

Estimating Reporting Burden for Statistical Surveys

Anna Hamelin

U.S Energy Information Administration, 1000 Independence Ave. SW, Washington, DC 20585

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research and Policy Conference

Abstract

Accurately estimating burden per response on establishment surveys is a difficult task. According to the Paperwork Reduction Act regulations in 5 CFR 1320(b)(1), nine components of burden need to be evaluated for estimating reporting burden. Some of the components are easily understood while other components are no longer applicable or confusingly similar to other components. Respondents are able to provide quantifiable responses to some overarching components while other discrete subcomponents are not easily measurable and result in respondent confusion. This paper discusses the challenges that arise from attempting to measure the components of reporting burden and how to accurately estimate burden per response in an electronic data collection environment.

Introduction

Federal agencies are required to provide respondents with an estimated time it takes to complete a statistical survey.ⁱ The Office of Management and Budget (OMB) guidelines in 5 CFR 1320.3, defines the components for estimating the total time, effort, and financial resources expended by persons to report on statistical surveys.

“(1)**Burden** means the total time, effort, or financial resources expended by persons to generate, maintain, retain, or disclose or provide information to or for a Federal agency, including:

- (1) Reviewing instructions;
- (2) Developing, acquiring, installing, and utilizing technology and systems for the purpose of collecting, validating, and verifying information;
- (3) Developing, acquiring, installing, and utilizing technology and systems for the purpose of processing and maintaining information;
- (4) Developing, acquiring, installing, and utilizing technology and systems for the purpose of disclosing and providing information;
- (5) Adjusting the existing ways to comply with any previously applicable instructions and requirements;
- (6) Training personnel to be able to respond to a collection of information;
- (7) Searching data sources;
- (8) Completing and reviewing the collection of information; and
- (9) Transmitting, or otherwise disclosing the information.”

However, each component of burden is not easily quantifiable or capable of measurement. For example, most data systems collect, process, and validate information as a single function so breaking this process into two separate steps is difficult to measure. Research shows that the elements that are described in 5 CFR 1320.3 may not be able to be accurately measured as discrete and separate elements.

Background

The U.S. Energy Information Administration (EIA) conducted three cognitive research studies to estimate reporting burden for three surveys in preparation for submitting Information Collection Requests to OMB for review in 2017. We conducted three studies discussed in this paper to test the effectiveness of using an expanded protocol versus a simplified version to measure the time it takes respondents to complete EIA survey forms.

Data and Methods

The results from 83 cognitive interviews across 3 cognitive studies were analyzed and compared to identify useful approaches for estimating the burden per response for respondents completing and filing a statistical survey with EIA. The three cognitive research projects are useful examples that show the problems with attempting to estimate each element as a discrete unit. They also show some useful approaches on how to collapse elements to generate accurate estimates for total reporting burden. These approaches were developed through a series of protocols for three survey instruments that collected information on the components of burden outlined in 5 CFR 1320.3. Each of the three protocols had a unique way to ask the burden questions and each delivered different results. The three energy surveys studied vary in frequency, energy source, and type of companies required to report. The studies

included Form EIA-63C, “Densified Biomass Fuel Report;” Form EIA-886, “Annual Survey of Alternative Fueled Vehicles” (AFV); and Form OE-417, “Electric Emergency Incident and Disturbance Report.”

Survey Methodology

Form EIA-886 is an annual survey that collects data on the number of alternative fuel vehicles (AFVs) made available, the distribution of AFVs in use, and alternative transportation fuels (ATFs) consumed. The cognitive study for Form EIA-886 was designed to verify the reporting burden estimate of 4 hours as well as assess the respondents’ ability to report data on a new set of questions being proposed. The companies interviewed ranged from small businesses that had two AFVs to companies with fleets over a 1,000 AFVs. Participants interviewed were broken into six different groups, state governments, local governments, fuel providers, transit agencies, private fleets, and Original Equipment Manufacturers (OEMs). In this study, 51 participants were asked to answer three separate questions on how much time it took to gather the data, fill out the form, and then the total time it takes them to complete the form. They were also asked how long it would take to report information on the newly proposed data elements. The responses to the burden questions were complex and hard to aggregate because the respondents varied greatly by respondent type, size, and types of data systems used to complete the form.

