Implementing a Mixed-Mode Design for Collecting Administrative Records: Striking a Balance between Quality and Burden

Marilyn Worthy[1], Danielle Mayclin[1]

[1] U.S. Energy Information Administration

Administrative records are increasingly being used by Federal agencies in lieu of, to augment, or as a quality control tool for survey data. As agencies' use and reliance on administrative records grow, it becomes increasingly important to understand how the data are assembled, organized, managed, and reported. This paper will focus on answering the question: "How does the mode in which administrative records are reported or obtained from record-keeping entities affect the quality of those data?"

We analyze the data from the Energy Supplier Survey (ESS) of the Residential Energy Consumption Survey (RECS). During the ESS, administrative records are reported by companies who supply energy to the households surveyed for the RECS. Respondents self-select mode from a fixed set of mode options, the most common being online teleform, electronic file submission, and paper/pen. Reported data were edited by analysts using tools such as bills collected during the household interview. In this paper we explore whether the reporting mode had an effect on the number of edit failures and errors in the data, by controlling for factors such as number of administrative records reported. We find that errors in the data did differ by the mode selected by the respondent.

Introduction

This paper discusses data quality disparities across modes in a mixed-mode collection of administrative records. Data quality is evaluated by studying nonresponse and data changes made during post-collection editing. Additionally, we provide an analysis of respondent mode preferences, and conclude with recommendations for improving data quality in future collections of administrative records.

Federal agencies are increasingly using administrative records to enhance, replace, or evaluate survey data. While many agencies maintain records on individuals, households, and businesses, private entities also keep records. Accessing and utilizing nongovernment administrative data widens the breadth of information available to federal agencies to potentially improve survey estimates and ultimately federal programs. The survey described in this paper collects household energy records from U.S. energy companies. Records are collected via a mixed-mode design where respondents are allowed to self-select mode.

A key decision in the survey design process is the selection of data collection mode(s), which requires weighing a variety of factors. While a mixed-mode design affords greater flexibility for respondents and potential cost savings, it can also lead to an array of problems stemming from questionnaire or survey

design effects, such as data quality differences across modes (Dillman 2000). Mixed-mode surveys may also necessitate additional processing effort to combine data submitted in different formats. However, data collectors may decide that benefits such as increased response rate or reduced respondent burden ultimately outweigh such concerns.

Data

The Residential Energy Consumption Survey (RECS) is a survey of housing units conducted by the U.S. Energy Information Administration to measure energy-related characteristics, consumption, and expenditures in U.S. homes. The RECS in-person household interviews are followed by the Energy Supplier Survey (ESS), a mandatory survey of the energy suppliers for each household. The ESS collects data directly from suppliers because households cannot recall or supply 20 months of energy records. For example, in the most recent RECS (2009) only 53 percent of respondents produced an energy bill during the household interview.

Historically, the ESS used a mail self-administered questionnaire. The 2009 RECS had tripled the sample size of previous rounds, which triggered a re-evaluation of survey mode. As a result of outreach to potential respondents, multiple mode options were created with expectations of reduced respondent burden, improved data quality and reduced costs.

Methods

The ESS contact protocol started with a recruitment telephone call to introduce the survey and locate the appropriate respondent, followed by a mailing to provide access information to the secure data collection website. Upon login, a welcome screen presented an introductory letter and a link to the survey instructions. The instructions explained site navigation, required data items, the deadline for completion, and presented the available mode options.

Electricity and natural gas companies were offered three mode options: paper forms, online forms, or Excel template. Propane, fuel oil, and kerosene companies were only offered a choice of paper forms or online forms because they had fewer cases to report. Data was requested for a 20-month period: September 2008 to April 2010. There were three critical items for each period: the end/delivery date (mm/dd/yy), amount used, and total dollars spent.

> *Paper form:* With few updates, the 2009 paper form closely resembled the forms used in earlier cycles. Through the website, respondents could view an online list of cases (including addresses and account numbers) and print forms pre-filled with this information. Completed forms could be mailed or faxed.> *Online form:* EIA designed the online form to closely resemble the paper form. Additionally, the online form had limited built-in edits, such as edits for date formats.> *Excel template:* The Excel template option allowed companies to download a spreadsheet,

Table 1   Number and percent of ESS companies and cases by response mode.

