

This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matter from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's spring 2006, meeting with the Committee to be held April 6 and 7, 2006.

# Preliminary Research Results on Respondent Cut-off Dates for EIA Electricity Data Collections

Howard Bradsher-Fredrick, E-mail [Howard.Bradsher-Fredrick@eia.doe.gov](mailto:Howard.Bradsher-Fredrick@eia.doe.gov) Telephone: 202-287-1721, Statistics and Methods Group, Energy Information Administration, EI-70

Alethea Jennings E-mail [Alethea.Jennings@eia.doe.gov](mailto:Alethea.Jennings@eia.doe.gov) Telephone: 202-287-1716, Statistics and Methods Group, Energy Information Administration, EI-70

## Introduction

In order to achieve high response rates on establishment surveys, EIA expends significant resources in administering non-response follow-up to those surveys. For example, on the EIA-861 survey<sup>1</sup>, *Annual Electric Power Industry Report*, while a large portion of the total volume for electricity sales is reported within a relatively short period of time, non-response follow-up continues for over four months following the final deadline in order to achieve its characteristic very high response rates. Considering tightening budgets, the issue can be raised as to whether EIA will be able to continue to expend significant resources to achieve near 100% coverage by volume.

In order to make rational decisions on this issue, it is advisable to study past data collections to first assess when data had been submitted and to then determine the character of the respondents and lost respondents associated with an array of alternative cut-off dates. This paper summarizes these preliminary analyses on 2004 submissions of the survey data obtained on the EIA-861, *Annual Electric Power Industry Report*, and EIA-860<sup>2</sup>, *Annual Electric Generator Report*. In addition to overall summaries, some analyses were also conducted on various strata important to data users.

In addition to discussing the work already completed, we would like to discuss plans for future analyses with the committee and any caveats or advice they can provide concerning such work.

---

<sup>1</sup> The EIA-861 is an annual census of electric power industry participants involved in generation, transmission, and distribution of electricity. The frame includes wholesale and retail marketers, as well as those providing bundled services. The form collects generation, wheeled energy, and sales by customer sector, by state, and by green pricing and net metering. The survey currently has about 3,300 active units. The frame is relatively stable (except for short-term marketers) and serves as the frame for the EIA-826. The EIA-826 is a cut-off sub-sample from the EIA-861. Monthly data for those not "sampled" are estimated using each company's prior year EIA-861 data. Hence, the estimation procedures would need to be revised if some facilities were dropped from the EIA-861.

<sup>2</sup> The EIA-860 is an annual census of all existing and planned electric generating facilities with a total generator nameplate capacity in excess of 1 megawatt. The survey collects information on the plant design and capabilities, but not on actual operations. Data are collected at the generator level. In 2003, the EIA-860 was filed by 2,700 companies that operate 5,700 (both existing and planned) plants containing over 16,000 generators. The EIA-860 frame also serves as the frame for other EIA electricity surveys, including three monthly surveys (423, 906, 920) and one annual survey, the EIA-767. The EIA-906 and EIA-920 have the same threshold (1 MW) as the EIA-860, and hence could be affected by any changes to the frame.

## **Overall Goal**

The overall goal of this work is to determine quantitatively the impact of various alternative policies regarding either surveying only a subset of the complete frame or cutting off the non-response follow-up earlier than has previously been EIA's policy. For research purposes, the EIA-860 and EIA-861 will be used as test cases to evaluate alternative policies. To the extent possible, we would also like to investigate any other issues related to early versus late survey respondents (e.g., data quality of the data received from early versus later respondents).

## **Short Term Objectives**

We began this study not really knowing what could be realistically investigated and how difficult it would be to conduct the analyses of interest to us. A great benefit associated with using the EIA-860 and EIA-861 data is that each respondent's data record has an associated submission date. Thus, we at least have the beginnings of the data needed to conduct analyses of interest to this project; data can be successfully stratified by submission date. Thus, we began with the following set of short-term objectives not knowing if data were available to conduct these studies to fully investigate:

1. To determine the effect on coverage (overall and stratified by fuel or sector) if EIA decided to cut-off the survey response stream on the EIA-860 on a variety of different dates.
2. To determine the effect on coverage of large producers/consumers on the EIA-861 by using a variety of cut-off dates.
3. To determine the tendency of the relatively largest producers (or consumers) to remain the largest over time.
4. To determine the tendency of the relatively smallest producers to remain the smallest over time.
5. To determine whether or not earlier respondents submitted data requiring less follow-up than later respondents.