Form OE-417 is an emergency survey form that collects information on electric emergency incidents and disturbances for the U.S. Department of Energy’s (DOE) use in fulfilling its overall national security and other energy management responsibilities. The cognitive study for Form OE-417 was designed to verify the reporting burden of 2.16 hours for one homogenous group of power producers. The participants varied by size, market share, and frequency. The frequency of this report varies based on disturbances to the electric grid. This study asked 20 respondents to answer four questions on the burden to gather the data, fill out the form, how long they spend on training and the total time to report their information. Participants were also asked a follow-up question on if they had to work with others.

Form EIA-63C is a monthly survey that collects information on pellet fuel and other densified biomass fuel consumption, production, sales, and inventory levels in the United States. The data collected are used to estimate densified biomass fuel production, sales, and inventory at state, regional, and national levels. The cognitive study for Form EIA-63C was designed for another homogenous group to verify the current reporting burden estimate of 1.5 hours per response based on responses from 12 participants. This study used an expanded burden protocol incorporating seven separate questions to attempt to discretely measure the components of reporting burden shown in 5 CFR 1320.3.

Results

The results from these three studies revealed that not all of the components are able to be estimated and that some of the questions were confusing to the participants. The separate components pertain to the generic survey process but may not pertain to the reporting requirements for every survey. The components are identified by OMB as guidance and should be used, when relevant, to validate and/or comprehensively estimate the total burden per response for responding to a survey.

Reviewing instructions

The first component of reporting burden is reviewing instructions. Participants are able to easily understand and provide an estimate of how long it takes them to review instructions. When participants were asked how long they spend reading the instructions during the cognitive interviews for Form EIA-63C, 9 out of 12 participants said that they did not review the instructions. Three of the participants said that it takes between 1 to 5 minutes to review the instructions. These results showed that most participants do not read the instructions. Form EIA-63C is a monthly survey so respondents are familiar with the survey and do not need to reference the instructions each month they fill out the form.

Developing, acquiring, installing, and utilizing technology and systems for the purpose of collecting, validating, and verifying information; purpose of processing and maintaining information; and purpose of disclosing and providing information

OMB components (2), (3), and (4), developing, acquiring, installing, and utilizing technology and systems for the purpose of collecting, validating, and verifying information; the purpose of processing and maintaining information; and purpose of disclosing and providing information were very confusing to participants. Many participants believed these components were similar and mean the same thing. The participants viewed developing, acquiring, installing

and using software as one continuous process without any defined measurable points in time. During the cognitive interviews for Form EIA-63C, EIA asked the respondents, “How much time does it take you to utilize your software or technology to gather the information from your data systems?” The participants had difficulty providing quantifiable responses. Some did not understand the differences between the questions and the responses varied from zero time to 3 hours. Another reason participants were confused was because they viewed utilizing technology as a part of their daily duties, and they did not consider it a separate step in completing a survey form.

Adjusting the existing ways to comply with any previously applicable instructions and requirements

Asking a participant to estimate the amount of time to adjust the existing ways to comply with any previously applicable instructions and requirements is another challenging component for participants to provide a quantifiable response to. In all three studies, we did not ask the participants this question directly. Instead, the participants were asked two questions. First, participants were asked if they are able to directly access the data needed from company records to report on this form, or if they work with others to gather this data. Second participants were later asked how long they have filled out the survey form. These two questions provide an approach for estimating OMB component (5) of adjusting the existing ways to comply with any previously applicable instructions and requirements.

