Designing a Paradata Application in a CAPI Environment

Barbara C. O'Hare
Matthew E. Jans

Barbara C. O'Hare U.S. Census Bureau, barbara.c.ohare@census.gov
Matt Jans U.S. Census Bureau, matthew.e.jans@census.gov

Background

The increase in analysis of survey paradata has been motivated by interest and need from two directions – academic research and survey operations. Academic researchers are analyzing paradata as it relates to survey participation, data quality, and effect on survey estimates in a climate of declining survey cooperation and increasing barriers to full representation of a survey population. Survey practitioners' are using paradata analysis to respond to these environmental influences in order to continue to deliver high quality data under severe budget pressures.

These two perspectives are beginning to come together, as research analysts and survey practitioners work to develop new techniques for data-driven decisions during data collection. This paper provides the background context for work currently being conducted at the Census Bureau to actively incorporate survey paradata analysis into its surveys, describes the work completed to date, and outlines the further steps to imbed the use of paradata into its production environment.

Mick Couper is credited with coining paradata as the term for describing the data about the survey process in 1998[1]. Since early 2000, increasing activity across survey organizations has expanded the umbrella of what is included in paradata and varied the discussion of how these data can be used to inform survey operations. Responsive design (Groves and Heeringa, 2006)[2] and continuous quality improvement (Lyberg and Biemer, 2007)[3] both speak to the responsive and adaptive aspect of modern survey data collection. This work is increasingly based on response propensity modeling using survey paradata among other predictors.

Most of the major survey organizations around the world conduct paradata analysis in one form or another, and many are developing survey management applications based on paradata. Some of the more notable organizations using paradata for management and statistical adjustments include the Survey Research Center at the University of Michigan, the European Social Survey (ESS), the Institute for Employment Research (IAB) in Germany, Statistics Canada, Statistics Netherlands, and RTI International.

Framework for Paradata in Survey Outcome

It is helpful to consider the application of paradata analysis in the theoretical framework outlined by Groves and Peytcheva (2008)[4] to describe the relationship between nonresponse rates and nonresponse bias. They assert the possibility of a "common cause model" where auxiliary variables (Z) are related to both survey participation (P) and the survey responses (estimates) of interest (Y). The auxiliary Z variables we are most interested in are those that predict the likelihood to participate in the survey (P) and have strong association with the survey variables of interest (Y).

---

[1] "Measuring Survey Quality in a CASIC Environment." Mick P. Couper, *Invited paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, August*.
[2] Groves, R.M. and S.G. Heeringa (2006) 'Responsive design for household surveys: Tools for actively controlling survey errors and costs', *Journal of the Royal Statistical Society, Series A* 169, 439-457.
[3] Lyberg, L.E., **Biemer, P.P.** (2007) *Quality Assurance and Quality Control in Surveys: International Handbook of Survey Methodology* In Hox, J., de Leeuw, E. & Dillman, D. (Eds.), *Mahwah, NJ: Lawrence Erlbaum Associates / Psychology Press*
[4] Groves, R.M. and Peytcheva, E. (2008) 'The Impact of Non-response Rates on Non-Response Bias: A Meta-Analysis", *Public Opinion Quarterly,* 72, 167-189.

Groves and Peytcheva note that attributes of the survey design, such as incentives and mode of data collection, can be related to both response propensity and survey estimates. In their meta-analysis of the relationship of survey attributes to survey estimates, they found that interviewer-administered surveys tended to produce larger differences between respondents and nonrespondents than self-administered surveys.

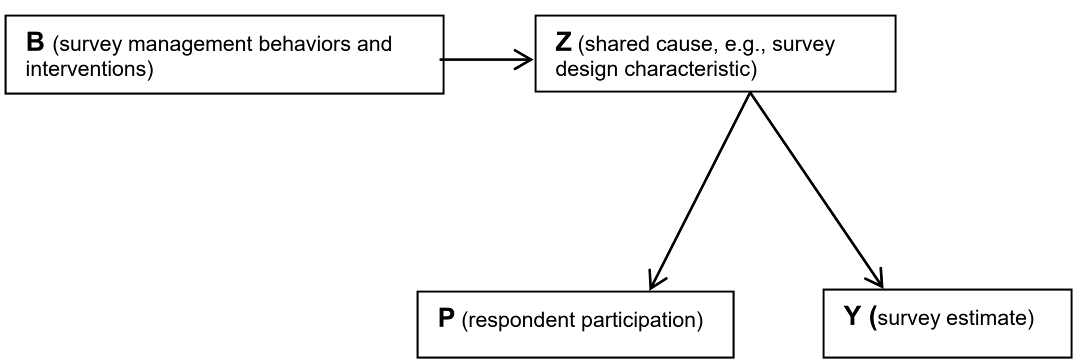
Kreuter (2011)[5] relates auxiliary data, in this same theoretical framework, to survey paradata and notes that there are two main routes for practitioners in using these auxiliary data:

