*This is the English manuscript of "Para deta gainen no tanjo to fukyu" by Mick P. Couper (translated to Japanese by Wataru Matsumoto), in *Shakai To Chosa (Advances in Social Research)*, 18:14-26, Tokyo: Shakai Chosa Kyokai (Japanese Association of Social Research), published here courtesy of the author.
*本稿は, ミック・P・クーパー, 松本渉訳, 2017,「パラデータ概念の誕生と普及」『社会と調査』一般社団法人社会調査協会 18: 14-26.の翻訳前の原文を、ミック・P・クーパー氏の許諾のもと公開したものである。

Birth and Diffusion of the Concept of Paradata

Mick P. Couper
Research Professor, Survey Research Center, University of Michigan

Introduction

In this paper I present a brief overview of the history and development of the concept of paradata. I begin with the initial development of the term and its expansion to a wide variety of different types of paradata. I review some examples of paradata, and the uses to which paradata has been put. This is a mostly personal reflection on how paradata has evolved since I first introduced the term in 1998. It is particularly gratifying that this special issue of the journal is devoted to paradata. This helps to demonstrate the universal use of this concept, and also to encourage greater use of paradata among researchers in Japan.

The Origins of Paradata

The idea of paradata or process data has been around for a long time, but the term itself dates back to 1998. It was developed in response to a specific situation, but has since been expanded to cover a wide variety of data types. I developed the term specifically to refer to the different types of data automatically generated by computer-assisted interviewing (CAI) software (Couper, 1998). At that time (and still) some software products called these keystroke files, while others called them trace files, audit trails, or log files. What they had in common was the capture of all keystrokes entered by interviewers as they administered surveys using CAI systems. They were all measuring similar things, so I proposed a common term to facilitate communication across users of different systems.

The original purpose of capturing these paradata was to detect and recover from system errors, but their utility for providing insights into the survey process was soon uncovered. The initial focus was on measurement error issues, providing an indirect record of the interaction between interviewers and computers as they conducted survey interviews (see Couper, Sadosky, and Hansen, 1994; Couper, 1995; Couper, Horm, and Schlegel, 1997). Couper, Sadosky,

and Hansen (1994, p. 849) wrote: “The keystroke files provide us with a wealth of data on interviewer-computer interaction during the survey interaction. Our work is focused on understanding such behavior, and finding ways to use these files in a systematic way to evaluate aspects of interviewer performance on CAI, and to identify needed improvements in CAI instruments or interviewer training to minimize the errors or inefficiencies committed by interviewers in using CAI.”

The development of interest in paradata paralleled the introduction of technology to survey data collection. The focus of the early paradata analyses was on how interviewers interacted with the CAI software. While computer-assisted telephone interviewing (CATI) had been around since the 1970s, computer-assisted personal interviewing (CAPI) was introduced in the late 1980s and raised concerns about the ability of interviewers to manage complex software on laptop computers while interacting with respondents. The early work looked at both interviewer-level behavior (to identify difficulties that interviewers may be facing) and at item-level behavior (to identify particular questions or functions that produced difficulties). For example, Couper, Hansen, and Sadosky (1997) found that interviewers’ computer experience was related to the use of various function keys to perform non-standard actions. At the item level, paradata analysis was used to reveal what items generated the most requests for help. These analyses were focused on the usability of the computerized survey instruments – either pointing to the need for additional training of interviewers, or to identify problematic questions or function that needed redesign. Couper, Hansen, and Sadosky (1997, p. 281) noted that “… keystroke files may reveal as much (or more) about the inadequacies of the instrument or system as they do about differences in interviewer performance of capabilities in using the system.”

While I first introduced the concept of paradata in my presentation at the Joint Statistical Meetings in 1998, I was discouraged (by my colleague and mentor, Robert Groves) from introducing the new term, so it did not appear in the published proceedings paper (Couper, 1998). The term “paradata” did not appear in print until two years later. In a paper on usability testing of survey instruments (Couper, 2000, p. 393-394), I wrote: “…one of the benefits of computer-assisted data collection is that a great deal of automated data is generated by the process and these data can be used in turn to evaluate the process. I term these sources of information ‘paradata’ (auxiliary data describing the process) to distinguish them from metadata (describing the data). Sources of paradata include case management information such as response rates, number of calls per case, average interview length, and so on. A most useful type of paradata for usability evaluation is keystroke or trace files (also called audit trails).”

At about the same time, analysis of CAI paradata was used to examine respondent use of audio computer-assisted self-interviewing (ACASI), as that new technology was being introduced (see Caspar and Couper, 1997). Interviewer observations were also being used to evaluate computer-assisted self-interviewing (CASI; see Couper and Rowe, 1996). The focus on

respondents in ACASI and CASI previewed the growth of interest in paradata with the development of Web surveys (see below).