Training personnel to be able to respond to a collection of information

During the cognitive interviews for Form EIA-63C, we asked if the participants had received any training. Only 3 out of the 12 had received training. The 3 participants stated that this training covered how to gather the data to fill out Form EIA-63C. Participants to Form OE-417 were able to provide an estimate for training time. A majority of participants interviewed said that their companies had practice drills simulating an electric disturbance, so they could practice filling out the form because the turnaround time for this form is an hour. Not all surveys have a training component. Form OE-417 is unique because regulatory rules require every company to provide 2 hours of training for their employees. For many surveys, it is hard for a respondent to quantify how long someone had to be trained for and if it was specifically to fill out EIA survey forms. With current technology and electronic data collection methods, training respondents is not as large of a component of burden for every survey. Participants may initially spend more time learning how to use software for the survey form, but then quickly become proficient as a regular user of an electronic data collection system in future collection cycles reducing the training time needed.

Searching data sources

Asking a participant to estimate the amount of time it takes to search data sources is an easy component to quantify when asked alone. This is a straightforward question that relates to a clear process that participants understand. During the cognitive interviews for all three studies, when participants were asked how long it takes to gather their data they were easily able to provide an estimate of the amount of time it takes them.

During the cognitive interviews for Form EIA-886, most participants were able to provide an estimate on how long it took to search their data sources. Table 1 shows the responses on how much time was spent gathering the data necessary to complete Form EIA-886. The responses ranged from 15 minutes to 40 hours. The weighted average time it took to gather the data was 2.7 hours. Of the 51 participants, 26 stated they needed less than two hours to gather the data for the reports, 14 participants reported they spend between 2 to 4 hours gathering data, and 2 participants reported they spend between 8 and 8 hours gathering data. The remaining 6 participants indicated they spend more than 8 hours gathering data.

Table 1: Burden to Gather Data for Form EIA-866 (n=51)

Number of Participants	Amount of Time	Percentage
26	<2hours	51.00%
14	≥2-4 hours	27.50%
2	>4-8 hours	3.90%
6	>8 hours	11.80%
3	Do not know	5.90%

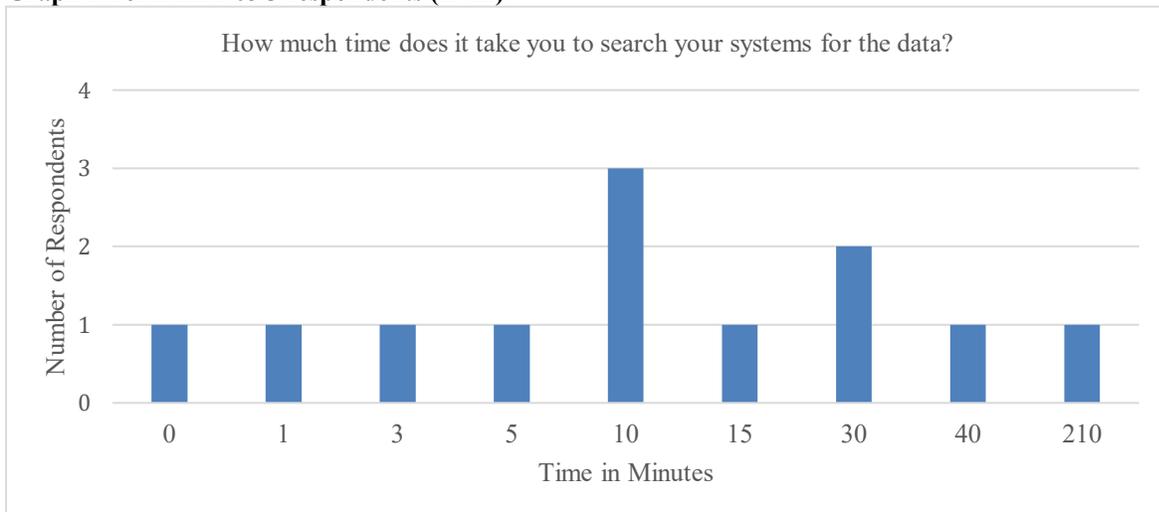
During the cognitive interviews for Form OE-417, EIA also asked the participants to estimate the amount of time it takes to gather the information to complete Form OE-417. Table 2 shows the burden estimates provided by the 20 participants. Only two participants estimated that it takes more than 2 hours to gather the information. The responses to this question had a smaller variation among the participants compared to the interviews for Form EIA-886.