Mode	ESS companies		ESS cases	
	n	%	n	%
Paper form	257	21	965	5
Online form	771	63	6701	38
Excel template	149	12	8762	49
Other electronic file	5	0	599	3
Nonstandard printout	37	3	721	4
Other	8	1	22	0
Total	1227	100	17,770	100

> populate it, and then submit it online. The Excel template included three tabs: variable descriptions, a list of cases, and a reporting template pre-filled with case information.

After submission, edit programs flagged cases for manual review when there was missing data, anomalous data (such as outliers or inconsistent patterns), or respondent comments. To determine whether data changes were necessary, editors used scanned energy bills, respondent comments, and data collected during the household interview.

The unit response rate was 90 percent for the 2009 ESS[1], a significant increase from 80 percent in the previous round. Table 1 shows the number and percent of ESS companies and cases by response mode. In addition to the three standard mode options, a significant share of cases (7 percent) was submitted via other electronic files or nonstandard printouts.

While online forms were the most common response mode for companies, they were only used for submitting about one-third of cases. In contrast, one-eighth of companies submitted half of all cases using the Excel template, suggesting a relationship between company "size" and chosen mode.

Results

Companies with larger data requests tended to use modes where they could submit many cases at once, or "batch-report" (Figure 1). Companies with smaller requests were more likely to report cases using individual forms, either paper or online. The largest companies, those with greater than 100 cases, most often chose the Excel template mode (65 percent) compared to only 8

1 The response rate was calculated using American Association for Public Opinion Research response rate 6.
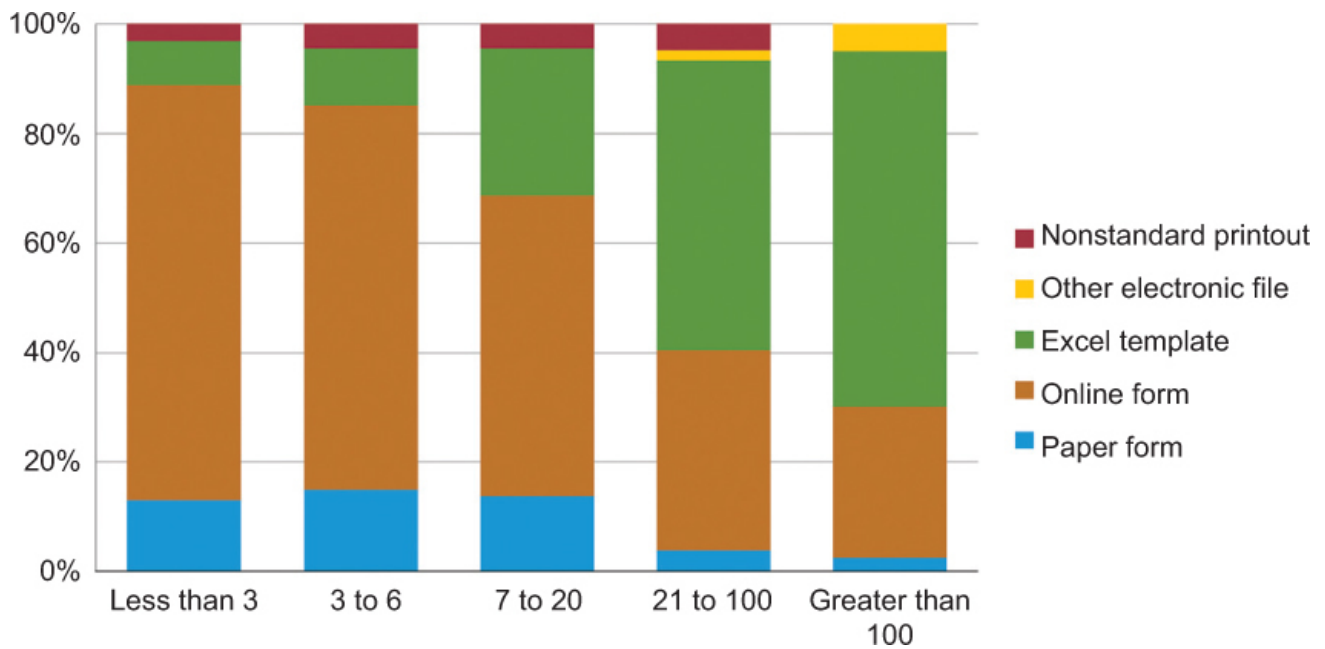
Figure 1   Response mode selected by number of cases requested.

percent of the smallest companies (less than 3 records). The smallest companies preferred the online form (76 percent), while the largest companies used this mode only 28 percent of the time.