The concept of paradata did not really gain traction or broad interest until two parallel developments in the late 1990s and early 2000s. The first was the development of Web surveys, and the second was the renewed interest in call-record data to manage and monitor large-scale surveys. I describe these developments in more detail below, before returning to the issue of the definition and meaning of the concept of paradata.

The Concept Expands

As noted above, two developments led to expansion of the use of the term paradata, and also expansion of the definition of what constitutes paradata. I describe each of these developments briefly in this section.

The first development was the introduction and rapid growth of Web surveys in the late 1990s. The kinds of keystroke data that could be captured in computer-assisted surveys could now also be captured directly from respondents as they completed Web surveys, giving us insights into their behavior that we could previously not obtain from paper self-administered surveys.

Jeavons (2001) was the first (to my knowledge) to apply the term paradata to Web surveys. As he wrote (p. 177), "The web provides researchers with a gold mine. Every click, response, page movement and delay can be logged and used for analysis." Heerwegh (2003; see also Heerwegh, 2008, 2011) is credited with distinguishing between server-side paradata and client-side paradata (CSP) and for making JavaScript code available for capturing CSP in Web surveys. Client-side paradata is the self-administered equivalent of keystroke files for computer-assisted interviewing. Heerwegh's contributions led to a rapid expansion of the use of paradata for Web surveys. Others have expanded on Heerwegh's code, and made generic scripts available for use in Web surveys (see Kaczmirek, 2009; http://kaczmirek.de/ucsp/ucsp.html). Web survey paradata is now an integral part of online survey data collection (see the paper by Noboru Ohsumi in this issue). Paradata used in Web surveys is mostly focused on data quality or measurement error, typically at the respondent or item level.

The second development that led to the widespread use of paradata was Groves and colleagues' use of the term in conducting the National Survey of Family Growth (NSFG). Groves et al. (2005) made extensive use of paradata to manage Cycle 6 of the NSFG in 2002. While call-record data had been used in computer-assisted telephone interviewing (CATI) for many years (see, e.g., Weeks et al., 1980; Greenberg and Stokes, 1990), the rise of computer assisted personal interviewing (CAPI) and the move from paper "cover sheets" (containing sample address information, on which interviewers recorded details of calls or visits) to computerized sample management systems, generated renewed interest in such paradata. Ironically, given Groves'

initial resistance to the term, it is partly his efforts to popularize the concept that led to its widespread acceptance in survey research.

Groves and Heeringa's (2006) paper on responsive design further served to bring the notion of call-record paradata into the mainstream of survey research. Given the increasing difficulties of contacting sample households and gaining cooperation from sampled persons, and the rising costs of survey data collection, researchers started making more use of existing paradata, and collecting additional observational data to monitor and manage surveys in real time. This approach – variously referred to as responsive or adaptive design (see Calinescu and Schouten, 2016; Couper and Wagner, 2011; Schouten, Calinescu, and Luiten, 2013; Wagner et al., 2012) – is now widely used in North America and Europe. Paradata used in interviewer-administered surveys is largely focused on coverage, sampling, and nonresponse errors, typically at the sample case level. As Groves and Heeringa (2006) noted, responsive or adaptive designs require enriched process data or paradata to guide decisions made during data collection. Such paradata serve as proxy indicators of costs or errors.

Definition and Meaning of Paradata

Given the evolution of the concept of paradata, let me return to the issue of the meaning and definition of the term. Here I briefly describe the origin of the term "paradata" and discuss how its meaning has evolved over time. The prefix "para" used to create the word "paradata" comes from the Latin, meaning "next to" or "beside." It has also come to mean "auxiliary" or "secondary" (as is paramedic, paralegal). I chose the term to mean "auxiliary" or "additional" data, to add to the already existing terms, "data" (what surveys are designed to collect or produce) and "metadata" (data about the data, including survey descriptions, codebooks, value labels, etc.). I coined the term to focus on the automatic byproducts of computer-assisted systems, i.e., auxiliary data that provides information about the survey process.

Despite its widespread use, the term "paradata" was never formally defined. Its more narrow conception as automated data about the process of interviewing has grown to encompass a variety of other types and sources of data. For example, as Olson (2013) notes, "Paradata have expanded to include any information recorded by interviewers that is external to data from the questionnaire itself." Thus interviewer observations, whether on the neighborhood, the housing unit, household members, or the respondent, are now viewed as paradata.