Table 2: Burden to Gather Data for Form OE-417 (n=20)

Number of Participants	Amount of Time	Percentage
11	< 1 hours	55%
7	≥ 1-2 hours	35%
2	> 2 hours	10%

During the interviews for Form EIA-63C, when participants were asked how long it took them to search their data sources, the results varied substantially. This may have been caused by confusion among the participants about the burden components. Prior to asking this question, participants were asked many other questions about reporting burden and they may have believed they had already answered this question. They were not able to differentiate this question from the other questions asked. Graph 1 shows the reported burden estimates ranging from taking no time to 210 minutes.

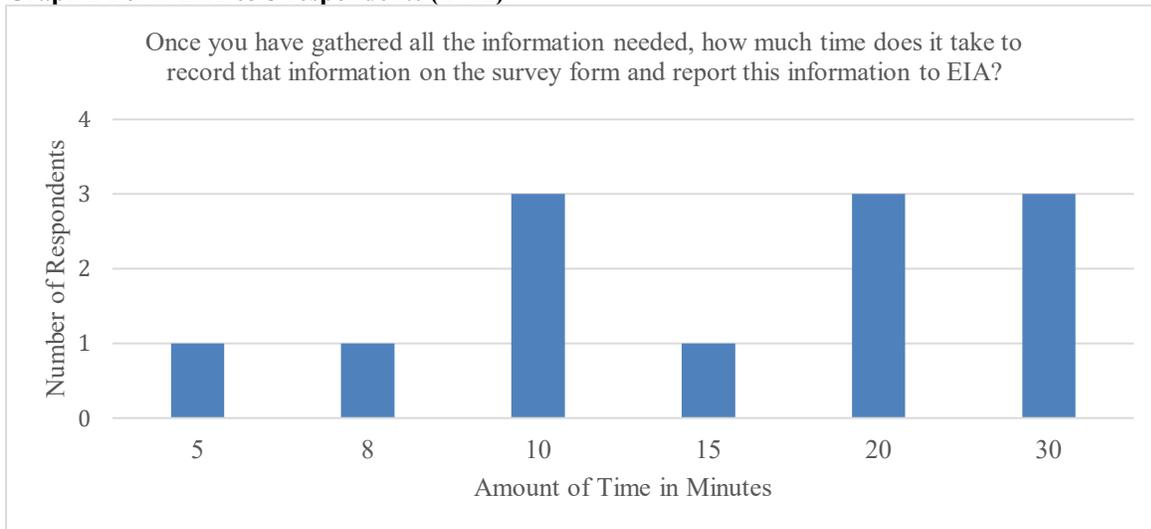
Graph 1 Form EIA-63C respondents (n=12)



Completing and reviewing the collection of information

During the cognitive interviews for Form EIA-63C, when we asked participants the burden of completing and reviewing the collection of information, they were able to provide an accurate estimate. Graph 2 shows that Form EIA-63C participants understood this question and were able to give a quantifiable response. This question was also asked during the cognitive interviews for Form’s EIA-886 and OE-417 and produced accurate and easy to understand results. Asking the participants how long it takes them to complete and review their collection of information is a very important component of burden.

