ESRC National Centre for Research Methods Review paper

Survey Paradata: A review

Gerry Nicolaas
National Centre for Social Research (NatCen)

January 2011
National Centre for Research Methods
NCRM/017

Content

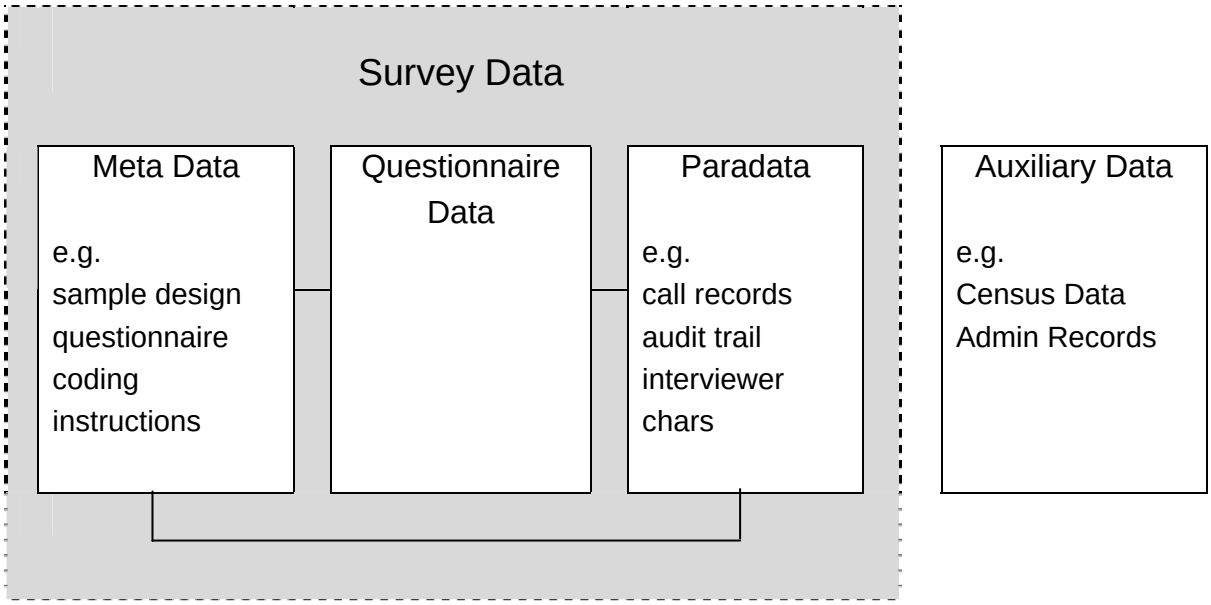
1 What is paradata?

There is currently no consensus over a standard definition for paradata. Couper (1998) was the first person to introduce the term “paradata” to the field of survey methodology. He initially used the term to refer to automatically generated process data, such as the audit trails produced by Blaise. Since then the term has expanded to cover all types of data about the process of collecting survey data such as interviewer call records, length of interview, keystroke data, interviewer characteristics. Although interviewer observations and information from interviewer questionnaires do not describe processes, this kind of information is also often referred to as paradata. Not included under the term paradata are the actual survey questionnaire data.[1] For the purpose of this review we also do not include metadata and auxiliary data under the term paradata.

Types of data:

1. Survey questionnaire data
2. Survey metadata = data about data; e.g. sample design, questionnaire
3. Survey paradata = data about processes; e.g. call records, audit trails
4. Auxiliary data = external data; e.g. information on sampling frames, Census area characteristics, administrative data

Figure 1.1 Diagram showing the three types of survey data and auxiliary data



[1] Exceptions are questionnaire items that collect information about the process; e.g whether the respondent consulted their documents when answering survey questions.