Groves and Heeringa (2006, p. 448) use an even-more expansive definition: "[Paradata] can include data collection administrative data, such as records of contact attempts on sample cases, travel distance and time to reach sample cases and hours spent by interviewers on different tasks during interviewing. They can also include sampling frame data that might be useful in allocating data collection resources, such as the level of urbanicity of a sample unit …, the existence of multiunit structures … or demographic information that might be recorded on the sampling frame. They can also include *de novo* observations taken on sample units that

might be proxy indicators of how difficult the sample case might be to gain survey participation, likely characteristics of the unit on key survey variables or whether the measurement errors of completed cases might be distinctive." In their view, paradata also include other types of auxiliary data (e.g., census data, aggregate-level commercial data) that are used to manage and evaluate the survey process. In this sense, paradata are defined by their use or purpose, rather than by their source. Some view paradata as a subset of a larger set of process data, while others see the two terms as equivalent (see Couper and Lyberg, 2005; Kreuter, Couper, and Lyberg, 2010).

The boundaries between data, metadata, and paradata are sometimes blurred. For example, survey completion times are often viewed as paradata, but if the resulting variable is included on the analytic dataset and used in substantive analysis, it could be considered survey data. Similarly, some survey metadata (e.g., response rates, average completion times) are aggregated from the raw paradata but are used to describe the survey as a whole. Scheuren (2001) drew a distinction between micro and macro paradata. He viewed micro-level paradata as detailed data about individual cases, calls, or items, while macro-level paradata are viewed as the summary descriptive statistics about survey data collection (such as response rates, breakoff rates, completion times, etc.). This distinction has not gained traction, and macro-level paradata are now usually viewed as survey metadata. In that sense, I erred by including "response rates" in my description of paradata (Couper, 2000). Response rates are derived from paradata, but are typically not viewed as paradata themselves.

Thus, paradata are defined in part by their source (i.e., data that are collected or generated as a byproduct of the survey process), but also in terms of their use (i.e., data that help us understand the process of data collection).

There continues to be disagreement over what is included under the rubric of paradata and what is excluded. The original meaning of the term focused on automatically-generated data. But call-record data includes a mix of such system data (e.g., date and time of the call) and information entered by interviewers (e.g., the outcome of the call, the reasons the sample person gave for not wanting to participate). Given this, other types of interviewer observations are now generally viewed as paradata. My own view is that these distinctions are not important, and a precise definition of paradata is not necessary, as the tools being used to conduct surveys are constantly evolving, and the kinds of data that can be generated by systems are expanding.

The broadest definition of paradata is found on Wikipedia: "The paradata of a survey are data about the process by which the survey data were collected" (see https://en.wikipedia.org/wiki/Paradata). I am comfortable with such a broad view of paradata, in part because constant change in survey methods and technology will render a more narrow definition obsolete in a few years. Thus, despite the varying and evolving definitions of paradata, the concept remains a useful one. As we see in the section below, the concept continues to expand and evolve.

Paradata Comes of Age

The notion of paradata has thus followed two largely separate but parallel strands. One strand focuses on interviewer observations and call-record data from interviewer-administered surveys, often generated through computerized sample management systems. These paradata often focus on errors of representation (including sampling, coverage, and nonresponse error), and on the efficient management of large-scale survey projects. The other strand focuses on server- and client-side paradata from Web surveys, and on keystroke-level paradata from computer-assisted surveys. These paradata are largely focused on measurement error.

In the latter part of the 2000s, several workshops and short courses were held on survey paradata that helped bring the two strands together and legitimize the collection and analysis of paradata to evaluate and improve the survey process. These events led to the publication of an edited volume (Kreuter, 2013) and a special issue of the *Journal of the Royal Statistical Society, Series A*, on "The Use of Paradata in Social Survey Research." These, together with a Wikipedia page (see https://en.wikipedia.org/wiki/Paradata) and an entry in the *Encyclopedia of Survey Research Methods* (see Heerwegh, 2008), helped establish paradata as part of the lexicon of survey research and methodology. Workshops and short courses on paradata are being offered at conferences on survey research methods (e.g., the Joint Statistical Meetings, the European Survey Research Association [ESRA] conferences, and the American Association for Public Opinion Research [AAPOR] conferences), and webinars on paradata are being offered by these and other professional associations.

As paradata became part of mainstream survey research data collection, we have seen efforts to enhance and extend the kinds of data that can be captured, and to make the data more useful for both management of surveys and for evaluation and methodological research.