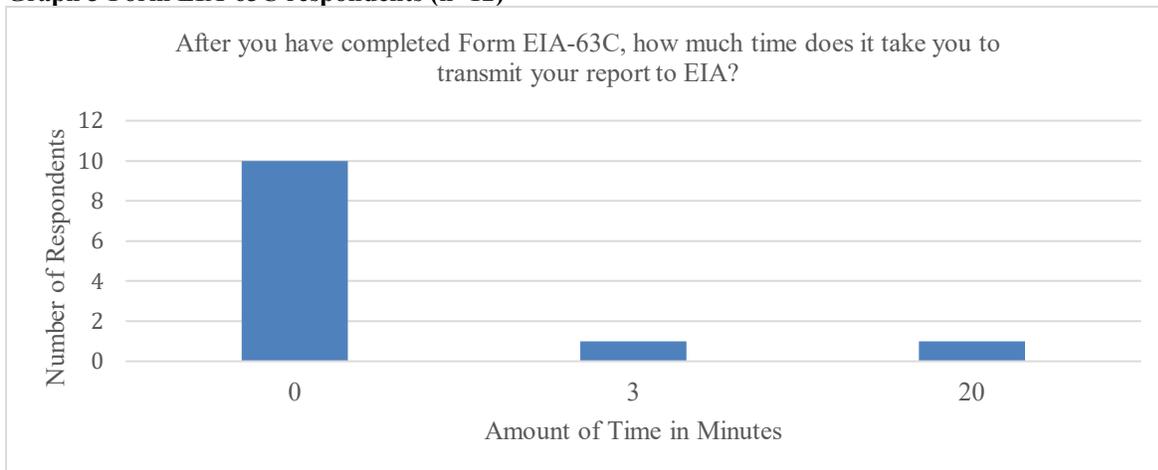
Graph 2 Form EIA-63C respondents (n=12)



Transmitting, or otherwise disclosing the information

Measuring the burden component of transmitting, or otherwise disclosing the information is less relevant in an electronic data collection environment. In the past, more time was spent filing and sending a hardcopy report than filing electronically. Currently, many respondents report electronically and they simply hit the “submit” button on their keypad. The participants who report via paper or excel sheet also do not think about the time that it takes to transmit the form because they simply put it in the mail. During the cognitive interviews for Form EIA-63C, our research showed that the participants’ said it took little to no time to transmit the data. Graph 3 shows the number of participants who believed it took no time to transmit their data. Only 2 of the 12 participant’s came up with a nonzero burden estimate of how long it takes them to transmit data. One participant stated it takes them 20 minutes to transmit data to EIA. This participant further explained their 20-minute estimate included the time it takes them to review the form once they filled it out. These results reflect the fact that this survey data is collected entirely electronically. This data suggests that as the federal statistical community relies increasingly on electronic data collection rather than hard copy, OMB burden component (9) may be less relevant in the future for measuring reporting burden.

Graph 3 Form EIA-63C respondents (n=12)



When conducting interviews for Form OE-417, we added a question to assess this burden element. We asked if the participants file the form online or by another method to break down the component of transmitting or otherwise disclosing the information. This question helped us understand how participants were sending in their survey but did not collect any information on how long it took for them to submit their form. A majority of the participants said

they either submit online or through an electronic reporting system or by email. One participant said they had to transmit their report by facsimile because they were not authenticated in the electronic collection system, and the other reporting methods were not working.

Combining the questions

Form EIA-63C

When the components of burden were separated from nine to seven during the cognitive research for Form EIA-63C, we found that participants were not able to provide information on discrete time units for each of the components to assess the burden. The responses showed that there is no uniformity of what participants understood each question to mean. Participants do not understand the difference between the questions and misinterpreted what the question is asking. Participants either understated total response burden or overstated the components of the total burden. Participants' interpretation of each burden question varied widely, especially for the questions, "how much time does it take you to utilize your software or technology to gather the information from your data systems" and "how much time does it take you to gather the information from all your data sources?" As a quality check, the responses to the burden component questions were summed and we found that they rarely equaled the value for the total time it took to complete the form from start to finish. People were able to provide estimates on some of how long it took them to complete different aspects. But when we asked them how long it takes to complete the form, many people gave a time estimate that was substantially less than the total summation of time they stated previously based on their estimates for the discrete aspects of filling out the form burden component.

