

# **Paradata and Blaise:**

## **A Review of Recent Applications and Research**

*Jim O'Reilly, Westat, USA*

Survey paradata, information about the process of survey data collection, has steadily increased in the breadth and depth of use. From modest beginnings in the 1980s as trace files of the fields entered, keys pressed, and timings during computer-assisted interviews, the value of the information for addressing significant challenges in the survey process has been explored and applied in increasingly diverse ways.

A number of survey organizations have recently implemented or enhanced major management systems with paradata as a central feature. As well, computer-assisted recorded interviewing (CARI) seems on the verge of becoming a powerful new form of paradata with the potential to qualitatively enhance important elements of the survey design and survey collection processes. In this paper we review these recent developments of paradata research and application.

At the 2009 FedCasic conference held in Washington DC a number of presentations focused on using paradata to monitor and control survey quality. Compared to similar sessions in prior years, the presentations this year seemed notable in terms of the maturity of the processes and the production experiences described.

### **PANDA**

Ari Teichman (2009) of the U.S. Census Bureau discussed his agency's Performance and Data Analysis (PANDA) system and its approach to improving survey quality. This comprehensive management system has been implemented for the 2007 American Housing Survey. Key goals of the system are to provide early warnings of interviewer difficulty with key survey concepts and possible falsification.

The PANDA system provides paradata-based reports to managers at every significant level of the project on key production metrics – interview duration, interviews by time of day, interview result, outliers, and more. Among the indicators of possible falsification are high rates of vacant housing units, small household size, interviews conducted at unusual times, short interviews and others. Management reports include summary level reports on Regional Office/Area Totals, Cumulative/Weekly Report, and Average Cases Per Interviewer; and interviewer reports on Highest Non-response Rate for Salary of Reference Person; Highest Number of Regular Occupied Interviews Completed in Less than 20 Minutes; Highest Number of Cases Completed 12:00 a.m. – 7:59 a.m.; Highest Vacancy rates, and many more.

The system also provides the field managers with the ability to drill down in trace files to examine details of interviews. Field managers reported that the most useful features for them were identifying cases completed between 12am and 8am, targeting potential falsification, and searches and access to trace files. Field managers also said they used the system to search for detailed information of interviewers' work, to address potential problems appropriately, and to identify interviewers needing retraining or cases requiring re-interviews.

Teichman said the system have been well accepted by staff at various organizational levels and is being implemented in other major Census surveys beginning with National Health Interview Survey.

### **Statistics Canada—Paradata as a Research Cornerstone**

François LaFlamme (2009) of Statistics Canada discussed her organization's major commitment to data collection research using paradata as the cornerstore. The goals are to understand the process, develop new efficiencies, evaluate new initiatives, and maintain and improve data quality. Data collection is a top concern

because it significantly determines data quality and accounts for nearly half to three quarters of total survey expenditures.

Statistics Canada has developed a paradata warehouse to store call and contact information for both telephone and in-person interviews and administrative and payroll information. As a result, the major cost of the paradata system, the data warehouse, is centralized, minimizing the burden on individual studies and customers, and insuring that all surveys are represented. This, in turn, enhances the research potential and benefits. As well, the paradata system provides in-depth and timely cost information for survey cost analysis.

LaFlamme discussed some early research efforts focusing on CATI and using paradata from a variety of surveys--RDD, cross-sectional, longitudinal, social and agricultural. Analyses were done on the time for contact attempts and system work, contact rates and calling patterns, and the production-cost relationship.

Strategically, Statistics Canada foresees important opportunities for improvements using the system to research such issues as:

- Better use of pre-collection information and information gathered during collection
- Methods used after the first contact
- Development of a responsive design framework combining active management and adaptive data collection
- Predicting collection requirements during collection based on progress metrics

LaFlamme reported on studies of production-cost relationships to provide more useful survey productivity indicators. One example is an analysis of the impact of capping the number of calls on survey cost, which found maximum potential cost savings from 3.1% to 4.2% for longitudinal surveys.

LaFlamme said research is underway on using responsive design for CATI surveys -- adapting the collection process dynamically based on on-going information on the productivity and costs of the surveys. Also, the calling process would be modeled and managed to maximize the likelihood of making contact and conducting an interview. Staffing would also be planned and predicted based on survey progress.

## **Statistics Canada -- POINT**

Mike Maydan (2009) of Statistics Canada presented a thoughtful discussion of the management challenges of the increasingly complex survey environment. In the Statistics Canada regional collection structure, there are multiple data collection, archive, and reporting systems with competing priorities, needs, and dimensions.

To improve coordination and integration, standardized survey reports has been developed based on by case-level paradata for CATI (from the Blaise telephone history file) and CAPI case management events--timing, routing, and outcome details. This provides managers and supervisors with measures of data accuracy with response/non-response rates, non-response follow-up, refusal conversion, and tracing.

