for difference between EIA and Census frames have been identified in Appendix I.

#### 1. Form EIA-860, "Annual Electric Generator Report"

##### **Frame (EIA-860)**

- Approximately 1,500 Establishments on EIA-860 frame
- Frame excludes inactive establishments
- Frame excludes establishments with a nameplate rating of 1 megawatt (1000 kW) or less
- Establishments with primary or secondary NAICS code in the manufacturing sector (31 to 33)
- Respondents self-report NAICS codes

##### **Matching**

###### *Criteria to consider*

- Company Name
- Establishment Name and address (city, state, zip code)
- Prime mover and cogeneration technology (MECS) used to generate electricity
- Fuel consumption by type of fuel
- 3-digit NAICS code (primary and secondary only for establishments with primary of NAICS 22)
- Electricity generated
- Cogeneration technology

**Coverage (Electricity Generated)**

U.S. Total (separately for cogen, noncogen, and combined)

$$\text{Count \%} = \frac{\text{Number of Matched Active Establishments on MECS}}{\text{Number of Active Inscope Establishments on MECS}}$$

*Unweighted*

$$\text{Volume \%} = \frac{\text{Unweighted Volume of Electricity Gen. for Matched Active Establishments on MECS}}{\text{Unweighted Volume of Electricity Gen for All Active Inscope Ests on MECS}}$$

*Weighted*

$$\text{Volume \%} = \frac{\text{Weighted Volume of Electricity Gen. for Matched Active Establishments on MECS}}{\text{Weighted Volume of Electricity Gen. for All Active Inscope Ests on MECS}}$$

In addition, similar information for Fuel Consumption for select Fuel Types would be of interest.

Disaggregated

Totals Use formula to provide percent of count/volume by geographic region and NAICS code to the extent possible.

**Analysis of Matched/Nonmatched Establishments by Count and Volume**

Cogen	Matched/Nonmatched	Count	Volume
Cogen	Matched respondents		
	On MECS frame, inscope and not on EIA frame		
	On EIA frame and not on MECS frame		
Non-cogen	Matched respondents		
	On MECS frame, inscope and not on EIA frame		
	On EIA frame and not on MECS frame		

2. Form EIA-3, “Quarterly Coal Consumption and Quality Report, Manufacturing Plants”

**Frame**

- 562 Establishments on EIA-3 Frame
- Establishments self-report NAICS code
- Crosswalk between EIA-3 and EIA-860 for over 100 establishments

**Matching**

*Criteria to consider*

- Company Name
- Establishment Name and address (city, state, zip code)
- 3-digit NAICS code
- Type of coal (Anthracite, Bituminous/Sub, and Lignite) consumed

**Coverage (Coal Consumed)**

U.S. Total

$$\text{For Count \%} = \frac{\text{Number of Matched Active Establishments on MECS that Consume Coal}}{\text{Number of Active Establishments Inscope on MECS that Consume Coal}}$$

*For Volume Option 1*

$$\text{\%} = \frac{\text{Unweighted Volume of Coal Consumed for Matched Active Ests on MECS}}{\text{Unweighted Volume of Coal Consumed for All Active Ests Inscope on MECS}}$$

For Volume Option 2  $\% = \frac{\text{Weighted Volume of Coal Consumed for Matched Active Ests on MECS}}{\text{Weighted Volume of Coal Consumed for All Active Inscope Ests on MECS}}$

Disaggregated

Totals Percent for both count and volume by coal type and if possible 3-digit NAICS code, to the extent possible

**Analysis of Matched/Nonmatched Establishments by Count and Volume**

Matched/Nonmatched	Count	Volume
Matched respondents		
On MECS frame, inscope and not on EIA frame		
On EIA frame and not on MECS frame		

3. For the following three forms, similar types of evaluation of coverage and analysis at US level only.

3a Form EIA-63a “Solar Thermal Collector Manufactures Survey”

**Frame**

- 29 Establishments on EIA-63a frame
- May be classified in NAICS 334413
- Establishments do not report NAICS code

**Matching**

*Criteria to consider*

- Company Name
- Establishment Name and address (city, state, zip code)
- Shipments of solar thermal collectors and associated revenue

3b. Form EIA-63b “Annual Photovoltaic Mod/Cell Manufacturers Survey”

**Frame**

- 22 Establishments on EIA-63b frame
- May be classified in NAICS 334413
- Establishments do not report NAICS code

**Matching**

*Criteria to consider*

- Company Name
- Establishment Name and address (city, state, zip code)
- Shipment of photovoltaic modules/cells and associated revenue

3c. Form EIA-5 “Quarterly Coal Consumption and Quality Report, Coke Plants”

**Frame**

- 24 Establishments on EIA-5 frame
- Establishments may be classified in NAICS 324 or 331
- Establishments do not report NAICS code

**Matching**

*Criteria to consider*

- Company Name
- Establishment Name and address (city, state, zip code)

## Appendix I: Potential Difference in Frames

The number of establishments on EIA's frame may differ from those on Census frame due to:

- **NAICS classification**
  - Respondents of EIA self-report NAICS codes for their establishments
  - Establishments on EIA-860 have both a primary and secondary NAICS code. It is possible establishments classified in manufacturing sector (NAICS 31 to 33) on the Census of Manufacturing have a primary code of NAICS 22 and a secondary NAICS code in the manufacturing sector on the frame for EIA-860
- **Status** of operations: EIA only includes active establishments whereas the Census of Manufacturing/MECS contain active and inactive establishments
- **Boundary** of establishments: The boundary of the establishments that report on the Census of Manufacturing/MECS may differ from boundary of the establishments that report on EIA surveys, especially EIA-860 and EIA-3. In addition, for establishments on both of these EIA frames, which match CM/MECS by address, it is possible the respondents will report different quantities of fuel consumption. This may, in part, be due to what the establishment considers to be "inside the boundary" For example, if the generator is offsite, then on the MECS and CM fuel consumed by the generator would be excluded whereas on the EIA forms this fuel consumed at the offsite generated (which provides electricity to the plant) would be included.

EIA-860	All electric generating plants, which have or will have a nameplate rating of 1 megawatt (1000 kW) or more. The operators of jointly-owned plants should be the only respondent for those plants. EIA will provide list of all establishments in <b>manufacturing sector</b> with primary or secondary NAICS 31 to 33.
EIA-3	Manufacturing companies that consume in excess of <b>1,000 short tons</b> of anthracite, bituminous, subbituminous coal or lignite for uses other than coke production during the year. This includes synfuel plants that use coal as a feedstock, other facilities using coal as a feedstock, and all facilities using coal for gasification/liquefaction.
EIA-63a	Solar thermal collectors may be a product of NAICS 334413. EIA does not collect NAICS codes.
EIA-63b	Photovoltaic modules/cells may be a product of NAICS 334413. EIA does not collect NAICS codes.
EIA-5	Census of Coke plants with potential differences due to difference in boundary of establishment.