- Nonresponse adjustment, i.e., weighting to adjust for probability of being measured in the respondent pool
- Monitoring and introducing design feature changes during data collection to effect survey outcomes of interest (responsive design)

To the extent that survey managers take actions to influence survey participation rates by manipulating survey attributes, we are affecting the observed relationship between Z, P, and Y. When the change in survey design is dependent on interviewer behavior, we introduce another source of variation, which is how well the interviewers execute the change in survey design.

Thus, interventions to modify the survey design based on paradata analysis introduce another domain of variables (B), the behaviors of the survey management team, influencing the interviewer administration of the survey. Interventions affect the Z outcome, and indirectly, survey participation and potential bias in the estimates, as illustrated in Figure 1. We have much to learn about the mechanisms that result in effective management interventions to change the Z auxiliary variable outcomes.

Figure 1. Framework for Understanding Paradata and the Survey Process



In this paper, we discuss the use of auxiliary variables (survey paradata) to create key performance indicators (KPIs) to monitor during data collection and suggest interventions to change field protocols. The framework discussed above is intended to provide context for the importance of systematic data-driven decision-making. As survey managers, we are trying to modify the outcome of Z that will result in the desired effect on P, and thus Y.

We need further study of the relationship between the intervention to modify the survey protocol and the effectiveness of the intervention. Successful use of paradata metrics to implement change will rely on improving our understanding of the mechanisms at work in this process. Earp and McCarthy(2011) [6] provided field staff with

[5] Kreuter, Frauke. (2011) Short course entitled 'The Use of Paradata and Process Data in Surveys' taught at U.S. Census Bureau, August 4-5, 2011.

[6] Earp, Morgan and Jaki McCarthy. (2011) 'Using Nonresponse Propensity Scores to Improve Data Collection Methods and Reduce Nonresponse Bias.' Presentation at the American Association of Public Opinion Research (AAPOR) annual conference, Phoenix Arizona.

likely cooperation of cases, based on tree analysis of known characteristics and found varying uses of this information by field staff.

Culmination of the current paradata zeitgeist leads to three fundamental questions to address:

- What paradata should we monitor?
- How do we affect an individual and organizational behavior change in our data collection process that will modify how participation (P) relates to our outcomes (Y)?
- What are the effects of these manipulations on our final variables of interest?

Census Bureau Paradata Initiative

In 2010, the Census Bureau dedicated funding for a developmental project to prototype the use of survey paradata. A small team, representing various directorates and skills, was constituted to complete the project.[7] The project included the identification and integration of operational process data across disparate systems and the drafting of charts and tables as new tools for summarizing and monitoring survey progress, cost, and quality. The purpose of the initiative was "to analyze historical survey process data and develop the survey monitoring quantitative tools that could be used in a field test of a paradata application in a demographic survey to manage survey cost and quality." (Project Charter, 2010). Important foundational steps were taken to accomplish this purpose, including:

Step 1: Identify survey process data in our current systems that can be used to monitor level of effort, cost and data quality during data collection.
Step 2: Establish how to access and compile these data.
Step 3: Define and prototype the metrics, in reports and graphics, to monitor during data collection.
Step 4: Discuss how to use the tools in field data collection and the implications for workload management.
Step 5: Provide specific prototypes of paradata reports to share with survey management and to field test.

This initiative was encouraged and endorsed by Dr. Robert Groves, Census Director. While we built on his experience and guidance from the National Survey of Family Growth (NSFG) at the University of Michigan, we recognized that the Census Bureau has particular challenges in incorporating survey paradata analysis into its operations, given the size and complexity of the organization and its survey operations.

To accomplish our overall goals, there were two constituencies involved in the discussions – the data analysts and the field operations staff. We formed a team with representatives from both areas. While initially we thought that the analysts and operations staff would concurrently collaborate to define the product, we found this difficult. We found that, in order to engage the operations staff, we needed specific examples of the new ways in which data could be combined and some specific examples of new metrics.

One of the constraints that the team quickly came up against was limitations on data accessibility across Census divisions. As a result, we decided to leverage work already completed in developing a financial reporting tool for the Current Population Survey (CPS) data. This became the survey for our project. A second constraint in trying to aggressively develop a field test of paradata was the time and funds needed to define a field test. As a result, we needed to reduce the scope of our original goals and fully address the data integration issues and the development of paradata analytic tools.

