The term paradata refers to auxiliary data collected in a survey that describe the data collection process (Beaumont 2005; Couper 1998; Couper and Lyberg 2005; Kreuter and Casas-Cordero 2010; Kreuter et al. 2010a). Common examples include the number of calls made to a case, or interview duration. The technology available to today’s survey researcher has enabled the collection of large volumes of paradata in a nearly passive manner. Given this widespread collection of paradata, there are many research areas emerging that could inform both the collection of paradata and paradata-driven innovations for years to come. Motivated by a roundtable discussion at the 2011 Joint Statistical Meetings (JSM) and a recent Survey Practice article on this topic (Lynn and Nicolaas 2010), this article reviews types of paradata, different ways that paradata are currently being used in practice, quality issues concerning paradata, and directions for future research.

TYPES OF PARADATA

The existing literature (see Kreuter and Casas-Cordero 2010) and the 2011 roundtable discussion suggest that there are numerous types of paradata. Importantly, care should be taken not to confuse paradata with more “traditional” auxiliary variables, such as stratum identifiers on a frame, demographic features of Census tracts, or auxiliary information from commercial data sources.

The simplest and most common type of paradata is likely call record data, including dates, times, and counts of call attempts (defined as phone calls in a CATI survey and household visits or phone contacts in a CAPI survey). Counts of call attempts and related measures, such as contact sequences (Kreuter and Kohler 2009), are sometimes referred to as level of effort measures (e.g., Olson 2006), which describe the difficulty of both contacting and obtaining cooperation from a sampled unit. Advanced computing applications also enable the collection of data on the durations of interviews and
administration of individual items (e.g., Couper and Kreuter 2011).

Also quite common is the collection of contact history data, using tools like the Contact History Instrument (CHI) developed by the U.S. Census Bureau, which allows interviewers to record refusal reasons and other household observations (e.g., Maitland et al. 2009). Response history profiles collected in longitudinal surveys (e.g., Kreuter and Jäckle 2008), which describe previous response patterns of units, also fall into this category, along with the various disposition codes (e.g., successful interview, hard refusal, non-contact, etc.) recorded for sampled cases. In establishment surveys, the position of the survey respondent within the establishment (e.g., accountant, information technology manager, farm manager, executive, etc.) or the different types of respondents providing information for a survey (e.g., accountant and executive) provide sources of paradata that could explain variance in survey responses.

Other types of paradata capture information about survey interviewers. The development of computer audio-recorded interviewing (CARI) applications (e.g., Hicks et al. 2010) has also enabled the collection of verbal paradata (e.g., Conrad et al. 2008; Ehlen et al. 2007; Groves et al. 2008; Jans 2010), describing features like pauses, changes in voice pitch, or incorrect reading of questions by interviewers during interviews. Verbal paradata can also be collected on respondents, but existing studies have primarily focused on using these data to study interviewer performance. Interviewers are also frequently tasked with recording various observations during data collection (e.g., Kreuter et al. 2010b), including features of neighborhoods (e.g., Casas-Cordero 2010), households (e.g., Pickering et al. 2003; Tipping and Sinibaldi 2010; West 2011a), and individuals (e.g., West 2011a). Interviewers may also be asked to judge features of respondents in telephone surveys (e.g., McCulloch et al. 2010). Current research is also considering the use of GPS applications to monitor interviewer travel patterns, as a supplement to hours reported by interviewers on timesheets for various tasks (e.g., Wagner and Olson 2011).

Advanced computer hardware and software also enable the collection of unique paradata describing respondent behaviors during the process of responding to a survey. These include indicators of respondent behavior during self-administered ACASI portions of interviews (Couper et al. 2009), eye-tracking measures (Galesic et al. 2009; Graesser et al. 2006), and keystroke data (e.g., the PANDA system at the U.S. Census Bureau; see Jans et al. 2011). Initial research on web browsing behaviors has also considered mouse-tracking measures (e.g., Arroyo et al. 2006; Guo and Agichtein 2008; Heerwegh 2003; Mueller and Lockerd 2001; Rodden et al. 2008), which may prove useful for studying response quality in web surveys. The collection of these paradata will likely offer insights into the behavior of survey respondents and improve the administration of survey questions.
**WHAT ARE PARADATA USED FOR?**

Similar to survey variables, paradata should be collected for some purpose. The collection and archiving of paradata in the absence of a clearly defined purpose (e.g., improving survey operations or data quality) is a waste of computing system resources. The roundtable discussion and various sessions at recent AAPOR and JSM conferences have revealed some interesting uses of paradata for attacking important survey problems.

When they are collected for both respondents and nonrespondents, paradata are used to model respondent behavior and predict response propensity (e.g., D’Arrigo and Durrant 2011; Durrant and Steele 2009; Kreuter and Kohler 2009; Lynn 1996). Accordingly, in a responsive design framework (Groves and Heeringa 2006), paradata are used to prioritize cases with high predicted response propensities (e.g., Lepkowski et al. 2011), saving costs and increasing response rates (e.g., Laflamme and Karaganis 2010b). Paradata associated with both response indicators and key survey variables could be used for post-survey adjustment of estimates for nonresponse (Kreuter et al. 2010b; West 2011a), and prior work has examined the possibility of using sequences of call attempts for nonresponse adjustment (Kreuter and Kohler 2009). When interviewers record paradata for respondents only (e.g., impressions upon completion of an interview), calibration methods may also prove useful for nonresponse adjustments (Kott 2006).

