Assessing Survey Quality through Streamlined Data Processing

Alicia Haelen[1], Donsig Jang[2], Flora Lan[3], Amy Beyler[4]

[1]Mathematica Policy Research, 1100 1st Street, NE, 12th Floor, Washington, DC 20002
[2] Mathematica Policy Research, 1100 1st Street, NE, 12th Floor, Washington, DC 20002
[3]National Science Foundation, National Center for Science and Engineering Statistics, 4201 Wilson Boulevard, Rm. 965, Arlington, VA 22230
[4]UnitedHeathcare, 9900 Bren Road E, Minnetonka, MN 55343

Abstract

Federal statistical agencies are continuo usly striving to provide high quality survey data in a timely manner. Adaptive survey design (Groves and Heeringa 2006) is one method they are using to help achieve this goal. This type of design draws on several data sources, such as paradata, frame data, and pro cessing data, in real time to help staff allocate resources effectively during data collection and make informed decisions about the closeout.

The technological advancem ents that make adaptive survey design possible also make it possible to streamline data processing. Survey management systems can now link data sources in real time, allowing statisticians to cond uct editing, imputation, and weighting during data collection. Researcher s can even m onitor key survey variables during data collection. (These measures, along with R-indicators and response rates, can serve as indicators of surve y bias.) Co mbining adaptive survey design with this streamlined process not only allows us to assess data quality and bias during data collection, but also expe dites data pr ocessing, because it enables us to p ut all data processing systems in place by the end of the collection period.

The development of this process was motivated by the National Science Foundation (NSF). In conducting the National Survey of Recent College Graduates for NSF, we replaced the customary sequential approach to data processing with this integrated approach. This allowed us to test our data processing procedures, including key SAS programs for autocoding, computer edits, a nd imputation. We produced and e xamined real-time quality m easures, bias indicators, and paradata and then assem bled a comprehensive quality profile and assessed nonresp onse bias. Monitoring the data in this manner also enabled us to correct proble ms as they arose. This paper presents our data processing framework, the measures we monitored during data collection, and the benefits and challenges of adopting this process.

Key words: National Survey of Recent College G raduates, adaptive survey design, paradata, streamlined data processing, survey quality measure, response rate, R-indicator

1. Introduction

Federal statistical agencies are continuo usly striving to provide high quality survey data in a ti mely manner. To reduce total su rvey errors, the agency should focus on the data quality from all aspe cts and in every stag e of data collection, including the design stage/pre-data collection (such as sample design, questionnaire design, interviewer training, etc), while fielding the survey (locating, call-backs and non-response follow-up, etc), and during the post-data collection stage (such as coding, editing, nonresponse adjustments, etc). These objectives were the principal motives behind the data processing procedure implemented for the 2010 National Survey of Recent College Graduates (NSRCG). The approach we took to accomplish our objectives mirrored an adaptive survey design, drawing on paradata, survey data, and processing data in real time to monitor and evaluate quality of both our procedures and data. With thorough planning, efficient allocation of staff resources, and use of technological advancements, the process proved effective and contributive to the practice of survey research in several ways. For instance, during data collection, we were able to track response propensities and make informed decisions about collection efforts and closeout. At the same time, we were able to evaluate data quality, estimation bias, and potential causes of nonsampling errors attributable to data processing.

Although data processing errors associated with coding, editing, imputation, and weighting are rarely quantified and potentially larger than other sources of error, this adaptive approach helped to lessen the magnitude of their impact. Our evaluation was facilitated by continuous monitoring and assessment (daily, weekly, and at pivotal time points during collection) of several survey quality indicators, performed at various stages during data processing (before editing, after editing, and after imputation). It further helped us identify, troubleshoot, and correct issues in a timely manner and without disrupting data collection efforts. As a result, combining adaptive survey design with this streamlined process expedites data processing because it enables us to put all data processing systems in place by the end of the collection period.

Survey administration and management currently are experiencing declines in response rates, increased data collection costs, and tight time constraints to disseminate high quality data; it is increasingly important to pay keen attention to both resource allocation and informed decision making for data collection closeout. This paper entails the many components of the data processing procedure implemented for the 2010 NSRCG in efforts to address these rising concerns. We present our data processing framework, the measures we monitored during data collection, and the benefits and challenges of adopting this process.