2 Why are we interested in Paradata?

2.1 Introduction

There is growing awareness among data producers and data users of the potential benefits of using paradata to tackle some of the challenges that we are currently facing in this field such as declining response rates, increasing risk of non-response bias and measurement error, and escalating costs of survey data collection. The collection of survey paradata is not new but the range and detail of paradata being collected has increased substantially over the years. This is mainly due to the increasing computerisation of the entire process of collecting survey data, not just the collection of questionnaire data.

2.2 Methodological Research

Paradata can be used by survey methodologists in their pursuit of reliable, replicable and generalisable knowledge about survey methods and practice. The overall aim is to use that knowledge to minimise survey error.

To date, methodological research using paradata has mainly focussed on non-response error. For example, there is a long history of analysing call records to gain knowledge about optimal calling patterns in face-to-face surveys (Swires-Hennessy and Drake,1992; Campanelli et al, 1997; Durrant et al. 2009). Survey methodologists have also recognised the value of using paradata to study non-response bias; in particular paradata items which are available for all sampled cases and are correlated with both the likelihood of participation and the survey variables of interest.

Survey methodologists are also using paradata to explore measurement error (Bassilli and Scott, 1996; Caspar and Couper, 1997; Heerwegh, 2003; Draisma and Dijkstra, 2004; McGee, 2007; Yan and Tourangeau, 2008). For example, question timings, keystroke files and audio-recordings can provide indication of respondent difficulty in answering questions. Paradata can also being used to examine the effects of interviewer characteristics on measurement.

In addition to non-response error and measurement error, there may also be potential for using paradata to assess other sources of error such as processing error. For example, coding quality could be assessed using indicators of when coders consult code books or other material, indicators of when coders change their mind, and coding latencies.

Although the use of paradata holds much promise, more methodological research is required to identify the key paradata items to be collected and the best ways to use them.

2.3 Management of Data Collection

Over the years surveys have become more challenging. Surveys have become more complex due to declining willingness among the public to take part in surveys, a growing interest among survey sponsors in hard-to-get groups, an increasing demand for longitudinal studies, and a push towards the use of multiple modes of data collection. The design and costing of such complex surveys tend to be based on assumptions that carry great risk to costs and data quality. Dependent on the appropriate tools being available, paradata could be used to improve the control of data collection costs and the quality of its survey estimates.

Paradata can be used to improve the management of data collection in the following ways:

- To achieve a better understanding of the data collection processes and to identify opportunities to improve survey operations;
- To evaluate new data collection initiatives;
- To monitor data collection and identify problems in a timely fashion (e.g. problems with the data collection instruments, problems with interviewer performance, possible interviewer falsification);
- To produce quality control metrics for internal use but also for sponsors;
- To produce more accurate information about costs and data quality;
- To inform survey design decisions about trade-offs between fieldwork costs, data quality and time;
- To make changes to the survey design during data collection in order to optimise the trade-off between costs, quality and time (i.e. responsive design).

2.4 Adjusting for Survey Error

As well as minimising survey error through a better understanding of the causes of survey error and the improved management of data collection, paradata can also be used after data collection has been completed to adjust for error in the survey estimates.

Data from most large-scale surveys are weighted for non-response bias. On the whole, these weights are based on information from auxiliary data sources (e.g. sampling frame information, Census data) and tend to only adjust for differences in demographic characteristics between the population of interest and the responding sample. However, if these demographic characteristics are unrelated to the survey variables of interest and the propensity to respond then weighting will not reduce non-response bias.

Recognising this limitation, attention has now shifted towards the use of paradata to improve non-response adjustment. There is some evidence to suggest that paradata can add explanatory power to non-response adjustments (Blom, 2009). But the challenge currently facing methodologists is to identify paradata items (either existing paradata or new paradata) that are correlated with both the likelihood of participation and the survey variables of interest. This may require a range of paradata items that could be used for

bespoke non-response adjustments rather than a single adjustment strategy for all estimates.