The use of the term paradata has also expanded beyond the survey domain. With the widespread increase in digital transactions, an increasingly amount of process data are being captured and analyzed. When such data are used for surveillance – for example, the U.S. National Security Agency's analysis of telephone log files (who called whom, duration of call, etc.) – or big data analytics, they are usually referred to as metadata. Similarly, Web server log files or user behavior on social media platforms are increasingly serving as a source of data, both to improve the site and to understand user behavior. But when such data are used for research purposes (often in combination with surveys), they are often referred to as paradata. One example is from the field of online interventions (see, e.g., Glasgow et al., 2007; Strecher et al., 2008; Resnicow et al., 2010). Paradata are used to evaluate the extent of participants' engagement and use of the website, to see how long they spend on the site, how many pages they visit, which optional features they use, and so forth. This is useful in the same way that measuring drug dosage or exposure to a treatment is in a controlled trial, and is used to evaluate reasons for the success of such online interventions.

A second example comes from the field of learning resource analytics. The National Science Digital Library adopted the term paradata in 2010 to refer to user interactions with digital learning objects (see https://en.wikipedia.org/wiki/Paradata_(learning_resource_analytics)). The NSDL distinguishes between paradata and metadata in similar ways to that in the survey domain (see https://wiki.ucar.edu/display/nsdldocs/Paradata). The methods and tools used are very similar to those in Web survey paradata (see, e.g., Kilburn and Earley, 2015). However, despite the increased use of Web server log files and cookies to track and study user behavior, the term paradata is still largely limited to the survey domain.

Challenges and Opportunities

While the utility of various types of paradata for evaluating and managing surveys has been demonstrated, there are still a number of challenges that limit the broader use of survey paradata.

One of the barriers to greater use of paradata is that the format of the paradata is still peculiar to the particular software or system being used. In other words, there are no universal standards for paradata. In a way, this is similar to the early days of audit trails, trace files, log files, and the like, where every system generated its own version of paradata. For Web surveys, this is still true of server-side paradata (which is managed by the survey software), but the client-side paradata (using standard JavaScript queries) is increasingly generating data that can be compared across systems and surveys. Two examples of this are Heerwegh's client-side paradata (CSP) code (see https://perswww.kuleuven.be/~u0034437/public/csp.htm) and Kaczmirek's universal client-side paradata (UCSP) code (see http://kaczmirek.de/ucsp/ucsp.html). Specific tools, such as Diedenhofer and Musch's (2014) PageFocus code, are being made available on GitHub, a code-sharing service (see https://github.com/deboerk/PageFocus). Developments like these will encourage greater use of client-side paradata in Web surveys. For interviewer-administered surveys, the paradata capture is still specific to the survey software and case management systems used, but there is increasing sharing among users of specific software (e.g., Blaise's audit trail and session systems; see http://blaise.com/).

Another barrier to more widespread use of paradata is the relatively unstructured nature of paradata. This is true of both call-record paradata and keystroke paradata. For call-record data, the observations are often hierarchical, with some outcomes (e.g., interviews, refusals) being conditional on others (e.g., contact, eligibility). There are also an uneven number of observations per case, with some cases being interviewed on the first call attempt, and others taking many call attempts to reach a final disposition. Recent advances in call-record analysis are addressing these challenges (see, e.g., Durrant, 2016; Durrant and D'Arrigo, 2014; Durrant,

D'Arrigo, and Steele, 2013; Durrant, Maslovskaya, and Smith, 2015; Durrant and Steele, 2009; Steele and Durrant, 2011).

For keystroke-level paradata or client-side paradata from Web surveys, the raw data is often in string format, and requires careful parsing to extract useful information. A single Web page (for example) could generate a wide variety of different behaviors across respondents – different numbers of mouse clicks, scrolling behavior, and even repeat visits to the same page, with or without changing answers. Determining the desired level of aggregation (e.g., page visit, page, survey, respondent) is an important step in utilizing such paradata.

This necessitates a great deal of pre-processing to prepare the data in a form suitable for analysis. It also requires different analytic tools. Progress is being made on both of these fronts. As the capture of paradata is becoming more standardized, and as greater use is being made of paradata, tools are becoming available to extract the relevant information. Increasingly survey organizations are building computerized case management systems to exploit the paradata being captured, and to automate the processing of such paradata for real-time management of surveys (see, e.g., Hubbard et al., 2010; Thalji et al., 2013). The Michigan Survey Management System (MSMS) developed at the Survey Research Center at the University of Michigan is built around capturing and processing paradata to management mixed-mode survey data collection. Similarly, developments in analytic methods are benefitting the analysis of paradata (as noted above). Several of these efforts are documented in Kreuter (2013).

A related challenge is finding ways to make survey paradata more readily available to outside researchers to encourage secondary analysis. While preparing paradata files for analysis by others takes considerable time and effort, making such files available may increase the use of paradata and lead to the development of standards for extracting and aggregating paradata, which in turn will further stimulate analysis.

