Managing Survey Costs: Using Mobile Technology Coupled with Performance Dashboards to Improve Field Operations

By Brad Edwards, Richard Dulaney, Abie Reifer, and Tamara Bruce

All authors are at Westat, 1600 Research Blvd., Rockville, MD 20850

Introduction

The cost of traditional household surveys continues to rise and challengers increasingly question their value in the face of nonprobability studies and big data. In face-to-face surveys, interviewer labor and expenses drive the project budget, yet little is known about how interviewers spend their time when they are not conducting interviews. Although CARI can provide some insight into interviewer behavior during interviews, it often reveals how poorly many field interviewers ask questions (Hicks et al, 2010; Suchman and Jordan, 1990), and little of their behavior when not conducting interviews (how they spend the majority of their time) has been observable. However, technology exists today that can remedy that. Smart phones can tell us where interviewers are and what they are doing. Speech to text, smart small screen design, and internet connectivity enable them to record and transmit outcomes of contact attempts on the doorstep. With text and instant messaging the supervisor and interviewer can remain in almost constant contact. Performance dashboards can give the supervisor at-a-glance views of the region's current activity, with drill down and drill through capability, and with alerts based on statistical process control that highlight problems of greatest concern each day.

In the last decade we saw the concept of a virtual telephone survey research center take shape, enabling telephone interviewers to work at home, but with the equivalent supervision and coaching provided in a brick and mortar center. The capability exists today to create a virtual field center, with levels of control that come close to those we have come to expect in CATI call centers. This paper presents a proof of concept for a virtual field center. We used task analysis and qualitative research with this dispersed work force to inform an attempt to create a user-centered design for a real time field management system, deployed it on a large field survey, and conducted a cost-benefit analysis in evaluating its deployment.

Background

Face-to-face interviewing (FTF) is the most expensive survey mode. Telephone is the next most expensive. Modes that do not require interviewer labor are generally much less expensive than those that do. Groves (1989) shows data from one survey comparing the cost of FTF with phone that suggests FTF is more than twice as expensive as telephone. For national surveys, the cost of developing the infrastructure to mount FTF surveys is so high that most survey organizations are locked out of the FTF market. Because FTF interviewing is local, interviewer wages are tied to the local labor market, and depending on the economic health of an area and the complexity of the interviewer's task, it may be difficult to find a match between the amount the survey can afford to pay and the skills of the workers that amount can attract. However, FTF still maintains some key advantages over other modes. It achieves higher response rates – an interviewer at the door is harder to ignore or refuse than a voice on the telephone, an envelope in the mail or a message in the in box. And with FTF it is easier to collect environment and biomarker data.

FTF costs seem inextricably linked with quality. Reducing costs without reducing quality is very difficult. Field interviewers are a dispersed work force. They work out of their homes and seldom have face-to-face contact with supervisors or co-workers. They are not usually connected to the Internet. Reaching them can be difficult: they may use email only sporadically, they may not be in regular phone contact, and they may not be able to receive text messages. Compliance with protocols for contacting sample assignments, working priority cases, and dealing with uncommon situations can be problematic. To the extent a survey includes non-interviewing activities in the respondent's home, such as collection of biomarker or environmental data, the interviewer's job may overtax her skills. Although paradata are increasingly used in monitoring interviewer behavior, no paradata exist for many activities. Compliance with standard protocol for asking questions is often poor. Observers of field interviewer behavior have been struck by the degree of departures from script and leading probes (Suchman and Jordan, 1990), especially when compared to telephone interviewers. Because so much of the field interviewer's work is unobserved, falsification is a considerable threat to data quality.

It may be useful to review the conditions in a typical survey research telephone center. Supervisors constantly monitor samples of interviewer performance in real time, in an unobtrusive manner. They evaluate adherence to protocol and compliance with standardized interviewing practice, asking questions exactly as worded, probing neutrally, and recording answers accurately, without the interviewer being aware of the supervisor's presence. The risk of falsification is low, because the chances of being caught are great. The supervisor is able to give frequent feedback to the interviewer, even sending text messages while an interview is in progress with hints or tips for handling tough interactions. Interviewers communicate frequently with supervisors during each shift. This tight control and ability to build a close sense of teamwork enables telephone centers to achieve a high level of standardization and quality.