2. Overview of the National Survey of Recent College Graduates (NSRCG)

The National Survey of Recent College Graduates (NSRCG) is sponsored by the National Science Foundation (NSF), National Center for Science and Engineering Statistics (NCSES) as part of its mission to promote the progress of science; advance national health, prosperity, and welfare; and secure the national defense. The purpose of the NSRCG is to provide high quality data on the demographic, educational, and employment characteristics of recent recipients of bachelor's and master's degrees in science, engineering, and health (SEH) fields. For policymakers, the data provide an indication of the relationship between education and career opportunities. For employers in all sectors (education, industry, and government), the data predict employment trends and salaries for recent graduates in SEH fields. Evaluations of the effectiveness of equal opportunity efforts can also use the data.

Originally known as the New Entrants Survey when it was first conducted in 1974, the NSRCG has been conducted every two to three years since then. The 2010 survey is the 19th and final in this series[1]. The 2010 NSRCG collected education, employment, and demographic information from graduates who received a bachelor's or a master's degree in a science, engineering, or health (SEH) field from an eligible college or university in the United States or one of its territories between July 1, 2007, and June 30, 2009. Eligible graduates must have been age 75 or younger, living in the United States or a U.S. territory, and not institutionalized as of the survey reference date. The NSRCG has a two-stage sample design: in the first stage schools are selected from the population of schools granting degrees in science, engineering or health; in the second, sample members are selected from the list of all graduates obtained from the selected schools. Three modes of data collection were used for the 2010 survey: (1) a mailed questionnaire, (2) a web survey, and (3) computer-assisted telephone interviews (CATI), primarily for nonresponse follow-up, but also upon request.

3. Overview of Data Processing for the 2010 NSRCG

In conducting the NSRCG for NSF, we replaced the customary sequential approach of data processing, performed after the field data collection period is ended, to a process including an integrated, streamlined approach. This process involved two fundamental and interrelated tasks: (1) testing of data processing procedures, including key SAS programs for autocoding, computer edits, and imputation, and (2) calculation and assessment of real-time quality measures, bias indicators, and paradata. To facilitate these efforts, we conducted rigorous quality assessments of each processing step on an ongoing basis throughout data collection. We also assembled a comprehensive quality profile, and assessed nonresponse bias. After the close out of data collection, the complete survey data file underwent a final sequential processing procedure, with each step followed by a rigorous quality control procedure. The end result was the timely dissemination of the final, high quality survey data. Figure 1 illustrates the sequence of processing steps, which include data collection, merging of differently sourced data, coding, machine editing, imputation, weighting, and finally variance estimation. While not indicated in this roadmap, a series of quality indicators are measured in parallel to this process between each processing stage and throughout collection.

[1] The NSRCG was discontinued after the 2010 round.