Some organizations are making summary paradata available for secondary analysis. Interviewer observations are often made available as part of the survey data. More detailed call-record data is something made available on request, or released (in summary form, with variables suppressed or collapsed for disclosure reason) as a public-use dataset. One example is the Paradata File for the National Health Interview Survey (NHIS; see http://www.cdc.gov/nchs/nhis/nhis_2014_data_release.htm). This public-use data file contained key variables from three sources: 1) the Contact History Instrument (CHI), that collects data from the interviewer about each contact attempt (i.e., call-record paradata), 2) debriefing questions completed by the interviewer after the interview (i.e., observational paradata), and 3) data and time information for each module of the instrument (i.e., timing paradata). These are aggregated up to the case level. The increased availability of paradata, whether in public-use forms or as restricted data, will help facilitate broader use of paradata by researchers.

With the increased use of paradata – especially involving interviewer observations or requiring interviewer recording of call records – have come concerns about the quality of such paradata. Many observations require subjective judgements by interviewers, and there is evidence of variability in how well interviewers make these judgements (e.g., Casas-Cordero et al., 2013; Eckman, Sinibaldi, and Montmann-Hertz, 2013; Sinibaldi, Durrant, and Kreuter, 2013; West, 2013). A recent line of research is focusing on evaluating the quality of interviewer-generated paradata and on efforts to improve the quality of the data by improving interviewer training and feedback (see, e.g., West and Kreuter, 2013, 2015, 2016). These evaluation efforts have led to improvements in the quality of the paradata being provided by interviewers, whether in the call-record paradata or in interviewer observations.

Another challenge with the use of paradata is questions about informed consent. While call records and interviewer observations don't involve interactions with sample persons or respondents, the recording of reasons for non-participation (for example) and observation of aspects of the interview or the respondents home could raise questions of informed consent, especially if such observations become part of the public use data set. For keystroke-level data and Web paradata, the argument is made that the capture of such log files is a routine procedure to recover from system failures, with the increasing analytic use of the data for other purposes, and with the enhancement of paradata to capture additional details of the process, some have raised ethical concerns about the use of these data for research purposes.

Heerwegh addressed this issue early on in his development of client-side paradata (see https://perswww.kuleuven.be/~u0034437/public/csp.htm). He wrote: "Use CSP only for genuine (methodological) research needs. Do not use CSP simply to 'spy' on your respondents, and never use the information from CSP to replace the final answers given by the respondent on the web survey." Singer and Couper (2011; see also Couper and Singer, 2013) attempted to explore the impact of informing respondents about the capture of paradata. Their work demonstrated the difficulty of explaining to respondents what paradata are and how they are used.

The current view on this issue is that such paradata don't reveal anything personal about the respondent above and beyond what is already provided in the survey, and as long as the data are being used for methodological purposes (rather than for evaluating respondents), use of paradata without informing respondents is acceptable. This is similar to the use of cookies to improve the Website experience for respondents without tracking them across sites. But as timing paradata are increasingly used to classify respondents as "speeders" or "satisficers" (i.e., those who may not be putting effort into reading and answering the survey questions), the issue of whether (and how) to inform respondents of paradata collection may resurface.

The above barriers or challenges to broader adoption of survey paradata can also be seen as potential opportunities. As more researchers make use of paradata and share their research, the demand for paradata will increase. This will likely lead to efforts to develop standards and

systems for the collection, processing, and analysis of paradata. This in turn will make it easier for researchers to document the paradata captured and share paradata files with others, further increasing their use.

A final challenge – and opportunity – is that of bringing together the two strands of paradata use. Some evidence of this is being seen in panel and longitudinal studies, where paradata focused on measurement error (e.g., item missing data rates, response times) is being used to predict subsequent nonresponse or attrition. Similarly, some researchers are exploring the trade-off between increasing response rates by converting reluctant respondents and the quality of the data provided by those respondents. Again, more work along these lines will serve to bring the two strands of research together, fostering a broader total survey error perspective rather than a focus on a single source of error.

The Growing Value of Paradata

As responsive or adaptive design continue to be used, the demand for paradata will continue to grow. Similarly, as computers are increasingly used for all aspects of survey data collection, the amount of paradata generated will continue to expand, further fueling the growth of paradata. Paradata are required to manage complex multi-mode surveys in real-time. As paradata become an increasingly integral part of the process of large-scale data survey collection, we will see continued efforts to evaluate the quality of paradata, and to extend the amount and type of paradata that can be collected and used, not only to evaluate sources of error, but also for nonresponse adjustment. The initial focus of paradata on measurement error has expanded to include coverage error, nonresponse error, sampling error, and survey costs and effort.