In the past decade, with advances in telephony and IT infrastructure, it is possible to move outside the bounds of a brick-and-mortar center. Interviewers can work from their homes, using their own devices, but working within the organization's central telephony and call management system. This affords the same level of close monitoring, communication and teamwork that can be found in a brick-and-mortar center, and enables an organization to break away from the local labor market and access an almost unlimited pool of workers who prefer to work from home or are unable to easily leave their homes. (CITE ALLEN) Unfortunately, the market for telephone surveys is much smaller than two decades ago. It has diminished greatly because of plummeting response rates and the rapid rise of the much less expensive Web mode.

Vision of a Virtual FTF Research Center

Contemporary society has benefited greatly from the combined effect of two technological principles. Moore's Law says that information technology capability doubles every 18 months. Perhaps less well known is Metcalf's Law, which indicates the value of a communications network is proportionate to the square of the number of users. When applied to a network like the Internet, which didn't exist 25 years ago, the benefits derived from Metcalf's Law are readily apparent. The combination of the two has led to products like the iPhone, which is much more powerful than a roomful of mainframe computers was 30 years ago, and which can place its user in the middle of networks like Facebook and Twitter, processing billions of messages from more than 10 percent of the world's population virtually instantaneously.

These IT and communications capabilities offer an opportunity to bring survey operations on FTF surveys under much greater control, reducing costs and increasing quality at the same time. The following elements can be combined to create a virtual FTF research center:

- Connectivity
- Mobile devices
- Rich paradata
- Infographics
- Management information systems that connect these elements

Such a center can be managed much like the operations of a centralized telephone facility, and can move FTF interviewing to much more acceptable levels of standardization. To realize this vision, we worked on two fronts: a mobile app for field interviewers, and a performance dashboard for field supervisors.

The Mobile App

We developed a mobile app for iPhones, with two parts: MyCases and MyDay. These were designed to address two major problems that are associated with FTF efficiency and quality. First, although great strides have been made in the use of paradata to understand the survey process and evaluate quality (Couper, 1988; Groves and Hueringa, 2006; Kreuter, 2013), much of the progress has been based on interviewer call attempt records, which are of notoriously poor quality (Biemer, 2013). A field interviewer might visit five households in her assignment in a two hour period. Each visit could have a different outcome. The interviewer might make notes about the visits, but often does not enter the outcomes into a management system on a laptop until later, at the end of a block of working time in the field or when at home later that day. Inevitably notes get lost, memory fades, and visits are not recorded correctly or not at all. Second, FTF interviewing has a major time and cost component that telephone interviewing lacks: travel. (Groves, 2004) The ratio of interviewing time to total interviewer time is much lower for FTF than phone, and travel is the major reason. However, other FTF activities (planning work for the day, calls with the supervisor, etc.) also take time, and may be less efficient in the field than in a telephone center. Yet little is known about how interviewers spend their work time when they are not actually interviewing. Attempts to derive this information from interviewers breaking out their time into categories for payroll purposes have not been very successful, because it has been difficult for interviewers to keep track of their time in small increments on paper while they are on the go.

With MyCases on the iPhone, interviewers could enter the outcome of each visit on the doorstep. The design simplified the traditional record of contacts to fit the small screen, and took advantage of the IOS user interface and the inherent capabilities of the device. The date, time of day, and interviewer ID were obtained from the device, not from interviewer entry. The interviewer chose the contact mode, and used a flywheel to select the outcome. Notes could be dictated using Siri, and converted instantly to text. Certain outcomes (appointment, refused) invoked screens that collected more detail. This approach made the interviewer's task much easier. The design included a list of all active cases in the interviewer's assignment. The history of contact attempts for each case could be accessed. Tapping on the address for the case generated a map showing the most efficient route from the current location to the case. Tapping on the phone number initiated a call. The device also yielded important new paradata: GIS information on the location of activities like interviewing, and on the interviewer's travel. With this user centered design, we sought to encourage the interviewer to record contact attempts immediately, thus improving paradata quality.

With MyDay on the iPhone, interviewers could more easily break their time into increments. They were encouraged to think of the phone as their portable office, to turn the app on when they began their workday, and to indicate which of three activities they were doing (field, travel, or administrative tasks). Typically the first activity of the day would be administrative. As they switched activities (e.g., entering their car to drive to an assignment), they were encouraged to tap the appropriate button (e.g., travel). A running count of the time they spent on each activity was displayed throughout the day. They were also able to enter the miles they had driven. At the end of their block of work time (their "shift"), they would tap the stop button.