Lesson Learned: Conduct parallel efforts to define the metrics of interest and to understand the operations data available. Use knowledge of current operational procedures and data to identify gaps in information and possible new measures of survey progress and efficiency.

Data Integration and Database Development

We went into the project with a broad understanding of the types of survey process measures we might want, based on literature reviews, our practical knowledge of our survey operations, and discussions of cost and data quality

---

[7] All members of the project team were critical to the success of the project. Members, in addition to the authors, included Cathy Buffington, David Morgan, Chandra Erdman, and Tammy Adams.

concerns. However, we found that we needed to take more time to define the specific paradata measures of most interest to address our concerns. The temptation was to move quickly into merging data across multiple sources.

Lesson Learned: It may take considerable time to identify the operations data of importance, their availability, and the process to access them for integration.

The three major phases we worked through included:

1. Identify the survey performance outcome of interest (e.g., lower average cost per interview) and defining measures that will inform progress in meeting this outcome.
2. Identify the input data that will be needed to calculate the measures. What data are available in operational systems? Are they accessible? Are they accessible daily? Can the data be linked across systems?
3. Combine the data by converting operational data into analytic databases. Create appropriate data files to represent the level of analysis (e.g., a case or an interviewer).

While these steps may seem obvious, there is a bit of a tension in the process in that you want to avoid defining metrics for which you do not have data, but create new metrics by combining data in new ways. Therefore, you will want to know what data are available, although the process does not preclude identifying additional data you might want to collect in the future to improve upon your metrics.

The Director encouraged us to think of the survey process as similar to a production process. The "raw materials" are the cases and the completed, quality interviews are the product. In the process, we expend effort through our interviewers and allocation of time and money to complete the interviews, with this effort resulting in different outcomes along the process. Productivity is a measure of effort outcome relative to effort. Quality is a measure of the interviews that are produced.

This framework for thinking about the process was used to identify broad categories of data important to delivering a successful survey. These categories include:

1. Characteristics of cases (raw materials)
2. Effort by the interviewers to complete an interview
3. Output / case status during the process indicating the results of the effort
4. Productivity measures that relate output to effort
5. Quality of the completed interviews

At the same time that we were discussing the measures we wanted to develop, we were also identifying the operational databases that we might use. In our project, we were most interested in tracking and monitoring case outcomes over days in the field. We used the Contact History Instrument (CHI) database as our base on which to match data from other sources. CHI is a mechanism for the interviewer to record the attributes and outcomes of contact attempts on a case (Bates, Dahlhamer, Singer, 2008)[8]. This record of interviewer activity across days in the field became the focus of our database integration. Data from the operational case management system and from the payroll system were additional crucial sources for us to monitor survey progress, cost and quality.

Important to survey paradata analysis is the calculation of metrics that combine data *across* systems to summarize the relative interactions of all the elements of the production process – cases, interviewers, time, and quality. In order to accomplish this, we were interested in data from a number of disparate data systems, which included:

- Contact History Instrument: records of each contact attempt on a case
- Sample Control System used to manage interviewer assignments
- Employee data to capture interviewer characteristics, such as tenure and pay rate
- Payroll data that includes hours and miles per day, as reported by interviewers
- Survey response data to measure data quality, including incidence of missing items
- Survey instrument audit trails to capture interview length as a data quality indicator
- Sample frame data, including small area geographic identifiers
- Census block group data

[8] Bates, Nancy, James Dahlhamer, Eleanor Singer. (2008) '"Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Nonresponse," *Journal of Official Statistics,* 24(4): 591-612.

The process of creating a clean, matched data for analytic purposes was more time consuming than we expected.

Some of the "Lessons Learned" in the process included:

- "Available" data does not always mean accessible data. Challenges included gaining access to files from the database owners and obtaining permissible extracts.
- Paradata contain their own errors. Prepare for unusual values, duplicate and missing cases, data fields that may not be kept current, and discrepant data fields.
- There may not be common keys to link data across systems. Data may not be synchronized across sources by important units such as date or unit of geography, leading to inconsistencies.

Operations data are, by design, formatted and maintained to track survey progress *during data collection* and become quickly obsolete. Often, data such as case assignments or case dispositions are replaced with the most status. This has the consequence of losing the daily history of cases and interviewer activity, and may require specific arrangements to store the historical data for paradata analysis.

There are other logical considerations in developing the infrastructure for paradata analysis in near real-time. IT resources will need to be dedicated for data manipulation, storage and analysis. Software for data manipulation and statistical analyses need to be identified. We worked in SAS, which provided the data manipulation and analytic tools in one software package.