The participants became confused when they were asked seven consecutive questions about the time it took them to complete different steps in the reporting process. One participant stated it takes 1 hour to complete and file their survey form. This participant then said that it takes 30 minutes to search the systems, another 30 minutes to gather the data, and an additional 30 minutes to record that information on the survey form. This participant combined the time it takes to search the systems with the time that they use for gathering the data, which shows a misinterpretation of what the question was asking or confusion about each burden component. Another participant said it took 15 minutes to search the systems and 15 minutes to record that information on the survey form. When this participant was asked, "How much time does it take you to gather the information" this participant stated that they included this time in their response about how long it takes to search their systems and could not break this number down further. When asking the participants to break down the information into discrete parts they did not all have the same understanding of what was included in each burden component.

Form EIA-886

When creating the protocol for Form EIA-886 we used three questions to estimate the OMB components of burden. One question that was asked, "How much time does it take you to gather the information," covered OMB components (2),(3), (4) and (7), "*developing, acquiring, installing, and utilizing technology and systems for the purpose of collecting, validating, and verifying information; developing, acquiring, installing, and utilizing technology and systems for the purpose of processing and maintaining information; developing, acquiring, installing, and utilizing technology and systems for the purpose of disclosing and providing information; and searching data sources.*" We also asked, "How much time does it take you to fill out the survey form" that covered OMB component (8). Form EIA-886 participants had a very hard time providing estimates for these components of reporting burden since there was such a diverse group of participants and many of the participants were not the individuals filling out the survey instrument. Many participants overestimate their respondent burden due to filling out the form for multiple states. Asking these questions did not provide as accurate results as Form OE-417 because the questions did not align with how the participants perceived the reporting process.

Two questions were asked to cover OMB components number (5) and (6); first, "Are you able to directly access the data needed from company records to report on this form, or do you have to work with others to gather this data": and second, "How long have you been filling out Form EIA-886." When asked these questions most participants had to work with others to gather the information. We also found no relationship between the length of time and experience with filling out the form and the responses of the burden estimate. However, these were useful questions as an approach to estimate this component of reporting burden.

Lastly, participants were asked the question, how much time does it take to complete the survey form and report this information to EIA, as a measure for OMB. This question was used as a validation step that covered all of the nine OMB components of burden. Once those component questions were asked, we asked the participant to add the time

it takes them to gather the information and report it your data, what is the total amount of time it takes to file your report for this survey? If the estimate for the total burden did not approximate the sum of the estimates for the components the interviewer asked follow up questions relating to the estimates on the burden components to validate the response for total reporting burden.

Form OE-417

When creating the protocol for the cognitive research on Form OE-417 we expanded the questions and disaggregated the components of burden in more detail than the protocol for estimating burden on Form EIA-886. However, for analysis purposes, we collapsed the components of burden into three categories: gathering, reporting, and the total amount of time it took participants to complete the survey form. We used the more in-depth questions to validate a reason if a participant reported a burden estimate that was different from the others. For this project, the participants were able to easily and accurately report the amount of time that it takes them to complete the form. The research for Form OE-417 was the most straightforward compared to the burden questions asked in the other studies discussed. In this protocol, we added a question to break down the component of transmitting or otherwise disclosing the information by asking if the participants file the form online. Another question asked was if this is the usual amount of time to report this information. Responses to both of these questions showed that the more familiarity and the more years the participants have completed the form, the burden per response decreases.

The participants on Form OE-417 were asked main two questions to measure the components of burden. These questions were simple for the participants to understand. We also asked a question on training because Form OE-417 has a unique regulatory requirement for the companies to provide training. From these questions, we were able to estimate burden easily by breaking it down into two components; First, how long it takes participants to gather the information, and second how long it takes to record the information. Participants were also asked the total time as a validation check for the two components. The estimates given by participants on how long it takes to gather and report were similar to the total amount of time they told us it takes them to fill out the form. Table 3, Table 4, and Table 5 show how the participants had similar estimates when calculating burden.

