

Revising Data Together Across EIA: Issues and Opportunities

Renee Miller and Alethea Jennings

Statistics and Methods Group, EIA

Abstract (with URL for reference): We have spoken to the Committee (see item 9 of <http://www.eia.doe.gov/calendar/sum5.htm>) about how EIA has expanded its use of data collected in electric power surveys across the organization, such as natural gas, petroleum, and the integrated statistics publications. As a result, revisions to electric power data affect more than just the electric power publications. In addition, it is easier to revise data on the Web than in hard copy format. These two events have raised questions about the need to coordinate revisions across EIA.

We will present background on EIA's current revision standard (which the Committee helped us develop in the eighties) and recent events that led to its review. Representatives from each affected office are working together to determine the best way to meet user needs and interoffice goals in support of EIA's mission. We will present conclusions from recent discussions/debates and will ask for the Committee's guidance on coordinating revisions.

In addition to thinking about the coordination of revisions, EIA is examining the circumstances under which data are revised. In our attempt to reduce the frequency of publishing revised data, we are examining the situations and events that prompt data revisions. EIA may be revising data more often than is necessary to meet user needs. Various situations cause data to revise, such as resubmissions, benchmarking (to the Annual publication), data corrections (where problems were discovered with the data originally reported) and late submissions. We will present an example of data to illustrate revisions occurring over time. We hope to obtain suggestions and ideas for reducing the number of revisions that we publish.

To satisfy users who are interested in having the latest data available, it has been suggested that EIA provide the latest versions of the data via the Website and inform users of this option. We will discuss some of the advantages and disadvantages of this approach and ask for the Committee's thoughts.

Attachment A presents background on EIA's revision standard, current revision practices, and a recent event that led to a review of our revision practices. It also summarizes themes from discussions that took place last fall. Attachment A is a package previously prepared for discussion with Office Directors Nancy Kirkendall, Ken Vagts and Scott Sitzler entitled, "Revising Together." The paper has three attachments including the current revision standard.

Attachment B presents an update on discussions on revisions.

Attachment C presents data on residential natural gas deliveries for 2002 as first published in the *Natural Gas Monthly* in April 2002 and shows changes through the January 2004 *Natural Gas Monthly*. Data are examined at the State level to determine what triggered the revisions at the national level.

Attachment D summarizes the advantages and disadvantages of what has become known as showing the "latest/greatest" data on the Web.

Attachment E contains our questions for the Committee.

Revising Together

Now that EIA is using electric power data in natural gas, petroleum, and the integrated publications, revisions to electric power data affect more than just the electric power publications. In addition, it is easier to revise data on the Web than in paper publications. These two events have raised questions about EIA's revision policy. This paper summarizes EIA's current revision policy, describes what we are actually doing and a recent event that led to review of the revision standard, and presents conclusions from recent discussions on revisions. It serves as background for future discussions.

What is EIA's Revision Policy? EIA has a revision standard that states that we should establish a schedule for anticipated revisions and make it available to users. We should not plan more than two revisions of the same survey cell. The standard allows for unscheduled revisions due to undetected errors. It states that the threshold criteria be defined ahead of time and gives an example of a threshold; i.e. the revision will change the national level marginal total by more than 1 percent or a cell value by more than 5 percent. Unscheduled revisions may also be made at the discretion of the office director. (Attachment 1 presents the current revision standard).

What Are We Actually Doing? For the monthly data, we all publish a preliminary estimate and then a final one at the end of the 12-month processing cycle but we differ in what goes on in between. Petroleum Supply and Marketing have planned revisions. Electric Power and Natural Gas do not, but allow for the possibility of revisions if there are errors (or changes in data source) that they consider significant (Attachment 2 provides details).