Particular thought and consideration should be given to the file structure of the integrated data. The data unit will differ across files and the intent should be to maintain the integrity of these units, but in a database structure that allows referencing across sources. Translating the data in operational format into data that is ready for statistical analysis requires planning. In our project, we defined three files for analysis: the case contact attempt, the interviewer-day, and the case. The latter two files summarized the contact attempts of an interviewer by day in the field and the contact history attempts by case, respectively.

Selecting the Key Performance Indicators to Monitor

At the same time that work was being completed to create the integrated databases, others were at work drafting the key performance indicators (KPIs) to be calculated from the data. Operational decisions will be made based on these data and displays, and thus defining indicators and the paradata to combine is a core operation. The intent is to monitor daily trends in key performance indicators relevant to survey progress to *anticipate* the success of the upcoming data collection. Points of change can alert the user to a potential issue to examine further. It is helpful to compare a trend line against a benchmark, such as a previous survey period. Comparisons across similar geographies or interviewers can be useful in identifying units outside of the general trend pattern. Forecasting and propensity models can be developed, the results of which can be used for planning and real-time responsive designs.

Using our production model framework, we identified KPIs representing level of effort, outcomes of the effort, and productivity. We did not pursue measures of data quality, through either survey response data or survey instrument time stamps, because of the difficulty in obtaining these data. To learn from the experience of other survey organizations, we studied the University of Michigan with the National Survey of Family Growth and RTI piloting of Adaptive Total Design dashboards.

<u>Dimensions of KPIs</u>

For any particular data element, such as completed interviews, there are multiple ways of computing a measure – counts, percentages, averages, ratios. One dimension is the numeric representation of the data display. Different end users may require different measures. Higher-level administrators may only be interested in total interviews and proxy costs, while field supervisors are more interested in the rates of completed interviews by field units.

A second dimension of KPIs is the unit level of display, or level of detail. National, regional, field teams and individual interviewer measures will be of interest to different end users. Further, access to discrete levels of detail may vary across end users, with appropriate access provided depending on the end user's role.

Effort Indicators

Survey effort indicators provide information on the allocation of resources and the activities related to getting completed interviews. Effort indicators are particularly valuable because effort can be easily increased or decreased, as appropriate, by the survey field managers. The interviewers are the "line workers" and directing their effort should have clearly measurable effects on costs and data quality.

Examples of Effort KPIs are:
- Number of hours worked
- Number of contact attempts
- Contact attempts during peak hours
- Interviewers working on any given day in field

Outcome / Case Status Indicators
For the goal of completing field operations, it is helpful to know the status of cases at any point in the field period. These can include completed interviews, non-contacts, contacts with initial refusals (if follow-up with refusals is allowed), appointments, or other situations where more information is expected to be gathered from the sample household during the survey period. The details of these field dispositions vary across organizations, but at the Census Bureau, these are known as "outcome codes" and "action codes." The Census Bureau's disposition types (e.g., Type A's (Refusals)) summarize outcome codes, and these codes may be one logical variable to monitor in a dashboard.

Examples of Outcome / Case Status KPIs are:
- Cases that are completed interviews and other disposition codes
- Cases that have not been worked
- Noninterviews, by type (e.g., no one home, refusal, etc.)
- Transmissions received (hours or cases)

Productivity Indicators
Productivity can be loosely defined as the relationship of effort to "product", in this case, a completed interview. A productivity indicator, such as hours per interview, reveals how much effort is needed to produce the interview. Some productivity measures can be considered measures of efficiency.

Examples of Productivity KPIs are:
- Hours per interview
- Contact attempts per interview
- Miles per interview
- Open days per interview

Quality Indicators

Under the concept of data quality, we include indicators of the accuracy and variability of the data collected relative to the target population for the survey. While there are other data quality elements like "fitness for use" and "timeliness," we focus here on data quality issues that affect the estimates made from the data. This quickly becomes a complex statistical topic, as data quality is ultimately apparent only when data are analyzed. Because data quality is a characteristic of individual data elements that in summary can create biased or highly variable statistics, monitoring data quality means anticipating the most common analytic uses of the data, and monitoring changes in the features of those uses (e.g., variance and bias in a statistics, or proportion of item missing data).

The response rate over days of data collection has been considered a data quality indicator because it is assumed a proxy for nonresponse bias. A higher response rate survey is assumed to have lower nonresponse bias, but this metric assumes that non-respondents are different from respondents on key survey variables. This is only an assumption. Recent research has brought that assumption into question.[9] Unit nonresponse is not the only potential source of error in data collection. Item nonresponse and measurement error are additional sources, discussed in the Total Survey Error perspective that might particularly contribute to survey quality during data collection (see Groves et al.[10] for more explicit definitions and examples of each error source).