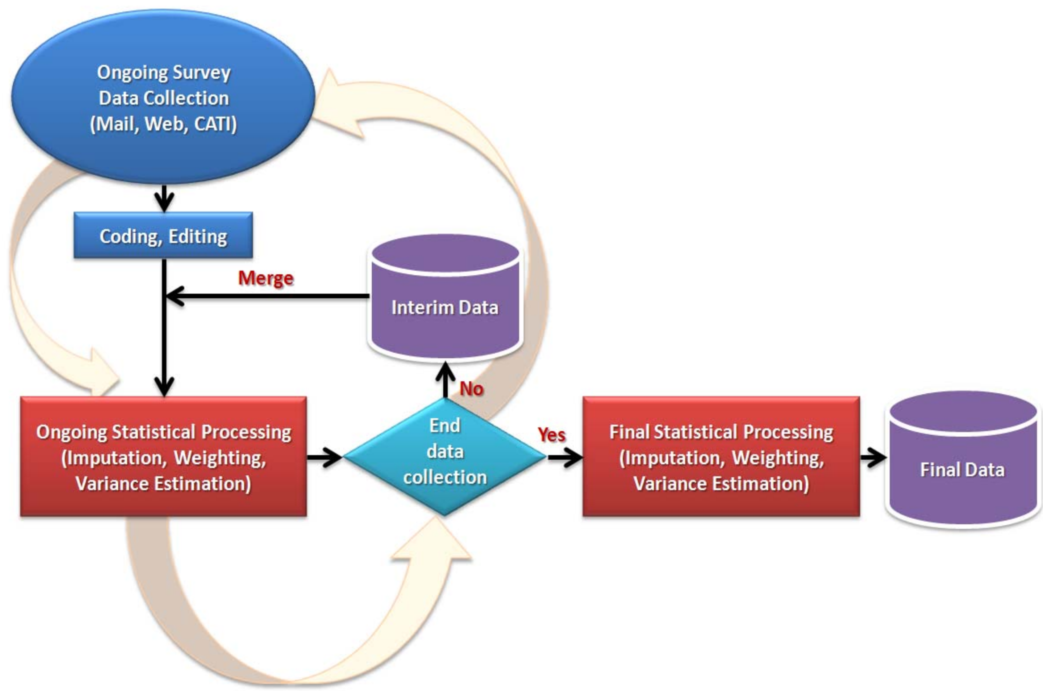

Figure 1: Integrated data collection and data processing

4. Survey Quality Measures

There is great demand for survey organizations to produce high quality data at lower costs (Biemer and Lyberg 2003). Unfortunately, how to define and assess the quality of a survey is not well established and can depend greatly on the type of survey being conducted. Juran and Gryna (1980) suggest that the "quality" of a survey be assessed through accuracy, timeliness, and accessibility. Of these three dimensions, accuracy is usually considered the most relevant—but it is also usually the most difficult to measure and evaluate. Recent research has attempted to evaluate survey quality in a variety of settings. Edgar and Gonzalez (2009) looked at it in terms of editing procedures for the Consumer Expenditure Quarterly (CEQ) Interview Survey. Laflamme et al. (2008) and Mitchell et al. (2011) considered using paradata (that is, data about the processes used to collect survey data) to evaluate survey quality. Galesic and Boznjak (2009) investigated how varying questionnaire lengths affected cooperation rates and other quality measures in web surveys. Although promising, these types of studies are still in their infancy and have yet to become well established in the survey community.

When feasible, it can be beneficial to directly define survey accuracy quantitatively. Survey accuracy is usually defined in terms of total survey error, which is a fairly general concept that is often calculated as the difference between a population parameter (for example, a mean or total) and the estimate of that parameter based on sample survey data (Biemer and Lyberg 2003). The idea, then, is that less total survey error leads to more accuracy and, consequently, higher survey quality.

Total survey error is a function of sampling error (error due to selecting a sample instead of the entire population) and nonsampling error (error due to mistakes that occur during survey implementation) (Biemer and Lyberg 2003). Sampling errors are usually calculated in a probabilistic manner given information about the survey design (Lohr 2010). For example, using the two-stage NSRCG survey design, we are able to calculate the selection probabilities of individuals in the sample and, consequently, reasonable standard errors for parameter estimates. Nonsampling errors—such as nonresponse, measurement, processing, and editing errors—are often more difficult to identify and quantify in probabilistic terms. When thinking about assessing the quality of the NSRCG data, we considered ways to identify and reduce nonsampling errors.

For the 2010 NSRCG, we produced and inspected real-time quality measures, bias indicators, and paradata consistently throughout data collection. Nonresponse and estimation biases were monitored to assess changes in data quality over time. In addition, domain-specific quality measures were produced for subgroups (defined by demographics, graduation year, major, survey mode, type of edit, and so on) in an effort to better identify the main factors that contribute to poor (and good) data quality.

We included the following interrelated survey quality measures in our quality assessment: (1) unit response and representativeness, (2) editing rates, (3) imputation rates, and (4) key survey estimates. With each measure we attempted to answer questions such as:

1. **Response rates**. Are there times in the data collection process when the unit response rate increases significantly? Are there specific survey items that are typically unanswered? Which subgroups have higher response propensities and which are underrepresented?