Whenever the interviewers were connected to Wifi or within range of a cell tower, data from the app was flowing into the management system and available to the supervisor. When out of range, the data were stored locally until a connection was established, then all the stored data were sent.

The Field Supervisor Dashboard

"Dashboard" became a popular business term in the 1990s, but its meaning was not well established. Stephen Few (2013) conducted hundreds of interviews with business executives to determine a common denominator. He did not find one, but from his review he formed a definition: "A dashboard is a visual display of the most important information needed to achieve one or more objectives, consolidated on a single screen so the information can be monitored at a glance." A market has emerged for dashboard software, but it is oriented primarily toward colorful visual displays, with little help in determining key performance indicators (KPIs) or identifying ways to create infographics with high signal to noise ratios.

Although some survey organizations began developing "at-a-glance" reports that pulled key information together in a tabular format in the 1980s, the earliest use of performance dashboards occurred in work on Cycle 6 of the National Survey of Family Growth at the University of Michigan. Dozens of statistical process control charts and reports were generated from paradata overnight, to identify instances of interviewer variance that could lead to biased survey data (Groves and Heuringa, 2006). They included visual displays and key performance indicators, but also many other reports on metrics that were not key and the information was not presented on a single screen. [More recently, the survey has adopted a more fully realized dashboard, with the most important KPIs presented on one screen (Kreuter et al 2010)] The Energy Information Administration developed a dashboard for their surveys about energy consumption in the U.S., but this was a strategic dashboard for senior management, not a performance dashboard (Reifschneider and Harris, n.d.). A literature search did not identify other mention of dashboards in surveys. This may reflect the thinness of the literature on survey operations in general.

We developed and implemented a performance dashboard for field supervisors on a national FTF survey in 2015 (Exhibit 1). We called it the Paradata Discovery Dashboard (PD3). The dashboard software architecture makes use of "portlets." A portlet is a pluggable user interface component that is displayed and managed on a portal or dashboard. The initial version of PD3 features 8 portlets, all visible at a glance on the screen:

- Two cost portlets, stacked bar charts of the region's average hours per completed interview by week, and the region's interviewers' average hours per complete to date;
- A GIS portlet, with an initial map showing the location of the region's interviewers;
- A "MyPeople" portlet, with photos of the region's interviewers above some information about each;
- A quality portlet, charting general interview quality from the CARI system;
- Two production portlets, stacked bar charts showing each interviewer's major case dispositions and goals for the region; and
- An anomalies portlet, a list of alerts signaling conditions that require the supervisor's immediate attention.