The use of paradata to adjust for measurement error lags behind that of paradata for non-response adjustment. Nonetheless there is some evidence showing that item-level paradata can provide useful information about the question asking and answering process. Such item-level information is lacking in current strategies for handling measurement error. The hope is that item-level paradata can be used to produce measurement error adjustments as well as being used in statistical models for data analysis. To achieve this, further research is required to attain a better understanding of the relationship between paradata and measurement error.

3 Overview of Paradata

3.1 Introduction

In this section we focus on the range of paradata items that could be collected for computer-assisted personal interview (CAPI) surveys. Most of these paradata items have their equivalents in other modes, but some items are specific to the mode (e.g. interviewer observations of the sampled unit is specific to CAPI where interviewers visit the sampled unit). Appendix A provides an overview of which paradata items are available in which mode.

3.2 Interviewer characteristics

In interviewer-administered surveys, the interviewer plays a crucial role in maximising response to the survey and minimising measurement error. The extent to which interviewers are successful in these tasks tends to vary across interviewers. The analysis of interviewer characteristics contributes to a better understanding of the causes of this variation among interviewers, which can be used to improve the recruitment, training and supervision of interviewers.

The range of interviewer characteristics that is routinely available tends to be quite limited. Most data collection agencies will have administrative records that include the age and sex of the interviewer. More recently, data collection agencies are required to record the ethnicity of interviewers to comply with equality and human rights legislation. Usually there is also some information available about the level of experience; e.g. number of years working for the agency, pay grade.

There is growing interest in collecting additional information about interviewers which are not available in administrative records. For example, research has demonstrated that response rates are strongly correlated with positive interviewer attitudes towards persuading reluctant respondents (Groves and Couper, 1998; Lehtonen, 1996). Survey methodologists have also shown interest in other interviewer characteristics such as personality traits and social skills (Sinibaldi et al, 2009). Because this information is not routinely recorded, a separate data collection exercise among interviewers is required. Nonetheless this kind of data is also referred to as survey paradata.

3.3 Call record data

It is considered best practice for data collection agencies to record detailed information about all the calls that are made to sampled cases. For face-to-face surveys this tends to include the day, time and outcome of every call made to the sampled unit. It is usually possible to identify which calls have been made by the same or different interviewer.

Call records are available for all issued cases, i.e. non-respondents as well as respondents. This makes call record data particularly useful for non-response analyses and non-response adjustments. The data can be used to analyse call patterns and identify

successful calling strategies. The data are also useful for monitoring interviewer performance, either post fieldwork with the aim of improving future interviewer performance or in real-time with the potential of intervening during data collection. Furthermore, call records can be used to inform decisions for two-phase sample designs and other types of responsive survey design.

The usefulness of call record data depends on the accuracy of the data being recorded. Biemer et al (2010) demonstrated that even small errors in the number and details of calls can seriously bias estimates. Although some error in call records is unavoidable, the process of collecting call records by field interviewers and how their performance is monitored can dissuade them from recording all call attempts. Wang and Biemer (2010) identified the following possible reasons for the under-reporting of calls:

- Underreporting may occur because of pressures to keep a case "alive." If too many unproductive visits are recorded, the case may get closed out.
- Field staff wishing to avoid being perceived as not using time effectively may also underreport.
- Failure to report "drive-by" visits seems to be a primary cause of underreporting.
- Another frequent cause of underreporting is the failure to report attempts to interview multiple sample persons within the same household.

The accuracy of call record data can to some extent be improved through better training and supervision of interviewers. Furthermore, new technologies such as hand-held devices and GPS can be used to reduce the burden of collecting call record information thus further enhancing the quality of the call record data.