Intra-case paradata—audit trails—enables evaluating the “act of collection”, enhancing the evaluation of interviewers. A relatively new application of this paradata is the POINT (pace of interview) system, which focuses on identifying irregular production calls. The system is designed to apply objective performance measures to help identify interviewers for possible retraining or other action. The two indicators are the pace of the interview (field changes per minute) and item non-response (don't-know and refusals).

The level for potentially significant irregular call levels is based on the survey's early collection experience--six hundred calls with at least twenty content field changes. The threshold is 175% of the early collection mean for field changes per minutes and greater than 25% item-non-response.

Reports based on paradata through the previous day provide statistics on the number of interviewers, calls, field changes, interviewing pace, item non-response, and irregular calls by site and interviewers. Audit trails are also used by managers for repairing technical data loss, for post-collection validation and verification, and for time per section report for post-survey analysis and budgeting.

# National Center for Health Statistics – Public Use Paradata and Research

Beth L. Taylor (2009) of the National Center for Health Statistics (U.S.) discussed the public use release of paradata on the data collection process along with the standard public use data file of the 2006 National Health Interview Survey. This appears to be the first time paradata has been included in a public use data release.

The National Health Interview Survey (NHIS) is an annual survey of 35,000 families conducted in-person with telephone follow-up. Detailed contact history information are collected on each visit attempt: description of the attempt, whether reluctance was encountered, and strategies used to complete the interview. For non-contacts the contact description and strategies are kept.

Also, the paradata includes the interview's language, cooperativeness of respondent, interview mode, reasons for interview breakoffs, type of non-interview cases, time of interview and module/section times. From the audit trail comes the time per question, dates, and interviewer notes.

To protect confidentiality and improve usability of the data, various items have been recoded to preclude individual identification. The paradata can be used alone or linked to health data. Examples of possible analyses include:

- Contact attempts and interview completion
- Time of day of interview
- Interview strategies and successful completion
- Characteristics of hard-to-contact families
- Modeling impact of interview mode on health outcome

With the release of the NHIS 2008 public use files, additional items will be added on visit attempts and use of function keys and language of interview. Also, working with the Census Bureau, which administers the NHIS, the PANDA system has been enhanced in the areas of interviewer performance, tracking, and reporting.

Public use of the NHIS paradata so far seems limited to survey methodology students. However, James Dalheimer of the NHIS group said in a recent conversation that internally there is extensive study of paradata around the issue of interviewer performance, using both case level contact histories and audit trail item and section times.

One current emphasis is the process of selecting cases for reinterview. This has been usually been done randomly with very few targeted supplemental reinterviews based on suspected problems. Research is underway on applying statistical predictors for re-interviews. Other research of interest based on paradata is:

- non-response adjustment to weights
- sub-unit response
- high-effort interviews and bias

## Summary

These presentations indicate how far the use of paradata has advanced. Not long ago, the main analytic efforts were on question and section timings, looking for outliers and anomalies that suggest problematic interviewer or instrument performance. Now, organizations are integrating paradata into their core management process, providing carefully developed metrics and reports to various supervisory levels, as well as giving direct access to detailed audit trail information for line supervisors.

Another indicator of the expanding importance of paradata is that for the first time there will be an invited-papers session, "The Use of Paradata in Federal Government Surveys", at the annual Joint Statistical Meetings in August 2009. The papers are:

- "Use of Paradata to Manage a Field Data Collection", Robert Groves (University of Michigan), et al.,
- "Using the Fraction of Missing Information to Monitor the Quality of Survey Data", James Wagner (University of Michigan)
- "Modeling the Difference in Interview Characteristics for Different Respondents", John Dixon (Bureau of Labor Statistics)

- “An Evaluation of Nonresponse Bias Using Paradata from a Health Survey”, Aaron Maitland (National Center for Health Statistics) et al.
- “Subunit Nonresponse in the National Health Interview Survey (NHIS): An Exploration Using Paradata”, James M. Dahlhamer National Center for Health Statistics and Catherine M. Simile (NCHS)

## **CARI -- Paradata’s Next Generation**

Computer-assisted Recorded Interviewing (CARI) provides a direct audio recording of the interviewer and respondent during the interview, enabling the review and analysis of a multitude of potential facets of the interaction, far beyond what audit trails and trace files provide through time stamps, data field changes and other keyboard actions. As with any significant new method CARI’s development and application has emerged and grown over a number of years. The most frequently used CARI capability has been for quality assurance (QA) focused on verification that the interview took place and evaluation of the quality of the interviewer performance.

Among significant recent CARI research is the U.S. Census Bureau’s rigorous CARI field test evaluation, (Arceneaux, 2007). The study is part of broader research effort toward implementing “CARI into all of the Census Bureau’s computer-assisted personal interview (CAPI) surveys.” The 2006 study was conducted with a sample of cases selected from three regions. CARI interviews were recorded for 423 cases. The test found that CARI did not impact production data, the CARI technology functioned properly, recording occurred without detection, and technical problems did not increase. Respondents were very receptive to CARI while interviewers were mixed with 39% comfortable and 23% opposed.