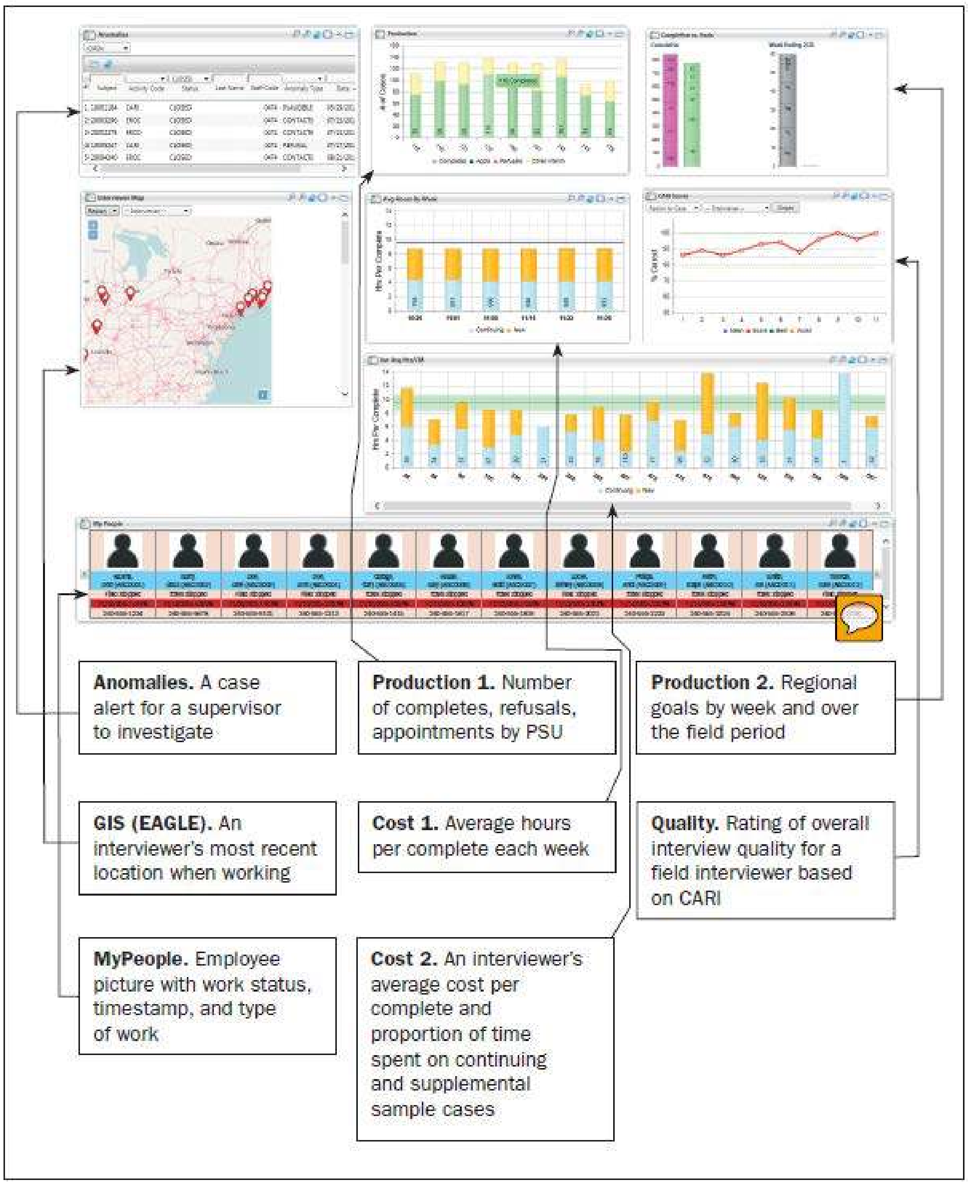

Exhibit 1. Annotated PD3 Dashboard

There is some overlap between the concepts of dashboard and portal. Several of the portlets we describe here also serve as portals into paradata databases, allowing deep dives into problems that surfaced in summary form on the dashboard. For example, the GIS portlet allows the user to zoom in and out from the initial view of the region's interviewers, to view all interviewers on a map, or to view individual interviewers and the distribution of their cases. Users can link directly to the GIS system for in depth analysis.

Cost Portlets. The region's hours per completed interview per week are displayed in a series of stacked bars, refreshed weekly. "Hours per complete" is the most commonly used metric at Westat to monitor data collection costs; it is easy to calculate and understand, and other cost elements such as interviewer mileage and wages track closely to it. In the first cost portlet, the "Region Average Hours per Completed Interview by Week," the height of each bar reflects the total hours interviewers charged to the project as of the end of the week divided by the number of completed interviews in the region. A horizontal line across the chart represents the budgeted hours per complete. Interviewers worked two sample types on the project, with different cost characteristics. The new sample was expected to exhibit much higher nonresponse rates, and therefore be more costly per interview in the field. The proportion of "old sample" completes and new sample completes in the region are shown in different colors in the bar. This gives experienced users a quick read on the relative difficulty of the completed work.

Individual Interviewers' hours per complete as of the most recent week were displayed in another stacked bar chart. Each bar in that chart represented one interviewer's total hours charged to the project as of the end of the current week, divided by the number of interviews the interviewer had completed so far. The proportion of new and old sample completes was shown with the same colors as on the region's chart by week.

The cost portlets were indicators of operations costs at the most summary levels. To diagnose cost issues and evaluate them in the context of production and quality, the supervisor could shift attention to other portlets on the dashboard.

GIS Portlet. The GIS portlet used data on latitude and longitude transmitted from the interviewers' smart phones. When the interviewers were connected to the internet, the GIS portlet showed their current location; otherwise, it showed their most recent known location. The initial view on the dashboard showed the location of all interviewers in the region. By zooming out, supervisors could see the location of other interviewers in nearby regions. This was useful in determining who might be in the closest position to travel into a region to fill staffing gaps. By zooming in, supervisors could see an individual interviewer's location and the location of their work. (Addresses of all cases in the sample and all interviewer homes were geocoded at the beginning of the survey.)