Recent Events. There was a revision to the April 2001 value for natural gas consumption for electric power that the *Electric Power Monthly* (EPM) and the *Monthly Energy Review* (MER) picked up, but that the *Natural Gas Monthly* (NGM) didn't due to the processing cycle in Natural Gas Division. It was small (the numbers differ in the last digit), but raised the issue of consistency in making revisions. At a meeting of the Inter-office Issues group, a comment was made that perhaps Electric Power shouldn't have revised and that we should talk about when we should make revisions. We decided that we would form a subgroup to work on this.

The subgroup¹ reviewed the current EIA revision standard and concluded that if the threshold for making a revision suggested in the current standard had been followed, this particular revision would not have been made. Nevertheless the subgroup realized that the problem with not all publications making a revision at the same time could have occurred if the revision had been larger and sought ways to clarify and strengthen the standard. Ron O'Neill had cautioned against making too many changes since this standard was recently developed and went through a rigorous and painful review process. From the initial response we got to our proposed changes to the standard, we saw his point and made suggestions for only a few changes that we thought would be clarifications. We also sought input from others: Scott Sitzer, Bob Schnapp and Betsy O'Brien from the Office of Coal, Nuclear, Electric and Alternate Fuels, Nancy Kirkendall of the Statistics and Methods Group, Beth Campbell from the Office of Oil and Gas, and Kitty Seiferlein from the Office of Energy Markets and End Use.

¹ Lou Schloss (CNEAF), Julia Hutchins (EMEU), Roy Kass (Natural Gas), Bill Horvath (NEIC), Ron O'Neill participating by e-mail (Petroleum) and Renee Miller (SMG).

Conclusions From Discussions. Several themes emerged from discussions. Renee summarized them as follows and presented them to the Inter-office Issues Group on October 1 (they don't necessarily represent consensus).

1. Electric Power needs to close first and stay closed.
2. Revisions after data are declared final should be harder to make than revisions to data that haven't been declared final.
3. If there is a catastrophic error and annual or "old" data need to be revised, petroleum and natural gas need the mechanism to make these revisions.
4. Establishing a revision schedule and making it available to both internal and external users is key.
5. Unscheduled revisions should be for catastrophic errors. In developing thresholds for what is a catastrophe, we need to clarify what is a "cell" value and what is a national level "marginal" total.
6. Perhaps we should concentrate on the series that we all use, fuel consumption for electric power sector and commercial/industrial combined heat and power plants.

At the October 1 Inter-office Issues meeting, the group agreed that distinguishing a "cell" from a "marginal" total was not always straightforward, depending on how the table was set up and that moving away from relying on thresholds as triggers for making revisions and moving towards schedules was desirable.



EIA REVISION STANDARD

ENERGY INFORMATION ADMINISTRATION STANDARD 2002-13

Title: Revisions

Superseded Version: 88-05-02

Purpose: To provide EIA customers with information about revisions in disseminated data.

Applicability: All EIA information products.

Required Actions:

1. Establish a schedule for anticipated revisions and make it available to users. Do not plan more than two revisions of the same survey data cell.
2. Scheduled revisions - The first dissemination of a data value in an information product should be identified as "preliminary" if revisions are anticipated in a subsequent dissemination. Scheduled revisions to these values should be identified as "revised" (or "final") the first time the changes are disseminated.
3. Preliminary and revised data must be identified through means such as data value labeling (i.e., data marked "P" for preliminary, and "R" for revised), or text in the product title, table titles, headers or footnotes, or other text accompanying the product.
4. When unscheduled revisions are required due to previously unrecognized errors or respondent resubmissions, develop threshold criteria (e.g., the revision will change the national level marginal total by more than one percent or a cell value by more than five percent) ahead of time to identify conditions under which revisions will be made.
5. In data cells with previously unrecognized errors that do not meet the established threshold criteria, unscheduled revisions may be made at the discretion of the sponsoring Office Director.
6. Data values changed as a result of unscheduled revisions must be clearly identified (see item 3 above) and the reasons for these changes should be communicated to data users.
7. Do not disseminate information if "errata sheets" are anticipated.