## **Data Submission Profile for EIA-860 Respondents**

Data were available to determine the date when each respondent actually submitted their form on the EIA-860. The results are shown below in four tables. The first table provides the overall totals for all submissions in 2005 (2004 annual data are being submitted) for all fuels when the respondent reported their generators operating. Tables 2 – 4 show the profile of submissions by date for several major fuel types: non-hydroelectric renewables, natural gas and coal, respectively.

**Table 1: Overall Summary – Generators, Operating Status Only, 2004 Data**

Submission Date	Count	%	Cumulative %	Nameplate Capacity (MW)	%	Cumulative %	Avg Capacity (MW)
< May 1	8,374	57.0%	57.0%	650,937	64.7%	64.7%	77.7
May 1 – 31	3,084	21.0%	78.1%	210,047	20.9%	85.6%	68.1
June 1 – 30	1,752	11.9%	90.0%	95,688	9.5%	95.1%	54.6
July 1 – 31	803	5.5%	95.5%	28,182	2.8%	97.9%	35.1
Aug 1 – 31	398	2.7%	98.2%	18,380	1.8%	99.7%	46.2
Sep 1 – 9	25	0.2%	98.3%	197	0.0%	99.8%	7.9
Non-resp	244	1.7%	100.0%	2,161	0.2%	100.0%	8.9
<b>Total</b>	<b>14,680</b>			<b>1,005,592</b>			<b>68.5</b>

**Table 2: Non-Hydro Renewables -- Generators, Operating Status, 2004 Data**

Submission Date	Count	%	Cumulative %	Nameplate Capacity (MW)	%	Cumulative %	Avg. Capacity (MW)
< May 1	712	52.5%	52.5%	10,623	60.2%	60.2%	14.9
May 1 – 31	246	18.1%	70.6%	2,310	13.1%	73.3%	9.4
June 1 – 30	267	19.7%	90.3%	3,138	17.8%	91.0%	11.8
July 1 – 31	64	4.7%	95.1%	602	3.4%	94.4%	9.4
Aug 1 – 31	34	2.5%	97.6%	572	3.2%	97.7%	16.8
Sep 1 – 9	2	0.1%	97.7%	98	0.6%	98.2%	49.0
Non-resp	31	2.3%	100.0%	310	1.8%	100.0%	10.0
<b>Total</b>	<b>1,356</b>			<b>17,653</b>			<b>13.0</b>

**Table 3: Natural Gas -- Generators, Operating Status, 2004 Data**

Submission Date	Count	%	Cumulative %	Nameplate Capacity (MW)	%	Cumulative %	Avg. Capacity (MW)
< May 1	2,960	63.0%	63.0%	255,081	65.4%	65.4%	86.2
May 1 – 31	875	18.6%	81.7%	71,239	18.3%	83.4%	81.4
June 1 – 30	538	11.5%	93.1%	48,534	12.4%	96.1%	90.2
July 1 – 31	207	4.4%	97.5%	12,268	3.1%	99.3%	59.3
Aug 1 – 31	64	1.4%	98.9%	1,555	0.4%	99.7%	24.3
Sep 1 – 9	3	0.1%	99.0%	4	0.0%	99.7%	1.3
Non-resp	49	1.0%	100.0%	1,230	0.3%	100.0%	25.1
<b>Total</b>	<b>4,696</b>			<b>389,911</b>			<b>83.0</b>