Data quality indicators for measurement error (i.e., errors of observation) are less clear, yet we can use operational data that are expected to be related to errors in responses to questions as indicators of measurement problems. For example, long or short average question duration can suggest that either interviewers are having difficulty delivering questions or respondents are having trouble answering them. Either of these indicators could be related to measurement error in reported data. Excessively short question durations in particular can indicate skipping questions or reading them too fast. Short questions durations for all questions in a survey might indicate falsification.

Examples of Quality KPIs are:

- Item missing or Don't Know responses
- Instrument section completion times
- Response rate

<u>Selection of Measures to Pilot</u>

Over a six-month period, the analytic team selected and drafted a set of key performance indicators to represent the different components of the survey process to measure and graphically display. Of paramount importance is the operational utility of the information. We wanted the information to provide insight and suggest survey management actions, such as case reassignments, that could improve survey performance. At the same time that we were concerned about relevance of the metrics we would provide, we also wanted to improve on the metrics currently available to field supervisors. Does the operations data that are presented provide new information to survey managers that better helps them identify actions they should take? Last, we wanted to display trend data to view daily. Trend charts provide an overview across the data collection period, which can assist in seeing unexpected changes.

Lesson Learned: Link KPIs to the production process (Effort, Output, Productivity, and Quality). Identify most important KPIs depending on the role of the end user.

Paradata Charts and Analytic Tools

Paradata displays will vary across user roles and needs. Data collection managers want to know the dispositions of cases, response rates, and other summaries of active data collection on a timely basis, when interventions are still possible – as near to real time as possible. Headquarters project managers and financial staff are interested in daily field production and cost estimates. Senior executives want to have quick access to the overall pulse of their organization with respect to field production and cost. We have developed tools that can work across all levels of the organization. Dashboard environments, in which users can see a display of timely data, can facilitate these user needs by streamlining access to survey process data (i.e., paradata) and cost data. The Census Bureau is in the process of developing a corporate data dashboard system, using near-real-time data for survey management.

Whether used in dashboards or not, paradata displays can be designed to follow the workflows and chains of command that drive organizational operations, streamlining work processes that currently require many steps for mission-critical work. Yet, having detailed production information at survey managers' and executives' fingertips via an easy-to-interpret interface gives them more time to devote to more critical management responsibilities.

---

[9] "Nonresponse Rates and Nonresponse Bias in Household Surveys." Robert M. Groves, *Public Opinion Quarterly, 70,* 5, 646-675, 2006. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." Robert M. Groves & Emilia Peytcheva, *Public Opinion Quarterly, 72, 2,* 167-189.
[10] "Survey Methodology." Robert M. Groves, Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau." New York: Wiley, 2007.

We pursued two avenues of analysis: 1) graphic and tabular presentations of daily performance indicators for field operations management and 2) multivariate analysis of the paradata to explain survey performance metrics. We will briefly highlight an example of each of these analyses.

The charts of key performance indicators can include sample dispositions and response rates for the survey as a whole or by any subgrouping available in the data. Based on operational experience and current practice at the Census Bureau, plots by day in the field period seem to be the most useful for survey management (i.e., where the x-axis in day in the field). Plots by number of contact attempts may also be helpful. There are numerous ways in which paradata can be displayed. To be useful, any chart should encompass critical characteristics.

Charts should:

- Be Actionable
- Use Benchmarks
- Highlight Variability
- Highlight Extreme Values and Outliers

Chart Examples

Figure 2 is an example of an Effort KPI chart. The KPI is calculated as the total hours worked by a field representative divided by the total contact attempts for any given day, representing the average number of hours a field representative takes to make a contact attempt. The graph displays the average of the field representative KPI, across all field representatives (solid line). The dotted lines show the upper and lower bounds around one standard deviation from the mean, and indicate the degree of variability across the FRs in their average time per contact attempt. For any given day, in this case Day 9, the codes of the FRs with the highest and lowest average hours per attempt are displayed in the shaded boxes, which alerts the field manager to interviewers that may require further investigation as to the reasons for the highest and lowest values.