2. **Data editing rates.** Are there specific survey items that require a great deal of editing or are being incorrectly edited? Could these editing rates be the outcome of other error (for example, instrument design)?

3. **Imputation rates**. Are there specific survey items that require a lot of imputation?

4. **Survey estimates**. Which data processing steps, if any, incurred the greatest amount of estimation bias?

Next, we define each measure, rationalize its utility in our assessment, and describe how it was operationalized. Then we will review the streamlined data processing procedure implemented for the 2010 NSRCG and how these measures were integrated into and facilitated our quality control and quality assurance procedures.

4.1 Unit Response Rates and Response Representativeness

Response rates have long been used to assess survey quality; the general notion is that higher response rates make for higher survey quality. Although this idea has faced scrutiny (Groves 2006; Schouten et al. 2009; and many others), response rates are still considered a standard means by which to assess survey quality; while extremely high response rates (close to 100 percent) will lead to little or no nonresponse bias, lower response rates (for example, less than 80 percent[2]) will signal researchers to be wary of nonresponse bias.

Moreover, response rates provide an indication of the quality of survey data, because the potential bias in a survey estimate is a function of (1) the response rate and (2) the magnitude of the true difference between the survey respondents and nonrespondents on each survey item. To that end, as part of our data quality assessment, response rates were calculated as unweighted and weighted rates. The unweighted rates are useful for considering the success of data collection procedures, whereas the weighted rates are useful in assessing the impact of nonresponse on data quality and reliability.

Response rates were also produced for the following sampling frame domains: cohort, degree type, degree major, gender, race/ethnicity, and sampled institution. Being able to observe and forecast response propensities meant we were able to tailor collection efforts to underreporting subgroups during collection (for instance, by customizing follow-up contact materials and incentives for groups of nonrespondents). Unit response rates were communicated with the project staff daily and, at key milestones, with NSF.

Although unit response rates are advantageous and informative in certain aspects of survey data collection, they are not necessarily sufficient indicators of response representativeness (Groves and Peytcheva 2008). Acknowledging this shortcoming, R-indicators were also applied to our quality profile as a more robust measure of response coverage. An R-indicator is fundamentally a "measure of the similarity between the response to a survey and the sample or the population under investigation" (Schouten et al. 2009). R-indicators directly assess the representativeness of response without the use of response rates. Instead, they utilize individual response propensities, modeled on a set of auxiliary or frame data, which is available for respondents and nonrespondents. "[T]he *response propensity* is defined as the conditional expectation of R_i [response] given the value of x_i of the vector X of auxiliary variables: $\rho_x(x_i) = E(R_i = 1|X = x_i) = P(R_i = 1|X = x_i)$" (Schouten et al. 2011 [italics original]). These response propensities are typically estimated with logistic regressions in which a set of auxiliary /frame variables predicts response. Calculations show that the standard deviation of response propensities $[S(\rho x)]$ is bounded by $\left[0, \frac{1}{2}\right]$. The overall R-indicator, $\hat{R}(\rho)$, is estimated by:

$$\hat{R}(\rho) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\frac{S_i}{\pi_i}(\hat{\rho}_i - \hat{\bar{\rho}})^2},$$

where S_i is the 0-1 sample indicator, π_i is the sample inclusion probability, $\hat{\rho}_i$ is the estimated individual response propensity, and $\hat{\bar{\rho}}$ is the weighted sample average of response propensities (Schouten et al. 2009).

For the NSRCG, we generated R-indicators for each week of data collection using the full set of frame variables and their interaction effects, and compared them to overall

[2] This threshold is taken from the Office of Management and Budget's Standards And Guidelines For Statistical Surveys" (http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf).

response rates for the weekly returns. These estimates were instrumental in determining whether the R-indicators provide superior information on sample representation.

With these measures, our primary goal was to evaluate the value of R-indicators to assess response representativeness (in comparison with response rates). It showed whether R-indicators should be used in combination with (or in lieu of) response rates when monitoring data collection. We hypothesized that the R-indicators would exhibit different patterns than the response rates across the data collection period, indicating that response rates are not necessarily good measurements of representativeness.