Paradata are also used for internal monitoring of data quality over the course of a data collection (e.g., Jans et al. 2011; Sirkis et al. 2011), studying possible measurement error after data collection (e.g., Bassili 2003; Knowles and Condon 1999), and evaluation of interviewer performance (e.g., Laflamme and St-Jean 2011; West 2011b). For example, the U.S. Census Bureau is currently considering the application of statistical process control techniques to paradata collected over time, to indicate possible issues with data quality requiring intervention (Sirkis et al. 2011).

Roundtable participants agreed that the monitoring of interviewer travel behaviors using GPS systems could allow interviewers to travel more efficiently. Notably, GPS tracking of interviewers in personal interview surveys would also enable improved classification of smaller area segments (e.g., urban / rural), especially in primary sampling units that are very heterogeneous in nature (where available auxiliary measures at higher geographic levels may not accurately represent the smaller area segments).

**STUDIES EXAMINING THE QUALITY OF PARADATA**

The collection of paradata may not be worthwhile if the resulting data are of reduced quality. Error-prone paradata could lead to biased nonresponse adjustments (Biemer et al. 2011; West 2011a), erroneous interviewer evaluations, increases (rather than decreases) in survey costs, and decreased quality of survey data. Studies examining the error properties of paradata are
slowly beginning to emerge, but more are needed to justify the large quantities of paradata that survey agencies are collecting.

Several studies to date have considered direct (i.e., using validation data) or indirect (i.e., reliability-driven) evaluations of interviewer observations (see West 2011a, for a review), finding that the accuracy and/or reliability of interviewer observations can range from quite low (<10%) to relatively high (92%). Although preliminary studies have suggested that computer-recorded call record data tends to be of high quality (Laflamme and Karaganis 2010a), other studies have suggested that call record data can have reduced quality, with under-reporting of call attempts or incorrect reporting of telephone contacts as in-person contacts by interviewers being fairly common (Biemer et al. 2011; Laflamme and Karaganis 2010a). Disposition codes may also be reported incorrectly by interviewers, leading to erroneous contact history profiles (Laflamme and Karaganis 2010a), and interviewers may charge time for particular tasks to different surveys. Initial work has also suggested that the inter-rater reliability of verbal paradata codes may be low (Jans 2010).

The extant work in this area therefore suggests that the error properties of paradata require a more consistent and dedicated research focus, but possible trade-offs between efforts to increase the quality of the paradata and the quality of the actual survey data collected also need to be a part of this research.

**FUTURE DIRECTIONS FOR RESEARCH ON PARADATA**

This is a critical time for survey researchers to rigorously examine the quality and utility of paradata. The roundtable discussion identified several important research questions that deserve future attention from survey methodologists and survey statisticians:

- What are the statistical implications of error-prone paradata for various nonresponse adjustments (e.g., Biemer et al. 2011), and how accurate do the paradata need to be to avoid attenuation of possible bias reduction?
- Is the type of respondent in an establishment survey predictive of response propensity and other key survey variables?
- Does an increase in response rate brought about by using paradata in responsive survey designs also lead to a decrease in nonresponse bias, given that these indicators are generally independent (Groves 2006; Groves and Peytcheva 2008)?
- Within a survey agency, are paradata being collected in a standardized manner and for well-defined purposes, with analysis plans for the paradata in place?
- Is the collection of additional paradata simply adding burden to interviewer workloads or information systems, without engendering increases in survey data quality and decreases in survey costs?
• Does GPS tracking of interviewers modify their behaviors?
• What role should post-interview observations / reports by the interviewer play in increasing data quality or improving survey operations?

A consistent concern arising in the 2011 roundtable discussion and mentioned by Lynn and Nicolaas (2010) was a lack of communication between survey managers and field staff about paradata. Managers need to emphasize the reasons why paradata are collected, because many interviewers have no idea why they are collecting additional measures that are seemingly unrelated to the survey. If more published studies and interventions can establish the value of paradata, field researchers need to be aware that the collection of this information may be as important as collection of the actual survey variables. One promising solution to this problem could be the presentation of agency-specific educational seminars on paradata. Both the U.S. Census Bureau (Kreuter 2011) and Statistics Canada (Laflamme, personal communication) have developed seminars in this area that have been extremely successful to date, and other agencies may consider building on these models.

This article has presented a review of current practice and research on the collection of paradata. There are a number of possible directions for future research, and the roundtable participants were excited about the possibilities that the collection of paradata represent and the interesting research that future years may bring. As a whole, the survey methodology literature would surely benefit from more published reports describing the utility of paradata across a variety of survey applications. Importantly, many interesting studies presenting applications of paradata have not found their way into the published literature. Agencies conducting surveys that were not explicitly mentioned or cited in this article are likely using paradata on a daily basis for their survey operations, and omission of references to examples at other agencies were not intentional.

ACKNOWLEDGEMENTS

I sincerely thank the participants in the roundtable on Measurement Error in Survey Paradata at the 2011 JSM, including Wendy Barboza (NASS), Jonaki Bose (SAMSHA), Nancy Clusen (Mathematica), Scott Fricker (BLS), James Harris (NASS), Matt Jans (U.S. Census Bureau), Francois Laflamme (Statistics Canada), and Roy Whitmore (RTI). I would also like to thank Frauke Kreuter, Francois Laflamme, and Matt Jans for extremely constructive thoughts and comments on an earlier draft of this article.
REFERENCES