Thus, across all interviewers, a contact attempt averages a little over one half hour at the start of the field period. Notice that the trend goes up a little over time (from about half an hour to almost an hour at the end). Notice also that the standard deviation bounds get wider over time, partly because the pool of outstanding cases without resolution decreases.
Key characteristics of this chart include:

- Highlights variability across FRs
- Highlights units at the ends of the distribution
- Actionable information on hours (proxy for cost) and interviewers

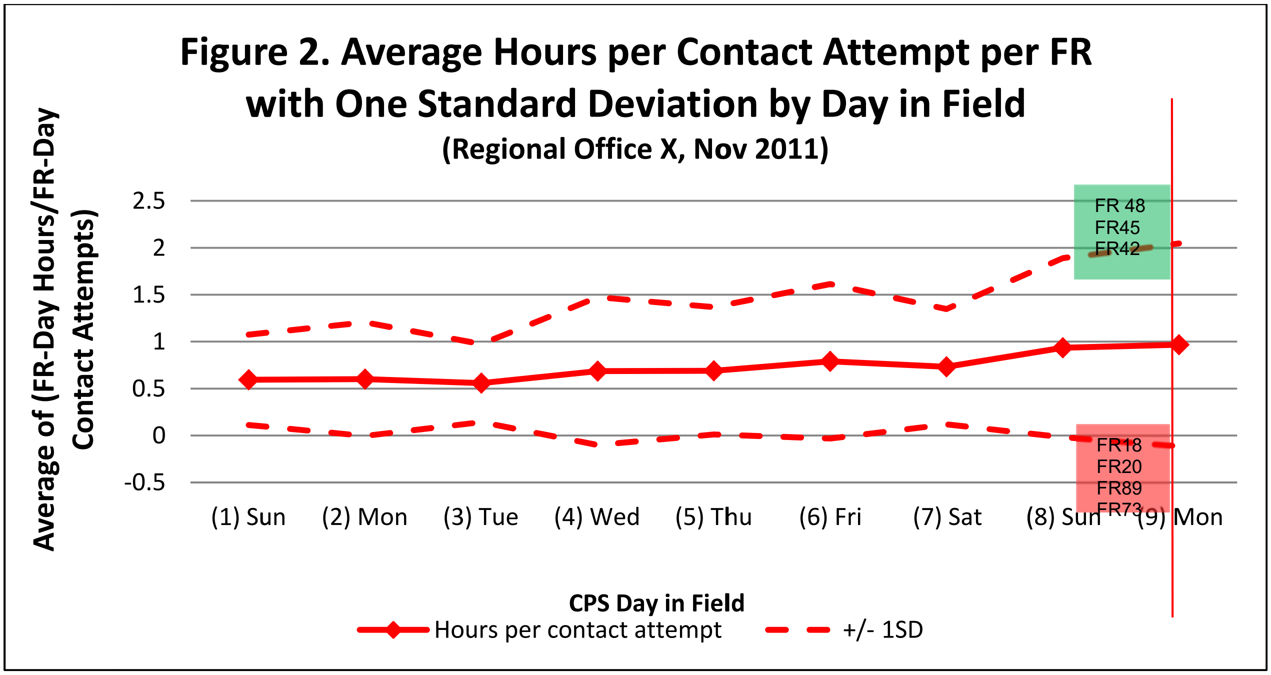
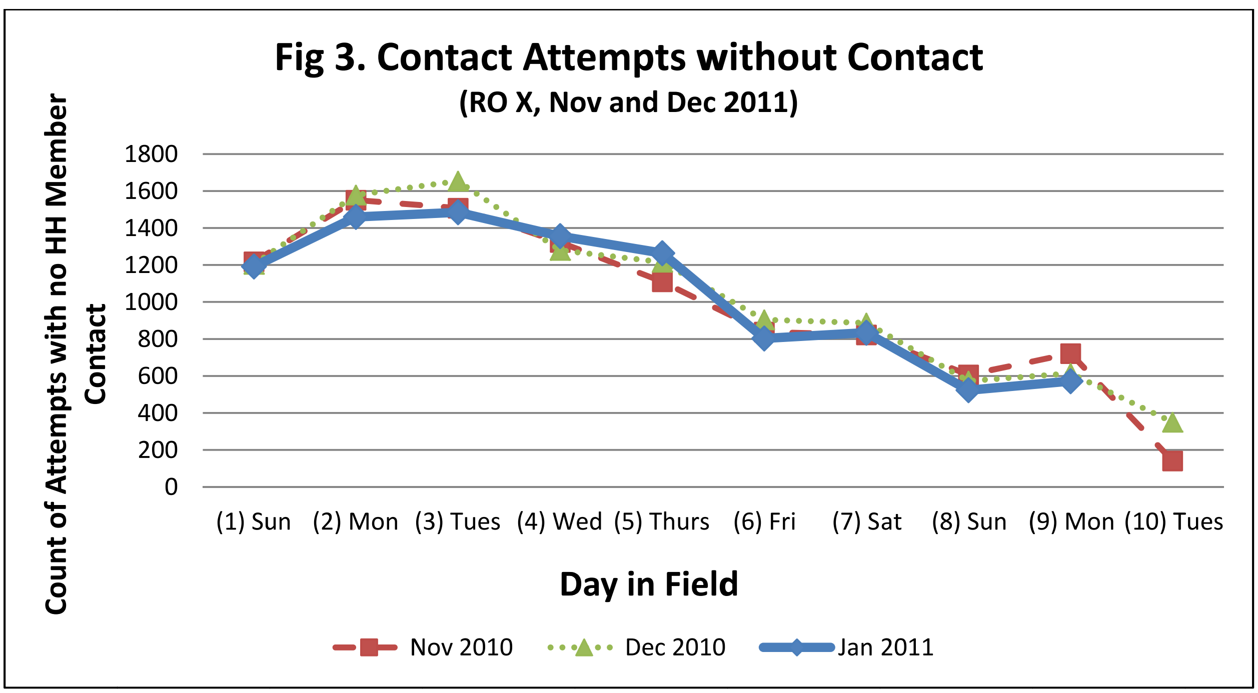