While response rates provided a rough indicator of response representativeness, the R-indicators provided slightly different slopes than response rates during the data collection, indicating response rate increments may not necessarily proportionally increase representativeness of the respondent set. Moreover, examining only the final response status does not fully demonstrate the advantages of R-indicators over the response rate; rather observing changes in all of these indicators through the data collection time line can.

4.2 Editing Rates

Data processing—the process of converting survey data from their raw state to a cleaned and corrected state—is an important part of the survey process that can vastly improve survey accuracy (Biemer and Lyberg 2003). A key step in data processing is data editing, where implausible responses are identified and corrected. If done improperly, data editing can cause errors and lead to poor survey quality (Biemer and Lyberg 2003), so it is important to make sure that editing is performed correctly and systematically. When done properly, data editing can provide insight into survey quality and help identify problems with the questionnaire design and data processing. For example, if a survey item requires a significant amount of editing, this may be a sign of one or more of the following problems:

- There may be an issue with questionnaire wording for that survey item
- There may be an issue with the editing procedure used for that survey item
- Estimates based on that survey item could be subject to significant error

As part of our data quality assessment, we computed overall (unweighted) editing rates for all survey items, where editing rate is defined as the number of cases that require editing for a particular item, divided by the total number of eligible cases for that item. As survey data were collected, they went through a series of edits (for example, for range, consistency, and logical skip), for each of which an editing rate was produced. In addition to edit-type rates, editing rates were calculated by overall occurrence (that is, frequency of treatment across all editing types and for all respondents), as well as by survey response mode, degree type, gender, and race/ethnicity. Further, several observations required manual review; any data items changed as a result of manual review and editing were also flagged but not necessarily included in these rates.

Throughout data collection for the NSRCG, all data available underwent machine editing once a week; at that time editing rates were calculated and evaluated. While we anticipated certain items to require more treatment than others, it was still imperative to our process to iteratively assess editing rates for all items and identify potential sources of error affecting data quality. This iterative procedure, which will be discussed in detail later, facilitated quality assurance efforts, the identification of errors and their sources, and the viability of the machine editing system itself.

4.3 Imputation Rates

Imputation procedures are used to fill in missing values in the data set due to item nonresponse, in an effort to reduce nonresponse bias while creating a complete rectangular data set (Lohr 2010). When there is greater item nonresponse for a particular item, there is also more imputation; this will likely diminish survey quality, because imputed values, although often reasonable, are not the true observed values. Consequently, in identifying items that require the most imputation (that is, items with high imputation rates), we are also identifying potential problem areas in terms of survey quality.

Similar to editing rates, we computed overall (unweighted) imputation rates for all survey items, where imputation rate is defined as the number of cases that require imputation (that is, the number of cases that have missing values that are eligible for imputation), divided by the number of total eligible cases for that item. After the data were edited, they underwent imputation in SAS; imputation rates were then outputted and evaluated alongside the editing rates.

4.4 Survey Estimates

In an effort to assess the overall impact the data processing had on the survey data, both weighted and unweighted survey estimates were calculated and presented graphically for key survey items at four time points: (1) before editing, (2) after editing, (3) after imputation, and (4) after weighting. Key survey items included unemployment rate, salary, and the proportion of cases with a temporary residency visa. Estimation was conducted on a weekly basis, subsequent to the completion of each editing, imputation, and weighting iteration. In addition to producing estimates over all respondents, estimates were produced by gender and race/ethnicity to provide some insight on potential nonresponse bias and the quality of survey item responses. We hypothesized bias in these estimates would incur (and vary across subgroups) at each point of estimation in part because our deterministic (machine edited) and theoretical (imputed) treatment of illogical and missing data.