Two negative findings were:

- Recordings were judged to be high quality (rated excellent or good) for 85.6% of cases, while the desired level was 96%.
- The HWS response rate was 81% compared to 90% on a comparable sample from the National Health Interview Survey.

Arceneaux mentions a number factors in the field test study that may mitigate these differences and recommends further research to advance adoption of CARI for “all of the Census Bureau’s CAPI surveys.”

In the 2008 American National Election Survey (Lupia, et al., 2009), CARI was used to verify that pre-election and post-election interviews were completed. CARI files were transmitted overnight to Research Triangle International (RTI).

“Cases were reviewed using a process that focused on address validation, respondent self-identification, confirmation of both respondent and interviewer voices, and consistency of voice(s) across the recordings. Any concerns with one or more of these dimensions prompted further review, including in-person validation in the field if needed.”

The percentage of cases reviewed was adjusted during different phases of data collection: 100% during the initial interviewer incentive period and then at least 10% of completed cases per interviewer. Indications of possible quality or validity problems prompted review of all completed cases for the interviewer. CARI refusal rates were also monitored closely.

Wendy Hicks (2009) and colleagues at Westat and the National Center for Health Statistics presented at the FedCasic Conference their application of CARI in the evaluation of questions and interview performance in a national health survey. They described a full system approach to CARI designed to meet varied research objectives.

In the central office system CARI recordings are linked to the survey data and audit trails. The system permits flexible review in different modes, in which coders can review one question across a series of interviews, or a series of questions in an individual interview. The system provides coding with a structured tool, links to the survey data and operational process data, other built-in quality control functions, and a variety of reports.

Hicks presented statistics on some important survey questions, examining whether a given question was read with a major change, whether the respondent asked for clarification, and whether the response did not match the format expected. Other results presented for selected questions included whether the respondent asked a question

or the answer did not match the format, comparing household and institutional types of respondents. Another report focused on interviewer skills and the percentage of interviewers with difficulty in terms of minor changes from verbatim, major changes from verbatim, not probing when needed, and using leading probes.

The Hicks presentation was greeted very positively by many at the conference. The combination of a comprehensive and effective system to integrate the management, processing, and coding, along with valuable analytic findings, appeared to interest many participants in the potential of CARI.

Another intriguing indicator of the rising interest in CARI as a major new tool for surveys came at the 2009 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. Officials from the Office of Behavioral and Social Sciences Research of the National Institutes of Health in discussing the NIH BIG Data Initiative for surveys of the future cited CARI as one of four notable developments (Mabry and Philogene, 2009).

Finally, the interest and application of CARI is likely to be enhanced by the about-to-be-released implementation of CARI in the Blaise system. As a result the primary survey interviewing system for complex surveys will have CARI built-in and will support highly flexible and dynamic CARI methods and functions.

## Final Thoughts

The evolution of paradata in survey research seems to be reaching an important threshold. Technology, organizational systems to integrate the key processes, and research demonstrating the value of paradata to better understand important and challenging survey elements seem to be aligning. The stage seems to be set for much wider adoption and utilization not only by the largest organizations, but also by other groups with similar needs.

## References

- Arceneaux, Taniecea A.. 2007. "Evaluating the Computer Audio-Recorded Interviewing (CARI): Household Wellness Study (HWS) Field Test". Proceedings of the Survey Research Methods Section, American Statistical Association
- Hicks, Wendy., Brad Edwards, Karen Tourangeau, Laura Branden, Drew Kistler, Brett McBride, Lauren Harris-Kojetin and Abigail Moss. 2009. "A System Approach for Using CARI in Pretesting, Evaluation and Training" *FedCasic Conference, Washington DC*
- LaFlamme, François. (2009) "Overview of CATI Data Collection Research Focused on Developing Operational Strategies for Process Improvement". *FedCasic Conference, Washington DC*
- Lupia, Arthur, Jon A. Krosnick, Pat Luevano, Matthew DeBell, and Darrell Donakowski. 2009. *User's Guide to the Advance Release of the ANES 2008 Time Series Study*. Ann Arbor, MI and Palo Alto, CA: the University of Michigan and Stanford University.
- Mabry, Patricia L. and G. Stephane Philogene, 2009. "Systems Science Methodologies To Protect and Improve Public Health". *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, Nashville TN*.
- Maydan, Michael J. (2009) "Using Paradata to Monitor Survey Quality in Statistics Canada's Regional Data Collection MIS Reports". *FedCasic Conference, Washington DC*.
- Taylor, Beth.L. 2009 "The 2006 National Health Interview Survey (NHIS) Paradata File: Overview and Application". *FedCasic Conference, Washington DC*.
- Teichman, Ari. 2009. "Panda: Using Paradata to Improve Data Quality". *FedCasic Conference, Washington DC*