Related Information:

1. [Standard 2002-7](#), Response Rates and Imputation

Approval Date: September 26, 2002

Summary of Revision Procedures

Monthly and Quarterly Estimates

All publish a preliminary estimate and then a final one at the end of the 12-month processing cycle. They differ in what goes on in between. Petroleum Supply and Marketing have planned revisions. Electric Power and Natural Gas do not, but allow for the possibility of revisions if there are errors (or changes in data source) that they consider significant. Coal does not revise the quarterly data.

Petroleum Supply

1st figure published: estimate from the weekly survey, called “Monthly-from-weekly”

2nd figure published: estimate from the monthly survey

Final: published at the end of the processing year

In between: If there are changes due to resubmissions, they are shown in Table C1 of the *Petroleum Supply Monthly*, but revised estimates are not shown. In this way, users can compute the revised estimates but EIA is not officially revising the data.

Petroleum Marketing

1st figure published: forecast for select cells

2nd figure published: estimate from monthly survey

3rd figure published: revised estimate from monthly survey

Final: published at the end of the processing year

Electric Power

Revisions are possible after the first monthly figure is published and before the end of the processing cycle if there are major errors that effect the national total.

Natural Gas

Revisions are possible after the first monthly figure is published and before the end of the processing cycle under different circumstances for different series. For production, revisions are possible when reported numbers are received; reported numbers supercede estimates. For consumption, there is a revision if current numbers differ from published numbers by more than 5% and the current number is not suspect. For storage, if large differences in either inventory or activity are reported, results are candidates for revision. Another trigger for a revision for consumption is if “NA’s” are removed.

Annual Estimates

Petroleum annual estimates are not usually revised. An exception is if there is a change in an alternate data source.

Natural gas annual estimates are revised if there is confidence in more recent information and the difference is meaningful. Revisions to data older than a year are possible.

Coal and Electric Power make corrections to data published as final if there is a difference of one percent or greater at the national level. Differences under one percent are left to the discretion of the Office Director.

Attachment B

Update on Discussions on Revisions

EIA's Inter-office Issues Group has continued discussing how to coordinate revisions. Following is a summary of the issues that have arisen.

What Other Agencies Do

We obtained information about 2 surveys that are similar to EIA surveys in that they are monthly surveys of establishments. The Bureau of the Census and the Bureau of Labor Statistics conduct the surveys. In both situations the agencies publish a subset of the data first and then present revisions of this subset along with data for other categories. Census revises the data two additional times and then freezes until data are available from their annual survey to benchmark. They perform an additional benchmark when data from their Economic Census become available. So, Census can present 6 versions of a number. Bureau of Labor Statistics does something similar but did not mention benchmarking in our discussions with them. It sounded as if they could have 3 versions of a number.

It was pointed out that our situation differs from Census' in that we publish facility-level data, whereas they do not publish at that level. Users can more easily observe problems with individual submissions and then if we correct the individual data it affects the aggregates. The argument that was being made here was that it might be more difficult for us to limit the number of revisions than it is for Bureau of Labor Statistics and the Census Bureau. But, it was noted that revisions were not costless because there tends to be a ripple effect.

Use of the Annual Energy Review and Other Integrated Statistics Publications as a Vehicle to Show Revised Data

Recently a user called to inform EIA staff that the 1999 kerosene data for Kansas shown in the *State Energy Data Report* looked like an outlier. A petroleum analyst checked into it and found that there was a reporting error, but since Petroleum Division doesn't revise once data have been declared final, they didn't republish the number. We then tried to address the question: Could the *State Energy Data Report* and the *Annual Energy Review*, which present long historical series, be used as a vehicle for revisions that cannot be made in other publications due to their revision policy.

We did not come to consensus on this issue. One participant thought that this approach could be an interim solution, but not necessarily the ideal solution and that perhaps we could come up with something better using the Web as a vehicle. We also talked about what would best serve our customers and whether there has been a storm of protests from users when we do not revise erroneous data. Experiences differed. Petroleum analysts did not hear complaints from users, but members of the Integrated Statistics staff did. The latter group also noted that their customers have not complained about the number of revisions EIA has made, but do complain when numbers are not consistent from one section of a publication to another.