**Table 4: Coal -- Generators, Operating Status, 2004 Data**

<b>Submission Date</b>	<b>Count</b>	<b>%</b>	<b>Cumulative %</b>	<b>Nameplate Capacity (MW)</b>	<b>%</b>	<b>Cumulative %</b>	<b>Avg. Capacity (MW)</b>
<b>&lt; May 1</b>	948	69.7%	69.7%	230,328	74.3%	74.3%	243.0
<b>May 1 – 31</b>	270	19.9%	89.6%	55,664	18.0%	92.3%	206.2
<b>June 1 – 30</b>	99	7.3%	96.8%	17,026	5.5%	97.8%	172.0
<b>July 1 – 31</b>	33	2.4%	99.3%	6,233	2.0%	99.8%	188.9
<b>Aug 1 – 31</b>	6	0.4%	99.7%	648	0.2%	100.0%	108.0
<b>Sep 1 – 9</b>	0	0.0%	99.7%	0	0.0%	100.0%	
<b>Non-resp</b>	4	0.3%	100.0%	73	0.0%	100.0%	18.3
<b>Total</b>	1,360			309,972			227.9

In perusing these four tables, it will be noticed that there was a steep drop-off in the response rate following July 1, 2005. For all fuels (as shown in Table 1), 11.9% of the respondents responded during the month of June while 5.5% responded during the month of July. In terms of volume (as measured by name plate capacity), the decline was from 9.5% in June to 2.8% in July. For August and until September 9, 2005, the steep drop-off continues with a 2.9% submission (and a 1.8% submission by volume). In Tables 2 – 4, it will be noticed that this phenomenon characterizes all of the major fuels studied. Non-hydro renewables had 19.7% of submissions (17.8% by volume) during the month of June, and only 4.7% of submissions (3.4% by volume) during the month of July. Similarly, natural gas fired generators showed an 11.5% (12.4% by volume) submission rate in June and a 4.4% (3.1% by volume) during the month of July, while coal fired generators showed a 7.3% (5.5% by volume) during the month of June and a 2.4% (2.0% by volume) during the month of July.

In studying the cumulative volume of generation, it will be noticed that overall 95.1% of volume has been submitted prior to July 1<sup>st</sup> with non-hydro renewables being characterized by a 91.0% cumulative response rate at that time with natural gas and coal being characterized by 96.1% and 97.8% response rates, respectively.

**Data Submission Profile for EIA-861 Respondents**

As in the case of the EIA-860, data were reviewed in a like manner for the EIA-861 to determine respondent submission dates. The four tables that follow display these findings. Table 5 provides a summary of Total Electricity Sales for all sectors combined. Data submissions are for the year 2004 and were received by EIA from January through December of 2005. Data for individual sectors (Residential, Commercial and Industrial) are shown in Tables 6-8.

**Table 5: Total Electricity Sales and Total Consumers, 2004 Data**

Submission Date	Respondent Count	%	Cum %	Sales (MWH)	% Sales	Cum %	Customer Count	% Customers	Cum %
<May 1	1,878	57.2%	57.2%	2,378,341,998	66.8%	66.8%	89,334,219	65.0%	65.0%
May 1-31	682	20.8%	78.0%	770,457,115	21.6%	88.5%	31,096,543	22.6%	87.6%
Jun 1-31	278	8.5%	86.5%	312,228,236	8.8%	97.2%	13,308,248	9.7%	97.2%
Jul 1-31	228	6.9%	93.4%	76,749,557	2.2%	99.4%	2,778,102	2.0%	99.3%
Aug 1-31	95	2.9%	96.3%	18,557,078	0.5%	99.9%	874,465	0.6%	99.9%
Sep 1-30	38	1.2%	97.5%	3,321,113	0.1%	100.0%	138,958	0.1%	100.0%
Oct 1-31	3	0.1%	97.6%	46,750	0.0%	100.0%	2,075	0.0%	100.0%
Dec-1-31	4	0.1%	97.7%	29,879	0.0%	100.0%	2,083	0.0%	100.0%
Undated Submission	5	0.2%	97.9%	27,085	0.0%	100.0%	3,792	0.0%	100.0%
Non-respns	71	2.2%	100.0%						
<b>Total</b>	<b>3,282</b>	<b>100.0%</b>	<b>100.0%</b>	<b>3,559,758,811</b>	<b>100.0%</b>	<b>100.0%</b>	<b>137,538,482</b>	<b>100.0%</b>	<b>100.0%</b>