Assessing the quality of survey estimates is not a straightforward task, however, primarily because of the potential bias introduced due to unit and item nonresponse. A consequence of nonresponse, and the fact that true values of estimates are not known, is that bias cannot be precisely quantified. Nonetheless, trends of key survey items observed over the course of data collection (for example, fluctuation followed by stabilization, or steady increase or decrease) could give us some indication of potential bias due to nonresponse. Moreover, as we collect more data, estimate trajectories helped to identify nonresponse bias, especially for those variables closely related to frame variables for which we have base estimates for the full sample that can function as benchmark values. During data collection, we were able to posit the question of whether, based on observable trends in key survey estimates, if more data were collected, would the values substantially change; based on this assessment we are in a better position to estimate the strength of the bias. Thus, we found it critical to both analyze estimates on a continuous basis throughout data collection and to operationalize auxiliary variables in this process to inform us of potential bias.

5. Data Preparation

As part of the 2010 NSRCG data preparation, the raw data collected in all three modes—hard copy, CATI, and web—underwent preliminary editing, coding, and cleaning to produce a final data file. The data were taken through the following preparation stages: pre-key editing; missing critical item data retrieval; data entry and verification; major, occupation, geographic, and Integrated Postsecondary Education Data System (IPEDS) coding; and computer data editing.

Although these procedures are not statistical in nature, statistics staff worked with the programming staff in part to gain knowledge and insight into the data preparation process in its entirety. In addition, statistics and coding/pre-key editing staff worked together on the editing specifications used in both machine and pre-key editing. This collaboration proved essential and time-efficient when addressing, understanding, and correcting system errors. For additional information on all data preparation steps implemented prior to machine editing, see Mooney et al. (forthcoming).

5.1 Machine Editing

For the 2010 NSRCG, the statistics team, which is responsible for all processing tasks, took on the machine editing for the first time. The premise behind this decision was threefold: (1) we would gain full understanding of both the raw and edited data, (2) we could better align editing rules with imputation methodology, and (3) we would be able to test the programs and assess data treatment outcomes and rates sequentially throughout data collection.

Machine editing programs were drafted in the beginning stages of data collection and implemented on a weekly basis (see Lin and Haelen (2011) for more details about the editing process and specifications). Initially, the data we received were raw, meaning they had not undergone coding and pre-key editing. Nonetheless, with these data, the editing team was able to test the SAS programs and evaluate measures and preliminary outcomes.

5.2 System Quality Control and Quality Assurance during Data Collection

Quality control and quality assurance were essential elements of the streamlined processing approach and this proved particularly true for data editing. After each run of the machine editing programs, the editing team reviewed the results of both the edited data and the editing rates. Every week, the editing team examined items with relatively high editing rates (generally 1 percent or more). In addition, as issues arose, we worked together to identify their source (such as machine editing programs, instrument design and wording, survey administration, or respondent error). To expedite these efforts, we included all necessary staff in the conversation to gain knowledge and multiple perspectives on both the issue and possible solutions. All issues, including their sources, their impact on data quality, and final resolutions were documented as they were addressed (Jang et al. 2012).

Editing rates, described above, as well as other output from the machine editing programs, greatly facilitated these ongoing quality assurance efforts. Even though there were few data from which to draw a firm conclusion in the beginning of collection, high editing rates provided bases to identify the source of the abnormality. Certain instances, however, where a group of related items all had high and similar editing rates was not alarming; this held true specifically for subseries of yes/no questions, where respondents tended to only provide "yes" responses where applicable and not mark "no" for the other items. Typically, we were prepared for such situations and editing algorithms were already in place to correct for this; nonetheless, without evaluating these rates early on and frequently, other discrepancies may have gone unnoticed until after closeout (or been missed entirely), resulting in poorer data quality. One example concerns a checkbox survey item that (if checked) indicated the respondent did not complete high school. If the respondent first provided a valid year of graduation, this item should not be checked. However, if the high school graduation year was not provided, an unchecked box could also mean a real missing value (if the respondent did in fact graduate, it should be unchecked and if the respondent did not graduate then it should be checked, but this we do not know). All unchecked boxes were mistakenly coded in the web instrument as missing, which yielded an alarmingly high item editing rate. To account for this, we implemented a pre-editing step to distinguish a true "no" response (meaning the respondent graduated) from a real missing, by setting values to "no" for cases with reported graduation years. After applying this rule, the editing rate declined to a reasonable rate.