Showing the Latest

Petroleum Supply Division staff members showed us how they limit the number of revisions they publish and also present information on resubmissions. They publish preliminary data, then revise the following month and then don't revise until the end of the year. They show resubmissions that occur during the year in Table C1 (http://www.eia.doe.gov/pub/oil_gas/petroleum/data_publications/petroleum_supply_monthly/current/pdf/tablec1.pdf). But, if errors are found in back years, Petroleum Supply doesn't revise.

Natural Gas and Electric Power Division staff members have pointed out that there is a difference between their data and the petroleum supply data. The petroleum supply data are based on censuses while some of their data are based on sample surveys. Since sample data cannot simply be tallied to obtain totals (weights and estimation procedures are required), they thought that publishing resubmissions would be problematic.

An Emergency Change Control Board: Can it Work for Revisions?

A participant who works in the Office of Information Technology said that in listening to comments made by the group the situation sounded similar to infrastructure maintenance. Infrastructure maintenance for information technology is usually scheduled and a two-week notice is given. They have an Emergency Change Control meeting when something comes up which is too important to wait. The participant thought we could do the equivalent. The emergency change control group could be the Inter-Office Issues Group or some subset that includes the major players. It should be composed of a representative from each affected Office. We wouldn't try to define what triggers the group to convene, but the idea would be if there were an unscheduled revision needed we would get the group together, make a decision on it and tell people what happened.

We are working on how to implement this idea. The group thought that addressing the issue of reopening data that have been finalized is where a Change Control group would be most beneficial. If someone wanted to make a revision after data are declared final, they would have to make a case for it. We are in the process of working out the details.

An Example of a Revision Trail: Residential Natural Gas Deliveries

Table C1 shows data representing residential natural gas deliveries for each month in 2002 as presented in the *Natural Gas Monthly (NGM)*. It tracks changes in the data through the January 2004 publication, when the data were benchmarked to the annual survey. For example, this table presents data for January 2002, shown initially in the April 2002 publication through the January 2004 publication.

The table illustrates that for residential natural gas deliveries EIA published a considerable number of revisions for the January 2002 reporting period. The data were revised several times over the next few months following its initial publication in April 2002. In total, the data were revised ten times during the twenty-two month period (the April 2002 through January 2004 publications).

A closer look at the estimate for January 2002 illustrates a common occurrence where the revised data often “mirror” what was shown in previous publications. This means that on a number of occasions the revised data match revisions that were submitted earlier during the twenty-two month period of review. This is apparent when comparing data corresponding to the following publication dates:

Jun-02/Oct-02 (^R821)
Jul-02/Sep-03 (^R820)
Aug-02/Dec-03 (^R819)

This calls attention to the fact that data at the national level often revise back to previously published values.

The national level data are not the only data of interest to our users. EIA also presents State-level data, of which the national total is comprised. When national-level data revise back to previously published values, it is due to state-level changes that may not have reflected the same bounce-back phenomenon. Also, because new sample members (respondents) are introduced in January, the January estimates are more likely to revise. A changing sample at the start of each year introduces this type of start-up phenomenon, which is common because of the time it takes to train new respondents and to resolve existing response issues. The apparent effect of the state-level components on the national total prompted a review of the State-level data to determine the impact of these resubmissions on the national total.

“NA” is published for this data series when the data reported at the State level fall below a predetermined percentage of total volume for that State (i.e., the decision to publish data at the State level is based on the percentage of total volume reported by respondents). The lack of coverage is typically due to late submissions from respondents. We wanted to identify how often a change from “NA” to actual reported data caused a change in the national total for that month. The data revealed that during the twenty-two month period of review there were only three occasions (May 2002, September 2002 and April 2003 publications) when revisions to the national totals for the January 2002 reporting period were caused by changes at the State level where “NAs” were replaced by actual data. Where NAs are shown, there are State estimates based on imputed values that are incorporated into the national total. These values are hidden behind the NAs. In

many cases the imputed values are very close to the actual reported data, so removal of the NA does virtually nothing to the previously hidden state estimate. Thus, no revision is generated.