Another example of the growing importance of paradata, along with the evolution of what constitutes paradata, comes from the world of Web surveys. As users are increasingly using smartphones to complete Web surveys, there is a growing demand for paradata to identify such users, and to identify problems encountered by mobile users as they attempt to answer surveys on mobile devices. This started with analysis of user agent strings to identify the devices, browsers and operating systems used by respondents (see Callegaro, 2010). But Web survey paradata has further expanded to include scrolling, pinch-and-zoom, orientation changes, and other user actions on smartphones (see, e.g., Couper and Peterson, 2016). Similarly, paradata are being used to detect respondent multitasking, particularly on smartphones (see Sendelbah et al., 2016; Diedenhofen and Musch, 2014). With the increasing need to optimize Web surveys for mobile devices, paradata are being used in real-time to deliver such instruments. As the devices and technologies available to respondents continue to expand, we are likely to see increased use of paradata, both to enhance the user experience, and to understand how effectively they are able to use complex survey instruments on a variety of platforms.

Despite the recent increase in use of paradata, it is important to remember that it serves a complementary role. Paradata do not replace other sources of information, including frame data and administrative or other auxiliary data. But paradata provides additional sources of information to evaluate and improve the design and conduct of surveys. Similarly, paradata do not replace other methods of evaluation or quality control. For example, paradata are useful in identifying interviewers who may be facing special challenges, or identifying geographical areas where extra attention is needed. Paradata can identify interviewers who may need additional monitoring using more expensive methods. Similarly, for questionnaire design, paradata help point to areas of potential problems. These issues can then be addressed using more intensive methods such as cognitive interviewing, eye tracking studies, or experiments. Thus, paradata provides an additional set of tools for the survey researcher. Paradata helps quantify some of the issues facing large-scale surveys, so that resources and attention can be directed at the most important challenges.

Summary and Conclusion

Over a period of nearly 20 years, the concept of paradata has grown from a narrow term used to describe a specific part of the computer-assisted interviewing process to a broadly-used term applicable to many modes of data collection – including CAPI, CATI, Web, and paper, as evidenced by the papers in this special issue. It has also grown from a focus on measurement error and usability of computerized survey instruments to cover many other aspects of the survey process. The concept of paradata is now a mature one and an accepted part of the lexicon of survey research. The collection, processing, and use of paradata continue to grow. As surveys become increasingly more complex, and as more digital data on the process becomes available, paradata will serve an increasingly central role in the design, management, and evaluation of surveys. A number of challenges still remain to maximize the utility of paradata for these purposes, but research and development in this area is moving the field forward.

References

Calinescu, M., and Schouten, B. (2016), “Adaptive Survey Designs for Nonresponse and Measurement Error in Multi-Purpose Surveys.” *Survey Research Methods*, 10 (1):35-47.

Callegaro, M. (2010), “Do You Know Which Device Your Respondent Has Used to Take Your Online Survey? Using Paradata to Collect Information on Device Type.” *Survey Practice*, 3 (6).

Casas-Cordero, C., Kreuter, F., Wang, Y., and Babey, S. (2013), “Assessing the Measurement Error Properties of Interviewer Observations of Neighbourhood Characteristics.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176 (1): 227-249.

Caspar, R.A., and Couper, M.P. (1997), "Using Keystroke Files to Assess Respondent Difficulties with an ACASI Instrument." *Proceedings of the Section on Survey Research Methods*, American Statistical Association. Alexandria: ASA, pp. 239-244.

Couper, M.P. (1995), "New Tools for Evaluating Interviewer Performance in CAI." Paper presented at the International Field Directors and Field Technologies Conference, Fort Lauderdale, May.

Couper, M.P. (1998), "Measuring Survey Quality in a CASIC Environment." Invited paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, August.

Couper, M.P. (2000), "Usability Evaluation of Computer-Assisted Survey Instruments." *Social Science Computer Review*, 18 (4): 384-396.

Couper, M.P., and Hansen, S.E. (2001), "Computer Assisted Interviewing." In J.F. Gubrium and J. Holstein (eds.), *Handbook of Interviewing*. Thousand Oaks, CA: Sage, pp. 557-575.

Couper, M.P., Hansen, S.E., and Sadosky, S.A. (1997), "Evaluating Interviewer Performance in a CAPI Survey." In L. Lyberg et al. (eds.), *Survey Measurement and Process Quality*. New York: John Wiley, 267-285.

Couper, M.P., Horm, J., and Schlegel, J. (1997), "Using Trace Files to Evaluate the National Health Interview Survey CAPI Instrument." *Proceedings of the Section on Survey Research Methods*, American Statistical Association. Alexandria: ASA, pp. 825-829.

Couper, M.P., and Lyberg, L.E. (2005), "The Use of Paradata in Survey Research." In *Proceedings of the 55th Session of the International Statistical Institute*, Sydney, Australia, April [CD].