Figure 3 is an example of an Outcome KPI chart, displays the total number of contact attempts with no contact, and compares across survey months.

The number of contact attempts without contact gives the survey manager a measure of the number of "actions" that are being taken without producing interviews. A long-term goal might be to reduce that raw number, perhaps by timing calls to fall within peak hours. The KPI decreases over the field period, as cases are closed out and the number of active cases diminishes.



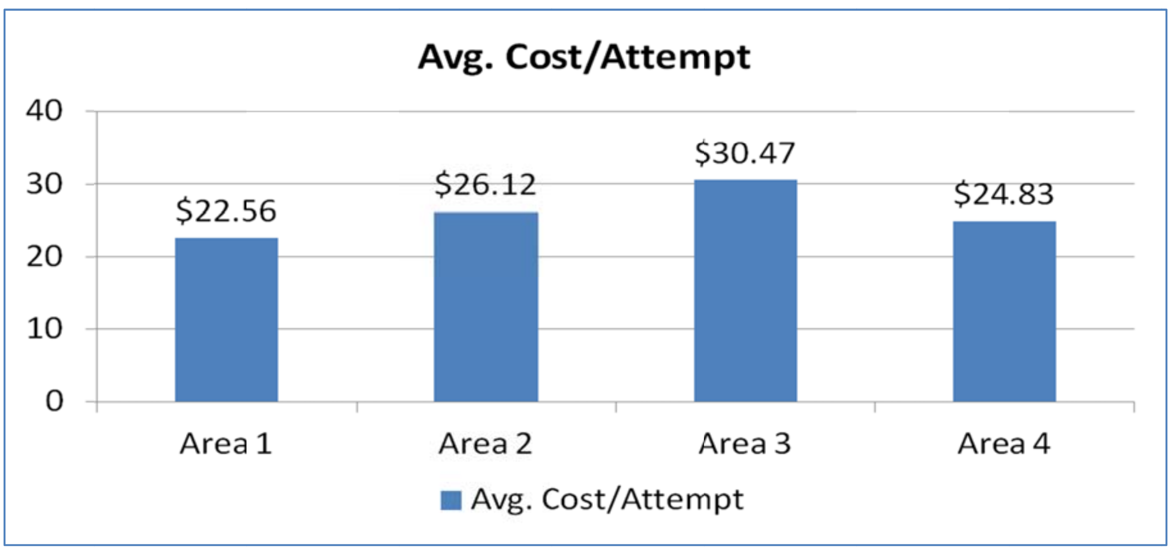
Data Modeling and Response Propensity Estimation

Predictive models based on paradata are another method for synthesizing data from multiple sources and for creating new management metrics. Modeled estimates can be useful for understanding survey outcomes, costs, and response propensities. At Census, we are estimating response propensities, to predict the likelihood of a completed interview on the next contact attempt. Model predictors include known characteristics of the housing unit and outcomes from previous contact attempts. Propensity estimates will be used to establish categories to flag cases for operations, and will be displayed to field managers. We are currently working on different ways to display this information, but the propensity information essentially becomes another known characteristic of the case for field operations planning, akin to neighborhood observations, contact history, or socio-demographic information about the neighborhood.