In considering State-level revisions, it is important to consider another factor. For States where there was enough coverage to publish data initially, there could be resubmissions or additional respondents reporting. Table C2 shows the actual changes in volume as submitted by State in the instance of both resubmissions and receipt of late submissions. The percent change column highlights the change in total volume for that State. Some of the revisions to previously published data that are submitted at the State level amount to ten or more percent of the total volume for that State, a change that we think would be considered significant to users of the State data. When the State data were revised, EIA revised the national total even though in some instances it did not result in a substantial change. This is done so that the national total will be the sum of the State components.

Suppose we were interested in presenting updated data for our State-level users while reducing the number of revisions for our national-level users. Would it ever be desirable to have the National total be different from the State total, even if users wouldn't see the difference (because the data would be hidden behind the NAs)?

As a data user, do you think there is a problem with our current revision pattern? It may have been a problem when people had to key in revised data, but in a world of downloads is this an issue?

Table C1
Natural Gas Deliveries in the United States
Source Publication: Natural Gas Monthly, Table 3

(Billion Cubic Feet)

Residential Sector

NGM Publication Date	Data Reporting Period for 2002											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Apr-02	829	-	-	-	-	-	-	-	-	-	-	-
May-02	R 818	696	-	-	-	-	-	-	-	-	-	-
Jun-02	R 821	R 703	660	-	-	-	-	-	-	-	-	-
Jul-02	R 820	703	660	417	-	-	-	-	-	-	-	-
Aug-02	R 819	R 706	R 666	R 418	259	-	-	-	-	-	-	-
Sep-02	R 817	R 704	R 665	418	259	164	-	-	-	-	-	-
Oct-02	R 821	704	R 666	R 419	259	164	128	-	-	-	-	-
Nov-02	821	704	666	419	259	164	128	118	-	-	-	-
Dec-02	821	704	666	419	259	164	128	R 117	125	-	-	-
Jan-03	821	704	666	419	259	164	128	117	125	250	-	-
Feb-03	821	704	666	419	259	164	128	117	125	250	490	-
Mar-03	821	704	666	419	259	164	128	117	125	250	R 489	773
*Apr - 03	821	704	666	419	259	164	128	117	125	250	489	773
May-03	R 823	R 706	666	419	R 258	R 163	R 127	117	125	250	489	R 771
Jun-03	823	R 707	666	R 421	R 259	163	R 128	117	125	R 251	R 490	R 769
Jul-03	823	707	666	421	259	163	128	117	125	251	490	769
Aug-03	823	707	666	421	259	163	128	117	125	251	490	769
Sep-03	R 820	R 717	R 665	R 420	R 258	163	128	117	125	251	R 487	769
Oct-03	820	R 718	665	R 417	R 256	R 161	R 127	117	125	R 252	R 484	R 773
Nov-03	820	718	665	R 416	R 255	161	R 125	117	R 124	R 251	484	773
Dec-03	R 819	R 717	665	416	255	161	125	117	124	251	484	R 772
Jan-04	R 816	R 713	R 661	R 415	255	R 160	125	R 116	124	251	R 483	R 771

TABLE C2

Residential Natural Gas Deliveries to Consumers - January 2002 Reporting Period				
State Level Revisions that Impact the National Total				
Natural Gas Monthly Publications from May 2002-December 2003 ¹				
(Million Cubic Feet)				
Publication Date	State	Data in Previous Publication	Data in Current Publication	Percent Change from Previous Month
May-02	WA	NA	11306	-
Jun-02	CT	3893	6197	59
Jul-02	ID	3735	3450	-8
	NV	6793	5871	-14
Aug-02	WA	11306	10931	-3
Sep-02	MD	NA	12872	-
	OK	NA	12761	-
Oct-02	NM	5256	6124	17
	PA	37202	40182	8
Apr-03	AR	NA	7944	-
	RI	NA	2925	-
	WV	NA	5912	-
May-03	GA	21969	22343	2
	MD	12872	12873	0
	NY	56231	57791	3
	OH	53775	54013	0
	PA	40182	39679	-1
Jun-03	AR	7944	7216	-9
	LA	8322	9300	12
Sep-03	MS	5860	5461	-7
	TX	42635	39565	-7
Oct-03	LA	9300	9833	6