Westat developed the Efficiency Analysis through Geospatial Location Evaluation (EAGLE) system initially to identify falsification. (Hasson, 2014) The location of the mobile device at the time the interview was conducted (as determined by CAPI timestamps matched to the GIS data) was compared to the sample address in routines that ran overnight; in almost all cases, the two locations were identical. Discrepancies greater than a few hundred feet were investigated immediately. More traditional means for identifying falsified interviews (e.g., very short interview times, or interviews conducted at unusual times of day) offer evidence that is less conclusive, or are applied to only a sample of interviews and may take much longer to process (CARI coding, short telephone re-interviews). Thus, GIS data can greatly reduce costs by identifying falsification much more quickly, before the interviewer can falsify more interviews.

MyPeople Portlet. The interviewers reporting to the supervisor are displayed in a ribbon across the bottom of the screen. Their ID Badge photos appear above their names. Data from the mobile device appear under the name: the current or most recent activity (field, travel, administrative), whether the interviewer's device is currently on, and the last date and time the device was used are displayed beneath the name. This gave the supervisor a much more immediate sense of how staff in the region were deployed.

A subsequent enhancement to the MyPeople portrait allows supervisors to click on the interviewer's photo to call up detail about the interviewer's work in the past 24 hours – the outcomes of all contact attempts, the hours worked broken out by activity, etc. – and to contact the interviewer by phone, text, or secure messaging. These data form the basis of a daily review with each interviewer, to provide feedback and plan next steps.

Quality Portlet. The CARI system can produce a huge amount of data on what happens during an interview, but it is also very scalable: the behavior coding can be sized to fit the resources available to the project. (Hicks et al, 2010) In addition to coding individual questions, Westat coders provided a summary evaluation of interview quality. This summary was the basis for a statistical quality control chart, showing the interviewer's performance on the first, second, and *n*th interview coded on a 5 point scale, compared to the national average. Experimental work at Westat has established the value of rapid feedback – both written and verbal -- on interviewer performance. Interviewers were invited to view a report and listen to an audio recording within three days of the interview, and to schedule an appointment with a CARI coder to review performance. Both positive and negative feedback was provided. For those whose performance was sub-par, the coder provided some retraining on the phone. The subsequent interview was coded and compared to the interview before feedback. Trends for individual interviewers were reviewed, and consistently poor performance could lead to termination. The supervisor could investigate specific question asking and probing behavior by drilling down through the quality portlet into the CARI database to call up specific instances of interviewer behavior.

Interviewer reactions to this rapid feedback approach have been very positive, even to negative feedback. They report an enhanced sense they are part of a team dedicated to quality improvement, and are more actively engaged in the process. We carried this approach forward to the trip routing feedback loop described above, with similar results.

Production Portlets. Two production portlets were displayed on the dashboard. The first, "PSU Case Status Summary," showed the number of cases by key dispositions (completed, appointments, refusals, other pending) by primary sampling unit in a series of stacked bars. The second, "Completion vs. Goals," contained two sets of stacked bars. The first set showed the goals for completed interviews for the region by month, compared to the actual production. The second set showed the goals for the week by day (reflecting past experience of the proportion of work completed each day of the week), compared to actual production. The actual production data were refreshed overnight.

The data in these portlets had been available in report format long before the dashboard was developed. The visualization in the portlets was intended to provide a quick read of the KPIs in less space than a report would require.

Alert Portlet. The alert portlet listed anomalies identified in the region's data. Specifications were developed for a limited set of conditions that would trigger an alert. The goal was to draw the supervisor's attention to the most important problems that required action each day. For example, an interview completed in a location far away from the respondent's home would generate an alert. The set of conditions was limited because we were at pains not to overwhelm the supervisor or crowd out routine activities that were necessary for managing operations in the region. However, we sought to identify key problems that, without an alert system, might require hours of supervisor time to review perhaps dozens of reports, and result in some problems escaping the attention of less skilled supervisors.

The alerts system gave the supervisor information about the nature of the problem and a field to record the status of work to resolve it (not reviewed, in progress, resolved, etc.). An enhancement to the MyPeople portlet allowed the supervisor to review all alerts associated with a specific interviewer. In addition, supervisors could choose an interviewer (rather than a region) view of all portlets in the dashboard at once, and field managers could choose a dashboard view of all their regions.