**Table 6: Residential Electricity Sales and Consumers, 2004 Data**

Submission Date	Respondent Count	%	Cum %	Sales (MWH)	% Sales	Cum %	Consumer Count	% Consumers	Cum %
<May 1	1,878	57.2%	57.2%	857,300,144	65.9%	65.9%	77,866,543	64.9%	64.9%
May 1-31	682	20.8%	78.0%	284,240,228	21.9%	87.8%	27,338,106	22.8%	87.6%
Jun 1-31	278	8.5%	86.5%	125,152,879	9.6%	97.4%	11,566,250	9.6%	97.3%
Jul 1-31	228	6.9%	93.4%	24,589,214	1.9%	99.3%	2,384,890	2.0%	99.3%
Aug 1-31	95	2.9%	96.3%	7,658,615	0.6%	99.9%	767,097	0.6%	99.9%
Sep 1-30	38	1.2%	97.5%	1,433,697	0.1%	100.0%	119,611	0.1%	100.0%
Oct 1-31	3	0.1%	97.6%	19,107	0.0%	100.0%	1,553	0.0%	100.0%
Dec-1-31	4	0.1%	97.7%	17,680	0.0%	100.0%	1,767	0.0%	100.0%
Undated Submission	5	0.2%	97.9%	16,660	0.0%	100.0%	3,470	0.0%	100.0%
Non-respns	71	2.2%	100.0%						
<b>Total</b>	<b>3,282</b>	<b>100.0%</b>	<b>100.0%</b>	<b>1,300,428,224</b>	<b>100.0%</b>	<b>100.0%</b>	<b>120,049,287</b>	<b>100.0%</b>	<b>100.0%</b>

**Table 7: Commercial Electricity Sales and Consumers, 2004 Data**

Submission Date	Respondent Count	%	Cum %	Sales (MWH)	% Sales	Cum %	Consumer Count	% Consumers	Cum %
<May 1	1,878	57.2%	57.2%	831,479,097	67.3%	67.3%	10,952,291	65.4%	65.4%
May 1-31	682	20.8%	78.0%	262,730,171	21.3%	88.5%	3,649,159	21.8%	87.2%
Jun 1-31	278	8.5%	86.5%	110,920,440	9.0%	97.5%	1,643,531	9.8%	97.1%
Jul 1-31	228	6.9%	93.4%	23,003,764	1.9%	99.4%	377,242	2.3%	99.3%
Aug 1-31	95	2.9%	96.3%	6,390,748	0.5%	99.9%	96,355	0.6%	99.9%
Sep 1-30	38	1.2%	97.5%	1,525,654	0.1%	100.0%	17,968	0.1%	100.0%
Oct 1-31	3	0.1%	97.6%	8,915	0.0%	100.0%	428	0.0%	100.0%
Dec-1-31	4	0.1%	97.7%	12,199	0.0%	100.0%	316	0.0%	100.0%
Undated Submission	5	0.2%	100.0%	10,425	0.0%	100.0%	322	0.0%	100.0%
Non-respns	71	2.2%	99.8%						
<b>Total</b>	<b>3,282</b>	<b>100.0%</b>	<b>100.0%</b>	<b>1,236,081,413</b>	<b>100.0%</b>	<b>100.0%</b>	<b>16,737,612</b>	<b>100.0%</b>	<b>100.0%</b>

**Table 8: Industrial Electricity Sales and Consumers, 2004 Data**

Submission Date	Respondent Count	%	Cum %	Sales (MWH)	% Sales	Cum %	Consumer Count	% Consumers	Cum %
<May 1	1,878	57.2%	57.2%	684,396,743	67.3%	67.3%	514,893	68.6%	68.6%
May 1-31	682	20.8%	78.0%	222,656,266	21.9%	89.3%	109,017	14.5%	83.1%
Jun 1-31	278	8.5%	86.5%	75,114,248	7.4%	96.7%	98,198	13.1%	96.2%
Jul 1-31	228	6.9%	93.4%	29,129,298	2.9%	99.5%	15,966	2.1%	98.3%
Aug 1-31	95	2.9%	96.3%	4,507,715	0.4%	100.0%	11,013	1.5%	99.8%
Sep 1-30	38	1.2%	97.5%	361,762	0.0%	100.0%	1,379	0.2%	100.0%
Oct 1-31	3	0.1%	97.6%	18,728	0.0%	100.0%	94	0.0%	100.0%
Dec-1-31	4	0.1%	97.7%	0	0.0%	100.0%	0	0.0%	100.0%
Undated Submission	5	0.2%	97.9%	0	0.0%	100.0%	0	0.0%	100.0%
Non-respns	71	2.2%	100.0%						
<b>Total</b>	<b>3,282</b>	<b>100.0%</b>	<b>100.0%</b>	<b>1,016,184,760</b>	<b>100.0%</b>	<b>100.0%</b>	<b>750,560</b>	<b>100.0%</b>	<b>100.0%</b>