¹ Taken from Natural Gas Monthly Publications (April 2002 through December 2003), Table 15, Natural Gas Deliveries to Consumers by State, 2001-2003.

Attachment D

Showing the Latest/Greatest on EIA's Website

Several times during discussions of revisions, better use of the Web has been suggested. Say, for instance, that EIA adopts the practice of scheduling revisions. Using the Web, there could be the opportunity to show resubmissions (and/or retabulated data based on the resubmissions) at more frequent intervals than the scheduling would allow. Thus, we would have a "latest/greatest" page comprised of the latest data available in addition to our official data based on the scheduled revisions.

Advantages

1. Errors would be corrected.
2. Users would have the latest data; EIA would not have to scramble to produce it when it was requested.

Disadvantages

1. The latest may not actually be the greatest. See, for example, the attached Table D1 showing the fluctuations in our natural gas price data for the electric power sector. This series is relatively new and probably is revised more often than most series, but is an example of how we can revise and then end up where we started.
2. Maintaining the site could be an added burden to EIA staff.

TABLE D1

**Natural Gas Prices: Electric Power Sector
Source Publication: Monthly Energy Review Table 9.11**

(Dollars per Thousand Cubic Feet)

MER Publication Date	Data Reporting Period for 2002 ^d											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Apr-03	R 3.13	R 2.87	R 3.38	R 3.81	R 3.82	R 3.62	R 3.50	R 3.43	R 3.73	R 4.20	R 4.41	R 4.77
May-03	R 3.11	R 2.86	R 3.35	R 3.78	R 3.80	R 3.59	R 3.48	R 3.41	R 3.72	R 4.19	R 4.38	R 4.76
Jun-03	3.11	2.86	3.35	3.78	3.80	3.59	3.48	3.41	3.72	4.19	4.38	4.76
Jul-03	R 3.13	R 2.87	R 3.38	R 3.81	R 3.82	R 3.61	R 3.50	R 3.43	R 3.72	R 4.20	R 4.41	R 4.76
Aug-03	3.13	2.87	3.38	3.81	3.82	3.61	3.50	3.43	3.72	4.20	4.41	4.76
Sep-03	3.13	2.87	3.38	3.81	3.82	3.61	3.50	3.43	3.72	4.20	4.41	4.76
Oct-03	R 3.10	R 2.86	R 3.38	R 3.80	R 3.78	R 3.61	R 3.49	R 3.43	R 3.72	R 4.19	R 4.35	R 4.72
Nov-03	R 3.13	R 2.87	R 3.38	R 3.81	R 3.82	R 3.61	R 3.50	R 3.43	R 3.72	R 4.20	R 4.41	R 4.76
Dec-03	R 3.13	2.87	3.38	3.81	3.82	3.61	3.50	3.43	3.72	4.20	4.41	4.76

^d The electric power sector is comprised of electricity-only and combined heat-and-power (CHP) plants with the NAICS 22 category whose primary business is to sell electricity, or

Attachment E

Questions for the Committee

1. When EIA developed its revision standard in the early 1980s, we were trying to establish our credibility and thought that limiting the number of revisions was a good practice. Is this still a good assumption? What is the Committee's point of view on limiting the number of revisions?
2. Specifically for the natural gas data for the residential sector, does the Committee have suggestions for another approach for presenting revisions?
3. What does the Committee think about showing "the latest/greatest" on the Web?