A Drill-down Chart and Reporting System
In addition to developing prototype charts to display key performance indicators, we wanted to provide an easy tool to access the new metrics. Two team members programmed a reporting system that allows users to view data at multiple levels of detail. This drill-down capability allows the user to see the underlying data to analyze. FieldSCAPES is the name of the census system, which is a series of tables programmed in Base SAS with ODS output, and organized in a drill-down format. SAS outputs HTML pages that are displayed in a browser. The pages are dynamic in that a user can click through several hyperlinked figures and tables, but all the pages are loaded from static HTML pages. While the Census Bureau is in the process of developing a large-scale corporate solution to paradata storage, access and analysis, FieldSCAPES is proving to be a useful tool for testing and evaluating metrics in a field environment.

Following is an example of FieldSCAPES reporting. A bar chart display compares the KPI averages across geographic areas. In this case, the KPI is average cost per contact attempt.

Figure 4. FieldSCAPES Chart of Average Cost per Attempt by Geographic Area



A user can click on any area figure, and display the next lower level of detail – sub areas, field representative teams within areas, or individual field representative data, as seen in Figure 5.

Figure 5. FieldSCAPES Report of Field Representative Cost per Attempt

FR	# of Attempts	Cost / Attempt	# of Personal Visit Attempts	PV Cost / Attempt	# of Telephone Attempts	Phone Cost / Attempt
T08	70	8.74	27	46.70	43	4.79
T22	56	24.94	15	35.0	41	17.63
W14	29	20.47	23	33.30	6	14.24
Y19	72	18.88	44	25.00	28	10.19
Y36	80	9.15	47	22.20	33	6.84

Incorporating Paradata into Field Operations

A critic might ask, "We've been using case outcomes, interviewer expense data, and case transmission data in operations for years! What's new?" While the statement is true, the novelty of paradata integration and display is 1) the display itself, looking at paradata in new ways and in new combinations, 2) the ease of creating of new indicators or old indicators more quickly due to data integration, and 3) the use of statistical forecasting to inform field practice.

Even if all staff is on board with the use of paradata and the organization supports it financially, concerns and challenges are still present. These include:

- Interviewer workloads: How will more proactive management with paradata or full responsive design affect interviewer workloads and expectations of the job? For example, the Census Bureau offers benefits to interviewers who work a minimum number of hours, which is evaluated quarterly. If changes in assignments can reduce the number of hours worked, are there changes in interviewers' expectations or job satisfaction?
- Multiple surveys: In an organization in which interviewers work on multiple surveys, how do changes in one survey affect workloads on others?
- Changing midstream: Most large-scale survey organizations operate in fast-paced, tightly planned environments. How do mid-stream changes in procedures affect plans and operations throughout the organization? For example, if Survey A monitors sample representativeness and decides they can stop data collection a week early, does that mean that the data will be delivered a week early when programming staff also have to prepare and deliver data for Surveys B-G.

Organizational challenges can remain as well. Most of these revolve around getting "buy in" from primary stakeholders (both within the organization and outside, e.g., clients and sponsors). Even if one class of stakeholders (e.g. clients) supports the use of paradata in practice, another (e.g., programmers) may not. Key players in the operations need to talk often and openly about challenges, concerns, and goals in order to make a project like this work. Yet it is possible if they are guided by the goal of defining new tools in ways that are helpful, provide consulting and training as needed, and offer guidance and setup feedback loops during implementation.

Lessons Learned in Field Implementation:

- Field staff and survey sponsors may gravitate to the usual metrics, making it challenging to introduce new ways of measuring survey performance.
- Analysts may not fully appreciate the time demands and management challenges of field operations.
- Implementation is an iterative and cooperative process.

The Future: Measuring, Adjusting, Modifying

The Census Bureau is still in the early stages of effective use of survey paradata in managing data collection operations. At this time, we continue to focus on bringing together the research on paradata metrics with the field staff need for information that is actionable and will improve survey cost and quality. Specifically, we are developing training and communications on how to use the new reports in the field, and providing a mechanism for research analysts to receive feedback. At the same time, the analysts are continuing analysis of the underlying data and proposing additional metrics for possible field use. We will evaluate and document the use of paradata in production.

Longer term we will want to measure the outcomes of the paradata interventions. What are the survey quality and cost benefits? Can we improve our understanding of the survey mechanisms at work? We expect that both "naturalistic" observation and experimentation can be conducted to measure how well the new tools are being used and what impact they have on operations and on data products. A responsive design test using paradata is already under discussion. We will continue our focus on CAPI surveys, but expect our next mode of study to be CATI surveys.

We will build on our knowledge as it becomes available and take an iterative approach to paradata analysis and production applications. Our tools and plans will continually be adapted, as we incorporate paradata into data-driven production decisions.