The structure of the Tables 5-8 includes respondent counts and percentages (both cumulative and by submission date) in columns 2-4. Columns 5-7 relate to volumes (sales in Megawatt hours). Columns 8-10 provide the number of customers reported by the respondents. The sum total of column 8 for each sector shows the total number of customers in that class.

The summary of Total Electricity Sales across all sectors is shown in Table 5. After the month of May, the response rate across sectors shows a significant decline. This pattern is quite similar to that of the EIA-860, which is discussed earlier. A drop-off in reporting begins with the month of June and is noted again in the month of July.

Approximately 78% of the total number of respondents submitted their data prior to June 1, reporting 88.5% of Total Electricity Sales. Subsequently, 8.5% of the total respondents reported 8.8% of Total Electricity Sales during the month of June. An additional 6.9% of the respondents

reported 2.2% of Total Electricity Sales during the month of July. This pattern of reporting appears to be common among the sectors and is examined closely by looking at Tables 6-8, which display the data for individual sectors. EIA received data submissions throughout the year, from January through December (with the exception of the month of November).

Note that Table 6, Residential Electricity Sales and Consumers, shows that data reporting 87.8% of the Residential Sales were submitted prior to June 1; while data reporting 9.6% were received during the month of June and 1.9% received during the month of July. Table 7, Commercial Electricity Sales and Consumers, illustrates a similar pattern: data reporting 88.5% of Commercial Electricity Sales were received prior to the first of June, 9.0% received during the month of June and 1.9% during the month of July. This pattern is again reflected in Table 8, Industrial Electricity Sales and Consumers. This table shows that data reporting 89.3% of Industrial Electricity Sales were submitted prior to the first of June; data reporting 7.4% were submitted during the month of June and 2.9% during the month of July.

Prior to June 1, most of the EIA-861 respondents are reporting for approximately 89% of the Total Electricity Sales. Therefore, it appears that June and July mark significant drop-off periods for respondent submissions, a pattern of reporting that is quite similar to the EIA-860.

### **Other Research Investigated**

Another area of investigation involved attempting to determine if the earlier submissions received were of higher quality than the later submissions. If this were indeed the case, an argument could be made that since the later data submitted were of lesser quality it might not be as worthwhile to collect it. However, it appears as though these data are not in a form to provide easy analysis, but they are available and could be analyzed with some extra effort if necessary.

Regarding EIA-860 data collections, there was also some question concerning whether or not there has been movement in the past from generators being ranked as one of the largest producers to becoming a lower ranked producer in terms its generation (as measured by name plate capacity). Upon investigation, it appeared as though this would be a difficult issue to investigate due to the fact that respondents can break their submissions in ways that would make it very difficult, if not impossible, to identify changes in such ranking on a year to year basis.

Regarding EIA-861 data collections, it is possible to analyze movement of utilities from being ranked as one of the largest providers of electricity to becoming a lower ranked provider. This could be investigated if thought to be a worthwhile endeavor.

### **Questions for the Committee**

1. Would it make sense to cover all of the larger producers or providers of electricity (on the EIA-860 and 861) with some rotation of surveying smaller producers/ providers while making sure that fuel coverage is sufficient (on the EIA-860)?
2. Would it make sense to utilize a cut-off date (e.g., June 30) to limit the number of non-response follow-ups for smaller producers/providers?

3. What other background analysis would you recommend in order to further illuminate pertinent issues?