Along with the editing rates, the machine editing programs flagged cases with violations that could not be resolved by any of the machine editing rules, and thus required manual review. As these cases were detected, they were well documented, communicated, and attended to by the appropriate staff and NSF. These cases, in particular, required a great amount of staff time and resources and likely would have delayed processing if they had not been resolved during, rather than after, the fielding period.

6. Statistical Processing and Estimation

After the data were edited, they underwent the following sequential processing steps: imputation of missing values, survey weighting, and variance estimation of the selected key survey items. These steps were also implemented weekly throughout data collection in that sequence. The primary objectives and purpose of including them as part of the streamlined process were (1) to detect additional data inconsistencies; (2) test, modify, and finalize the SAS programs prior to collection close out; (3) assess items requiring additional and significant amount of treatment; and (4) determine which processing step(s) incurred the most estimation bias.

Working toward the fourth goal helped us better understand how the data editing, imputation, and weighting procedures we had operationalized were affecting survey estimates. For example, when analyzing the weekly survey estimates for unemployment rate, the estimated rate rose substantially in a consistent manner after the data were edited. Based on the raw data estimates, we could thus conclude unemployment rate was initially lower than expected in part because it was subjected to substantial treatment during the coding and editing stages. Incidences such as this indicate the importance of not only tracking survey estimates over time, but doing so between the fundamental stages of data processing; failing to do so would leave us with misleading estimates based only on raw data. In addition, identifying trajectories of estimates by subgroup (in our case, by gender and race) provides valuable information to the survey research discipline in studying and improving response propensities, representativeness, and nonsampling estimation error.

7. Final Data Processing

For the 2010 NSRCG, data collection extended about six months. At the end of collection, the final response data went through a final and complete iteration of the statistical processing sequence (that is, coding and pre-key editing, machine editing, imputation, weighting, and variance estimation); each step was followed by a thorough quality assurance review. Compared to prior survey rounds, the processing team found data cleaning, editing, and processing of the final survey data to be less problematic, more efficient, and require fewer resources (time, effort, and processing technology). Again, this was made possible by our collaborative efforts with other survey staff in identifying and resolving data anomalies and system errors throughout collection. As a result, the processing team, as well as other survey staff and NSF, were aware of, had attended to, and had documentation on most issues by closeout, allowing us to minimize the time and resources required to prepare and disseminate the final data files.

8. Benefits and Challenges to Applying Adaptive Survey Design

The streamlined data processing procedure we implemented for the 2010 NSRCG demands diligent attention to detail and an effective, collaborative team effort. There are major advantages to employing this approach, which are often underestimated. First, we had the ability to continuously test and modify our SAS programs and perform routine quality control procedures of our systems. Identifying programmatic glitches early on allowed for more corrective options to be explored and gave us ample time to implement and then assess our decisions. This ongoing effort was strengthened by the opportunity to share and discuss ideas, perspectives, and innovations with all survey staff. To that end, all tasks and issues were viewed through a variety of lenses, and solutions often took a number of perspectives into account.

Second, with careful planning and sufficient staffing with a variety of expertise, we were able to continuously monitor survey quality measures. Allocating resources over a longer period of time (that is, before and during data collection) may seem exhaustive and inefficient; however, when done effectively, the upfront investment proves worthwhile in reaching the overarching objective: delivering high quality data in a timely manner.

As anticipated, implementing such an extensive system of processes is nontrivial and does not come without challenges. Nonetheless, we believe this adaptive and streamlined design is adoptable and feasible to other large-scale surveys under the appropriate conditions. When considering its application, five key inputs should be taken away from this report and our experiences: (1) planning, (2) quality control, (3) quality measures, (4) communication, and (5) collaboration.

First, any well-executed process requires comprehensive planning. A strong plan is key to this process as it establishes guidelines for infrastructure development, early testing of programs, and staff time and resources needed to maintain continuous effort. We advise task leaders to take a leadership role to establish a well-defined task structure extending through data collection and to ensure that staff understand and can commit to their roles (on average, we staffed two team members per task and met once a week with the entire processing team). As a group, we established weekly and long-term milestones, discussed problems, anticipated problems, and made improvements and adjustments as necessary.