Couper, M.P., and Peterson, G. (2016), "Why Do Web Surveys Take Longer on Smartphones?" *Social Science Computer Review*, available online, DOI: 10.1177/0894439316629932.

Couper, M.P., and Rowe, B. (1996), "Evaluation of a Computer-Assisted Self-Interview Component in a Computer-Assisted Personal Interview Survey." *Public Opinion Quarterly*, 60: 89-105.

Couper, M.P., Sadosky, S., and Hansen, S.E. (1994), "Measuring Interviewer Behavior Using CAPI." *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Alexandria: ASA, pp. 845-850.

Couper, M.P., and Singer, E. (2013), "Informed Consent for Web Paradata Use." *Survey Research Methods*, 7 (1): 57-67.

Couper, M.P., and Wagner, J. (2011), "Using Paradata and Responsive Design to Manage Survey Nonresponse." *Proceedings of the 58th Congress of the International Statistical Institute*, Dublin, Ireland, August.

Diedenhofen, B., and Musch, J. (2014), "PageFocus: A New JavaScript to Detect Cheating in Online Tests." Paper presented at the General Online Research Conference, Cologne, March.

Durrant, G.B. (2016), "Using Quantitative Paradata to Improve the Data Quality in Social Surveys." In R. Edwards, A. Phoenix, H. O'Connor, and J. Goodwin (eds.), *Working With Paradata, Marginalia and Fieldnotes: The Centrality of By-Products of Social Research*. London: Edward Elgar.

Durrant, G.B., and D'Arrigo, J. (2014), "Doorstep Interactions and Interviewer Effects on the Process Leading to Cooperation or Refusal." *Sociological Methods and Research*, 43 (3): 490-518.

Durrant, G.B., D'Arrigo, J., and Steele, F. (2013), "Analysing Interviewer Call Record Data by Using a Multilevel Discrete-Time Event History Modelling Approach." *Journal of the Royal Statistical Society, Series A*, 176 (1): 251-269.

Durrant, G.B., Maslovskaya, O., and Smith, P.W. (2015), "Modelling Final Outcome and Length of Call Sequence to Improve Efficiency in Interviewer Call Scheduling." *Journal of Survey Statistics and Methodology*, 3 (3): 397-424.

Durrant, G.B., and Steele, F., (2009), "Multilevel Modelling of Refusal and Non-Contact in Household Surveys: Evidence from Six UK Government Surveys." *Journal of the Royal Statistical Society, Series A*, 172 (2): 361-381.

Eckman, S., Sinibaldi, J., and Montmann-Hertz, A. (2013), "Can Interviewers Effectively Rate the Likelihood of Cases to Cooperate?" *Public Opinion Quarterly*, 77 (2): 561-573.

Glasgow, R.E., Nelson, C.C., Kearney, K.A., Reid, R., Ritzwoller, D.P., Strecher, V.J., Couper, M.P., Green, B., and Wildenhaus, K. (2007), "Reach, Engagement, and Retention in an Internet-Based Weight Loss Program in Three Healthcare Systems." *Journal of Medical Internet Research*, 9 (2): e11.

Greenberg, B.S., and Stokes, S.L. (1990), "Developing an Optimal Call Scheduling Strategy for a Telephone Survey." *Journal of Official Statistics*, 6 (4): 421-435.

Groves, R.M., Benson, G., Mosher, W.D., et al. (2005), "Plan and Operation of Cycle 6 of the National

Survey of Family Growth." National Center for Health Statistics. Vital and Health Statistics 1 (42), available at: http://www.cdc.gov/nchs/data/series/sr_01/sr01_042.pdf

Groves, R., and Heeringa, S. (2006), "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society, Series A*, 169 (3), 439-457.

Heerwegh, D. (2003), "Explaining Response Latencies and Changing Answers Using Client-Side Paradata from a Web Survey." *Social Science Computer Review*, 21 (3): 360-373.

Heerwegh, D. (2008), "Paradata." In P.J. Lavrakas (ed.), *Encyclopedia of Survey Research Methods*, *Volume 2*. Los Angeles: Sage, pp. 574-575.

Heerwegh, D. (2011), "Internet Survey Paradata." In M. Das, P. Ester, and L. Kaczmirek (eds.), *Social Research and the Internet*. New York: Taylor and Francis, pp. 325-348.

Hubbard, F., Couper, M.P., Wagner, J., Gu, H., Benson, G., and Schulz, P. (2010), "Creating Effective Paradata Dashboards to Help Reduce Survey Costs and Errors." Paper presented at FedCASIC, Washington, D.C., March.

Jeavons, A. (2001), "Paradata: Concepts and Applications." *Proceedings of the ESOMAR Net Effects 4 Internet Conference*, Barcelona, Spain, February. Amsterdam: ESOMAR, pp. 177-187.