Cost-Benefit Analysis

There are two cost components in the proof of concept for a virtual FTF field center: the development costs and the ongoing operations costs. The development benefited from two major systems that had experienced several years of design work and deployment in production environments before 2015: the CARI coding system and EAGLE. The mobile field operating system (mFOS) with its MyCases and MyDay app had been tested and refined on several projects before 2915, but was first deployed in production this year. The field supervisor dashboard was preceded by a dashboard built for a clinical trials coordinating center contract, but was first developed and deployed for survey operations in 2015. The activities that required the most time and resources were the development of the links between the paradata systems and the portlets; the development and refinement of KPIs and their visualization; the GIS articulation with other paradata; and testing the portlets. The mFOS and dashboard development combined required less than six month of reasonable effort, with input from senior operations and IT staff. Although some adaptation is required for use on other projects, the effort is scalable and generally less than the initial development of the overall system.

The ongoing operations costs are more significant. As deployed with iPhone 5S devices in 2015, the cost is comparable to the laptop computers used for interviewing on many surveys. The connectivity came at a relatively high price. The data plan is based on usage rates that are much higher than used in the Westat implementation; 100 times the level of usage would have been covered without any additional cost. However, as this paper is being written, the consumer market for mobile devices is undergoing a rapid change. Carriers are moving away from bundling equipment charges with data plan charges, and toward separate charges for the devices themselves (with iPhone 6 prices at $500 to $700) and for the data plan. Data plans are increasingly tailored to usage, and charges for low level usage by consumers may level out at about half of 2014 rates.

Various benefits can be identified for the virtual FTF center. Data collection costs were reduced through:

- Less falsification, and less time to identify falsifiers;
- More efficient trip routing;
- Features intrinsic to smart phones, including speech-to-text, internet access and search capability, translation applications, etc.
- Increased field supervisor efficiency; and

- Fewer and more effective contact attempts, due to mobile app efficiency and quality gains, quicker supervisor review and immediate feedback on performance.

It is difficult to quantify savings from some of these activities. No falsification instances were discovered during this proof of concept. Based on past experience, one might have expected to have discovered 2 instances, related to 5 to 10 interviews. Interviewers were informed about procedures to detect falsification, including use of GIS, and we suspect the information about GIS acted as a deterrent. We estimate cost savings at roughly 80 percent compared to conventional falsification detection measures.

We estimate that trip routing for about 20 percent of the field interviewers was improved, mainly for those interviewers with higher than average hours per complete but who were effective in other aspects of their job. Rapid feedback and routine monitoring of routes improved their efficiency by perhaps 1 hour per completed interview. This marks up to about $50 per month per interviewer (when spread across the entire interviewing staff).

A normal field supervisor's work week consists of 20 hours of report calls with interviewers (20 interviewers at one hour per call) and 20 hours of reporting to managers and investigating and resolving problems. We estimate that the alerts portlet saved about 2 hours of supervisor time per week, investigating problems and analyzing reports (which would therefore reduce supervisor costs by about 5 percent, translating into savings of about $20 per interviewer per month. Although this time could be saved by hiring fewer supervisors, in practice we allowed supervisors to shift their time to devote more attention to monitoring and coaching interviewers.

Analysis of contact attempt records using mobile devices compared to traditional entry on laptops is in progress but has not yet yielded preliminary results.

Quality was also improved through:

- Greater standardization of interviewing behavior;
- A higher level of compliance with protocols; and
- Heightened morale and tighter teamwork.

We were able to improve overall interview quality by about 12 percent, using the methods described in this paper. We were able to identify major departures from protocol (such as interviews shorter than 2 standard deviations from the mean) more quickly and take quicker action to correct them, with no intervening management layers. Interviewers reported elevated morale on employee surveys. Several spoke to the benefits of the GIS system, saying it was comforting knowing that someone at Westat always knew where they were in case something should go wrong, and that they appreciated the navigation feature on the iPhones.

Perhaps the greatest benefits of the system were in the new, real time information from the mobile devices and the ability to see multiple aspects of the operation at once, on a single screen. Field supervisors are themselves dispersed workers, and they seldom see the staff they supervise. Communication is typically limited to a phone call each week, and secure messaging. With data available from the mobile devices, constantly refreshed, supervisors could "see" what their staff were doing in real time. This was eye opening. Supervisors reported feeling empowered by this innovation, and were able to act on information much more quickly than with traditional systems. Furthermore, the ability to see data from separate paradata systems on cost, production, and quality in one place fostered a total survey error focus. It made the relationship between different error sources (e.g., nonresponse error, interviewer error) and cost more visible. The dashboard's innate ability to show data from different sources side by side led to some insights that would not have been possible with a series of conventional reports. Although it is not possible to quantify some of these quality improvements, the general sense was that quality improved.