3.4 Interviewer observations about the area and dwelling

Many surveys tend to use sampling frames with little if any additional information about the sampled cases. For example, the Postcode Address File is the most common sampling frame in the UK for surveys of the general population and it only includes address details. Consequently nothing is known about non-responding cases and to what extent the exclusion of these cases may increase the risk of bias. It has therefore become common practice for data collection agencies to instruct their field interviewers to record a selection of observations for all sampled cases, including the non-responding cases. For example, NatCen field interviewers are required to record four standard observations about the sampled address: i.e. the condition of the area, the condition of the sampled dwelling relative to the area, the type of dwelling and any physical barriers to gaining access to the sampled dwelling. Additional interviewer observations are also collected on specific surveys such as the Understanding Society study and the European Social Survey; e.g. evidence of litter or rubbish, evidence of vandalism, graffiti or deliberate damage to property, evidence of children living at the sampled address. Observations that are available for all sampled cases tend to relate to the area and the dwelling. It is also possible for the interviewer to collect observations about household members but such information is only available for contacted cases; e.g. age and gender of contacted

household member. Given the high contact rate for general population surveys in the UK, such observations can still be useful for assessing the risk of non-response bias.

The availability of observations for all sampled cases provides the opportunity to explore whether the non-responding cases are significantly different from the responding cases. This information can then be used to inform survey design issues (e.g. whether a non-responding case should be reissued for refusal conversion), to assess the risk of non-response bias, and to adjust the survey data for non-response bias. For this to be effective, the observations should be strongly associated with both the likelihood of the sampled case to respond and the survey variables of interest.

This need for information that is related to the survey variables of interest as well as the likelihood of responding suggests that survey-specific observations may be more useful than standard observations across all surveys. For example, observations about car ownership may be more informative about non-response bias on a travel survey than the condition of the area. Furthermore, the information provided by interviewer observations of the area could possibly be captured using other data sources such as small area statistics derived from the Census.

As well as identifying observations that are closely related to response outcome and survey variables of interest, it is also important to measure those observations well. Interviewer observations vary in their objectivity, ranging from easily identifiable observations to subjective speculations. Therefore the accuracy and reliability of the data can vary widely and have varying degrees of influence on the analyses in which they are used. Sinibaldi et al (2010) presented the results from five studies that looked at the measurement error of different types of interviewer observations and demonstrated that the measurement error can be large and that the causes of the error vary by the type of observation.

3.5 Doorstep interaction

A crucial step in gaining co-operation from sample members is the interaction between the interviewer and the contacted sample member. Some attempts have been made to collect summary measures of this interaction. For example, reasons for refusal are collected on a small number of surveys such as Understanding Society, the European Social Survey, and the Family Resources Survey. Specific reasons can be grouped under general headings such as too busy, concerns about confidentiality or the personal nature of the questions, attitudes towards the survey, family pressures not to take part and other reasons. This information can be used by interviewers to tailor their next contact with the sample member. However, this information has limited use for non-response analyses because it is only available for those cases that refuse to take part in the survey.

Summary measures of the interaction between interviewer and household members would be more useful for the analysis of non-response if the measures were collected for all contacted cases. It would then be possible to distinguish between common characteristics of cases that eventually become non-interviews from those cases that

eventually take part in the study. This information can be used to predict which contacted cases will subsequently refuse to take part. It has also been suggested that information about these initial interactions could possibly provide some indication of the quality of the subsequent interviews (Couper, 1997; Campanelli, Sturgis, and Purdon, 1997; Stussman,Taylor, and Riddick, 2004).

Not many studies have collected such measures of interaction for all contacted cases (Morton-Williams and Young, 1987; Groves and Couper, 1996; Campanelli, Sturgis, and Purdon, 1997). It is more than likely that the burden on interviewers and extra cost for collecting this information deters survey sponsors and data collection agencies from doing this. Nonetheless, this information is currently being collected in the UK on a small number of surveys, such as Understanding Society. Interviewers who work on this study are instructed to record whether household members asked any questions about the purpose of the study, the length of the interview, whether this would be a one-off request, who would see their answers, what's in it for them or some other question. Understanding Society interviewers also provide their own assessment of the respondent's initial resistance to co-operation (i.e. none, soft, moderate, firm).