A comparison of respondent job title and mode yielded a much weaker relationship. Two factors limit this analysis. First, the person who provided his or her job title may not have actually reported the data, playing only a data-coordinating role. Second, the reporter of the data may not have made the decision about which mode to use. For instance, it is possible that managers, who were most likely to use the online form at 65 percent, were also likely to only play a data-coordinating role, or choose the mode although they were not personally reporting the data. Analysts had the strongest mode preference, with 83 percent preferring the Excel template.

Data quality is quantified by tabulating: the level of unit and item nonresponse, and the level of data changes during editing.

Companies that chose the other electronic file and Excel templates were most likely to have some unit nonresponse, submitting less than 100 percent of the requested cases. However, these batch reporters were more likely to submit at least 75 percent of their cases, indicating they generally submitted the vast majority of their cases. The companies using these modes tended to have larger data requests and therefore: 1) may have been more likely to "miss" cases and 2) may have been less likely to search manually for records if electronic processes did not extract them from the database.

Table 2   Percent of cases with data changes by responses mode and size of data request.

Number of cases requested	Paper form	Online form	Excel template	Other elec. file	Non-standard printout
Less than 3	17	7	S	S	S
3 to 6	17	11	23	S	S
7 to 20	28	10	39	S	41
21 to 100	18	19	34	S	36
Greater than 100	15	14	30	7	15

S = Suppressed because less than 75 cases or less than five companies.

Item nonresponse was much lower among companies that used the standard modes (paper form, online form, and Excel template) than those using the nonstandard modes (other electronic files and nonstandard printouts). Those submitting in nonstandard modes were less likely to report the bookends of the requested period outside of the reference year, i.e., months in 2008 and 2010. The standard modes were formatted for reporting 20 months of data, and those submitting via these modes conformed to the requested modes and their formats. In contrast, those submitting via nonstandard modes conformed to neither. To explain missing data for months in late 2008, anecdotal evidence suggests companies archive records after a certain period of time, limiting access to older records. This underscores the importance of understanding record systems when designing data collection methods.

Data changes during editing were most common for Excel template submissions, with corrections made to 31 percent of these cases. This held true across all size categories (Table 2). Cases submitted via online form were the least error-prone, with only 14 percent requiring data changes. This was likely partially attributable to the built-in instrument edits. They were the most clean in all size categories except the companies reporting 21 to 100 cases. Overall, paper forms were fairly clean as well (19 percent). Other electronic files submitted by the largest companies were fairly clean but required extra processing effort in mapping constructs and formatting. Nonstandard printouts also required many data changes, in addition to manual keying of the data.

Across modes, the smallest companies had the cleanest cases. Respondents with fewer records to report may be less likely to make mistakes and more likely to check their work.

Conclusion

The quality of data submitted varied by mode. The traditional paper form and online form submissions were of similar quality, with high response and few data corrections needed, while the Excel template submissions did not perform as well. Compared to the standard mode options, data collected in the nonstandard modes had higher nonresponse and required more changes. Accepting nonstandard response formats shifts the burden of quality control

from the respondent to the data collector, but concerns about nonresponse bias made including non-standard data submissions more attractive than imputation.

Due to the high quality of the paper and online forms, they will continue to be offered in future rounds. The Excel template will also continue to be offered as it was preferred by respondents with the largest data requests, but efforts will be made to improve the data quality of the submissions. Two potential improvements are: 1) targeting data quality issues that may affect all cases submitted by a company and 2) immediately integrating Excel submissions with the case management system to allow quick notification about any missing cases.

The use of administrative records requires significant preparation, including understanding record structure, planning the collection process, and determining the appropriate mode(s). Data collectors may benefit from conducting data quality assessments throughout the process, especially when implementing a mixed-mode design. Planners of administrative record data collections should endeavor to strike a balance between quality, cost, and burden when designing their data collection approach.

REFERENCES

Dillman, D.A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons.

"Residential Energy Consumption Survey." 2009. Energy Information Administration. http://www.eia.gov/consumption/residential/data/2009/index.cfm?view=methodology.