In summary, we estimate that benefits exceeded operating costs by at least 70 percent, while improving quality. We view the proof of concept as successful: FTF virtual centers based on mobile systems and performance dashboards can tighten survey operations management and control costs.

Next Steps

Activities in the next year will focus on two fronts. Next steps for the dashboard at Westat include:

- Developing a **single sign-on** for field supervisors. Currently they must enter their log-in credentials each time they want to dive into a data base from the dashboard, and this interrupts the flow of tracking a problem through several data bases to reach a diagnosis.
- Developing a **strategic dashboard** for senior managers. Project directors and field directors have some needs that go beyond the field supervisors' operational needs, and their perspectives and goals relate more to strategy than to performance *per se*.
- Giving greater emphasis to total survey error, by showing data for sample domains and statistics for key variables during the course of data collection, and developing simulation exercises using the dashboard, to highlight relationships between error sources. This paves the way for implementation of adaptive designs and making them more transparent to field supervisors.
- For the mobile app, we anticipate more training enhancements, to increase acceptance more quickly by more interviewers, and adding more information to MyCases, paralleling information in the basic interviewer management system. The most significant change in development for the app is the simplification of MyDay, to become just a "start work" and "stop work" button. All other information on basic activities can be derived from paradata observations. The GPS, CAPI time stamps, and MyCases data can tell us how much time the interviewer spends traveling, contacting cases in the sample, and conducting interviews. We will infer that any remaining time is spent on administrative tasks such as calls to the supervisor.

One major aspect of telephone research centers was not included in the proof of concept: call scheduling. Advances in FTF paradata have allowed us to develop contact stopping rules informed by each contact attempt's propensity to become a completed interview, and adaptive designs that support assessment of each additional interview's value. However, attempts to guide interviewer activities with directions about case priorities have been plagued by compliance problems. Interviewers often believe their assessment of the cost effectiveness of each contact attempt is better than managers' judgments. We hope to mount an experiment that would compare these two approaches. If it can be established that the managers' judgment truly is more accurate and less costly, we can imagine serving up to the field interviewer the next case to be worked on a given work day, much like a telephone interviewer is served up cases by a central scheduling system during the shift. However, it is more likely that a combination of the two approaches, preserving the field interviewer's local knowledge and control of the work flow, but linking these with suggestions drawn from paradata analytics, will prove to be the most successful solution.

References

Biemer PP, Chen P, Wang, K. Using level-of-effort paradata in non-response adjustments with application to field surveys. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2013;176(1):147-168.

Couper MP. Measuring survey quality in a CASIC environment. Proceedings of the survey research methods section of the American Statistical Association; *1988*. p 41-49.

Few SC. Information dashboard design: displaying data for at a glance monitoring, 2nd ed. Burlingame, CA: Analytics Press; 2013.

Groves RM. Survey errors and survey costs. Vol. 536. Hoboken, NJ: John Wiley & Sons; 2004.

Groves RM, Heeringa SG. Responsive design for household surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society 2006;169(3):439-457.

Hasson M. Efficiency analysis through geospatial location evaluation (EAGLE). Paper presented at the Annual FedCASIC Workshop, Suitland, Maryland; 2015.

Hicks WD, Edwards B, Tourangeau K, McBride B, Harris-Kojetin LD, Moss A. Using CARI tools to understand measurement error. Public Opinion Quarterly 2010;74:985-1003.

Kreuter, F, ed. *Improving surveys with paradata: Analytic uses of process information*. Vol. 581. John Wiley & Sons, 2013.

Kreuter F, Couper MP, Lyberg L. The use of paradata to monitor and manage survey data collection. Proceedings of the section on survey research methods. Joint Statistical Meetings; 2010.

Reifschneider M, Harris S. Development of a SAS dashboard to support administrative data collection processes. Paper produced for the U.S. Energy Information Administration; n.d.

Suchman L, Jordan B. Interactional troubles in face-to-face surveys interviews. Journal of the American Statistical Association 1990;85(409): 232-241.

List of Exhibits