Table 3: How much time it takes to gather information for Form OE-417 (n=20)

Number of Participants	Amount of Time	Percentage
11	< 1 hour	55%
7	≥ 1-2 hours	35%
2	> 2 hours	10%

Table 4: How much time it takes to record information for Form OE-417 (n=20)

Number of Participants	Amount of Time	Percentage
19	< 1 hour	95%
1	≥1-2 hours	5%

Table 5: Total Response Burden for Form OE-417 (n=20)

Number of Participants	Amount of Time	Percentage
5	< 1 hour	25%
11	≥ 1-2 hours	55%
4	> 2 hours	20%

The amount of time participants stated it took them to gather information and report their information added up to the total response burden more consistently than the estimates from Forms EIA-63C and EIA-886 studies. This

shows that when asked to provide estimates for two main burden components participants understand the questions being asked and were able to provide reasonable estimates.

Conclusion

The three protocols used to assess respondent burden had different levels of detail to estimate the components of burden and produced varied results. The findings for Form EIA-63C demonstrated that participants did not understand all of the questions and were not able to provide separate time estimates for each component of burden. The protocol for the cognitive research for Form EIA-886 asked the same questions as the OE-417. However, the EIA-886 participants were spread across a diverse group of categories and the participants were not always knowledgeable on burden estimates because many did not fill out the form. The burden questions asked on Forms OE-417 and EIA-886 were simple and showed that they are useful for generating accurate estimates on the burden components.

Participants are not able to provide information on discrete time units for each of the components for estimating burden. Participants do not understand the distinction between the questions for estimating OMB components number (2), (3), (4), (5) and (9) we grouped these components into the category of the questions that cannot be easily understood. OMB components number (1), (6), (7), and (8) are easier for participants to understand and estimate. Participants either understate total response burden or overstate the components of the total burden when asked too many detailed questions.

Asking a question on how much time it takes to review instructions is a simple procedural step that participants can easily answer. The results from these three studies indicate that most participants do not read the instructions on a regular basis and did not consider it a component of reporting burden when answering the other questions.

Developing, acquiring, installing, and utilizing technology and systems for the purpose of collecting, validating, and verifying information; developing, acquiring, installing, and utilizing technology and systems for the purpose of processing and maintaining information; developing, acquiring, installing, and utilizing technology and systems for the purpose of disclosing and providing information are all confusing for participants because they tend to look at these components as one step or feel that they overlap and are redundant. Utilizing technology is hard for participants to estimate because most companies have had systems in place for years. The participants don't know the time it takes to write software programs and are not involved with this work. They do not include these components in their burden estimates when asked to assess how long it takes them to fill out the survey form.

Adjusting the existing ways to comply with any previously applicable instructions and requirements was not an applicable question for survey respondents. It is hard to quantify the process of adjusting the existing ways to comply with any previously applicable instructions and requirements because most respondents have a system to fill out our form either by a spreadsheet or other type of file. The setup time when the data is automatically pulled does not take long compared to the companies that do it all by hand. Asking this question to estimate this burden component does not provide useful information on the time it takes them to complete each form.

Training personnel to be able to respond to a collection of information is not common and is hard to measure. For some surveys, it is a component but it may not be a component on every survey. Searching data sources is an important component of burden when combined with gathering data to get an accurate representation of how respondents fill out the form.

Completing and reviewing the survey report is also an important step in estimating respondent burden. Respondents are able to provide estimates of how long it takes them to fill out the form. Transmitting, or otherwise disclosing the information may become less of a factor as the federal statistical community expands its use of electronic collection. A majority of EIA survey forms are submitted electronically so when participants were asked how long it takes to transmit, or otherwise disclose the information they said instantly or a minute or two for it to actually upload. These components were written when most survey forms were mailed in and it would take a few days to transmit the data to an agency.

The best approach to estimate reporting burden accurately is to collapse the 9 components discussed in the OMB guidance into three questions: How long it takes to search and gather their information; how long it takes participants to fill out the form, and what is the total amount of time to fill out the form. This third question is used

to validate the responses to the first two questions. Some subcomponents may play a role if there are unique circumstances or regulatory rules that apply to a specific survey. These other components may be used to validate the estimates of the main two components when the respondent provides a burden estimate that is an outlier value. The approach outlined in this paper is a general model to use for estimating reporting burden on a statistical survey.

1. 5 CFR 1320.8(b)(3)(iii).