Kaczmirek, L. (2009), *Human-Survey Interaction: Usability and Nonresponse in Online Surveys*. Cologne: Herbert von Halem Verlag.

Kilburn, D., and Earley, J. (2015), "Disqus Website-based Commenting as an E-research Method: Engaging Doctoral and Early-career Academic Learners in Educational Research." *International Journal of Research and Method in Education*, 38 (3): 288-303.

Krueger, B.S., and West, B.T. (2014), "Assessing the Potential of Paradata and Other Auxiliary Information for Nonresponse Adjustments." *Public Opinion Quarterly*, 78 (4): 795-831.

Kreuter, F. (ed.)(2013), *Improving Surveys with Paradata: Analytic Uses of Process Information*. New York: Wiley.

Kreuter, F., Couper, M.P., and Lyberg, L.E. (2010), "The Use of Paradata to Monitor and Manage Survey Data Collection." Invited paper presented at the Joint Statistical Meetings, Vancouver, August.

Olson, K. (2013), "Paradata for Nonresponse Adjustment." *Annals of the American Academy of Political and Social Science*, 645 (1): 142-170.

Resnicow, K., Strecher, V.J., Couper, M.P., Chua, H.F., Little, R.J.A., Nair, V., Polk, T., and Atienza, A.A. (2010), "Methodological and Design Issues in Patient-Centered eHealth Research." *American Journal of Preventive Medicine*, 38 (1): 98-102.

Scheuren, F. (2001), "Macro and Micro Paradata for Survey Assessment." Washington, D.C.: Urban Institute, unpublished paper.

Schouten, B., Calinescu, M., and Luiten, A. (2013), "Optimizing Quality of Response Through Adaptive Survey Designs." *Survey Methodology*, 39 (1): 29-58.

Sendelbah, A., Vehovar, V., Slavec, A., and Petrovčič, A. (2016), "Investigating Respondent Multitasking in Web Surveys Using Paradata." *Computers in Human Behavior*, 55 (Part B): 777-787.

Singer, E., and Couper, M.P. (2011), "Ethical Considerations in Web Surveys." In M. Das, P. Ester, and L. Kaczmirek (eds.), *Social Research and the Internet*. New York: Taylor and Francis, pp. 133-162.

Sinibaldi, J., Durrant, G.B., and Kreuter, F. (2013), "Evaluating the Measurement Error of Interviewer Observed Paradata." *Public Opinion Quarterly*, 77 (1): 173-193.

Steele, F., and Durrant, G.B. (2011). "Alternative Approaches to Multilevel Modelling of Survey Non-Contact and Refusal." *International Statistical Review*, 79 (1): 70-91.

Strecher, V.J., McClure, J., Alexander, G., Chakraborty, B., Nair, V., Konkel, J., Greene, S.M., Couper, M.P., Carlier, C., Wiese, C., Little, R.J.A., Pomerleau, C., and Pomerleau, O. (2008), "The Role of Engagement in a Tailored Web-Based Smoking Cessation Program: Randomized Controlled Trial." *Journal of Medical Internet Research*, 10 (5): e36.

Thalji, L., Hill, C.A., Mitchell, S., Suresh, R., Speizer, H., and Pratt, D. (2013). "The General Survey System Initiative at RTI International: An Integrated System for the Collection and Management of Survey Data." *Journal of Official Statistics*, 29 (1): 29-48.

Wagner, J., West, B. T., Kirgis, N., Lepkowski, J. M., Axinn, W. G., & Ndiaye, S. K. (2012), "Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection." *Journal of Official Statistics*, 28 (4): 477-499.

Weeks, M.F., Jones, B.L., Folsom, R.E. and Benrud, C.H. (1980), "Optimal Times to Contact Sample Households." *Public Opinion Quarterly*, 44 (1): 101-114.

West, B.T. (2013), “An Examination of the Quality and Utility of Interviewer Observations in the National Survey of Family Growth.” *Journal of the Royal Statistical Society, Series A*, 176(1): 211-225.

West, B.T., and Kreuter, F. (2013), “Factors Affecting the Accuracy of Interviewer Observations: Evidence from the National Survey of Family Growth.” *Public Opinion Quarterly*, 77 (2): 522-548.

West, B.T. and Kreuter, F. (2015), “A Practical Technique for Improving the Accuracy of Interviewer Observations of Respondent Characteristics.” *Field Methods*, 27(2): 144-162.

West, B.T. and Kreuter, F. (2016), “Strategies for Increasing the Accuracy of Interviewer Observations of Respondent Features: Evidence from the U.S. National Survey of Family Growth.” Draft paper under review.