Second, an adaptive processing design is not advantageous unless quality control procedures are applied, and applied often. The high quality data produced by this process play an essential role when estimating and evaluating key survey items and nonsampling error. To that end, each processing step requires careful specification and implementation, along with monitoring procedures that can be feasibly conducted on a routine basis.

Third, a comprehensive quality measure procedure is an essential ingredient to an adaptive design. Such procedures should entail determining which measures will be employed, how they will be calculated, when and how often they will be produced, and guiding principles to evaluate the results. To optimize their value, in addition to informing the overall quality profile, these measures should be operaterationalized to evaluate and inform data collection efforts on a real-time basis.

Fourth, it is not unlikely for staff to get over-involved in their specific tasks; however, survey management involves several interrelated steps where the outcomes of one process often impact the next. Thus, communication among data collection and data processing teams is vital: for us, it facilitated our success. In addition to weekly meetings, we diligently documented all processes and outcomes and communicated via email, phone, and memorandum as needed with other teams and NSF.

Finally, survey management entails a range of meticulous tasks and procedures. At times, teams must work both in parallel and together. We found collaboration to be the ultimate key to the success of this process. Staff needed to understand not only the bigger picture and all that went into the end product, but also how and why other tasks impacted their own. We found that many issues were from sources outside our realm or requiring outside information to resolve; opportunely, team members from each sphere brought their perspective and expertise to help make rational and well-informed decisions.

Our piloted process was not without its flaws, and we plan to use the lessons learned to improve programs, technologies, and procedures when adapting it to future surveys. Nonetheless, we deem adaptive design and streamlined methodology to be an innovative and exciting approach to advancing the objectives and production of survey management.

References

Biemer, P., and L. Lyberg. *Introduction to Survey Quality*. New York: Wiley, 2003.

Coopersmith, J. "Extending Quality Measure Research Using R-Indicators." Final Report. Washington, DC: Mathematica Policy Research, July 2013.

Edgar, J., and J. M. Gonzalez. "Correlates of Data Quality in the Consumer Expenditure Quarterly Interview Survey." *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 2009.

Galesic, M., and M. Bosnjak. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly*, vol. 73, no. 2, pp. 349–360. 2009.

Groves, R. M. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly,* vol. 70, pp. 646–675. 2006.

Groves, R. M., and S. Heeringa. "Responsive design for household surveys: tools for actively controlling survey errors and costs." *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 169(Part 3), pp. 439-457, 2006.

Groves, R.M., and E. Peytcheva. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." Public Opinion Quarterly, vol. 72, no. 2, pp. 167–189, 2008.

Jang, D., C. DeSaw, A. Haelen, and X. Lin. "2010 NSRCG Data Processing Experiences and Observations." Memo submitted to the National Science Foundation. Washington, DC: Mathematica Policy Research, May 2012.

Juran, J. M., and F. M. Gryna Jr. *Quality Planning and Analysis*, 2nd ed. New York: McGraw-Hill, 1980.

Laflamme, F., M. Maydan, and A. Miller. "Using Paradata to Actively Manage Data Collection Survey Process." *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 2008.

Lin, Xiaojing and A. Haelen. "2010 NSRCG Data Editing Specifications." Final Report. Washington, DC: Mathematica Policy Research, March 2012.

Lohr, S. *Sampling: Design and Analysis*, 2nd ed. Boston: Brooks/Cole. 2010.

Mitchell, S., L. Carley-Baxter, O. Day, A. Peytchev, and S. Sadler-Redmond. "Developing Standardized Paradata and Dashboards for Use Across Multiple Surveys." Presentation to FedCASIC, 2011.

Schouten, B., F. Cobben, and J. Bethlehem. "Indicators for the Representativeness of Survey Response." *Survey Methodology*, vol. 35, no. 1, pp. 101–113. 2009.

Schouten, B., N. Shlomo, and C. Skinner. "Indicators for Monitoring and Improving Representativeness of Response." *Journal of Official Statistics*, vol. 2, 2011, pp. 1–24.