3.6 Audit trails

Most large scale surveys carried out in the UK use computer-assisted interviewing (CAI) software systems. These CAI software systems tend to generate vast amounts of paradata along with the recording of questionnaire data. For example, the times an interviewer enters and exits a field can be automatically captured as well as a range of other actions carried out within the field such as the interviewer clicking on help keys and adding comments or remarks. Some CAI software systems (such as Blaise) record all keystrokes made during the interview. These paradata are stored in files called audit trails in Blaise but they are also known as keystroke files or trace files, and journal files for web surveys.

These audit trails have received increasing interest from survey managers, survey methodologists and analysts because they provide a wide range of information on all cases that can be used to evaluate the survey instrument, interviewer performance and data quality. For example, above average use of the help key for a specific field may suggest that there is a problem with that question. An interviewer who is spending significantly less time in certain fields than other interviewers could indicate that the interviewer is not administering those questions properly and may require additional training. Audit trails can also be used to reconstruct completed questionnaires that have been lost in transmission or corrupted.

However, the processing and analysis of audit trails are not straightforward. Audit trails are very large semi-structured text files consisting of a single record for each respondent for whom a questionnaire has been opened. The records are of varying lengths depending on the questionnaire routing for that particular respondent and the actions carried out within the fields. The analysis of these datasets can be time consuming and messy. There is a need for tools and techniques to harness and manage these vast amounts of

paradata. Practical methods for making use of the paradata, training of managers in analysing the paradata and development of key process variables are also required (Couper and Lyberg, 2005).

3.7 Audio recordings

Advanced CAI software systems also provide the opportunity to collect sound recordings during the interview. This is referred to as computer-assisted recorded interviewing (CARI). With respondents' consent, recordings can be made for the whole interview, a section of the interview, or random sections of the interview. In telephone surveys, supervisors can also listen in on live interviews (i.e. silent monitoring).

CARI provides the opportunity for survey managers and survey methodologists to listen to the interaction between the interviewer and the respondent and derive information that could be used to evaluate the instrument, the interviewer and data quality; i.e.

- To assess whether a question is working as intended;
- To assess the interaction of the interviewer with the instrument and how this affects the overall quality of the data;
- To assess the interaction of the interviewer with the respondent and how this affects the overall quality of the data;
- To reduce interviewer burden; e.g. CARI can be used to record long open-ended responses rather than having the interviewer type in the responses;
- To supplement or replace other quality control efforts; e.g. direct observation of a percentage of field interviews by field supervisors is costly and can be affected by the presence of the supervisor whereas CARI is relatively cheap and is less obtrusive;
- To supplement or replace interviewer validation; e.g. usually a fixed percentage of respondents are re-contacted to verify that the interview had taken place and was carried out as intended whereas CARI can be used to validate interviews without additional burden to the respondents;
- To provide more detailed information than written responses for coding and editing;
- To be used as an interviewer training tool.

The use of CARI is not routinely used on surveys in the UK, mainly due to operational issues and cost. Audio files tend to be large and need to be deleted from interviewers' hard drive space. Furthermore, such files being transmitted by a large field force can also take up a considerable amount of server space. Finally the analysis of the recordings requires behaviour coding which can be lengthy and costly. So the feasibility of using CARI depends on finding practical solutions to the space problems (e.g. automated methods for deleting sound recordings from hard disks after transmission, deletion of audio recordings after a period time, storing only a sub-sample of audio-recordings) and demonstrating that CARI can deliver cost savings.

3.8 Other paradata items

A range of other paradata items can be collected within the data collection instrument. Data collected within the actual questionnaire are generally not considered to be paradata. Exceptions, however, are data items which describe the process of asking and answering questions. For example, whether the respondent consulted documents when answering detailed questions about costs and expenditure; how certain the respondent is about the accuracy of their answer; how certain the interviewer is that the respondent provided an accurate answer; other interviewer remarks. Such paradata can be used to assess the risk of measurement error at the item level.

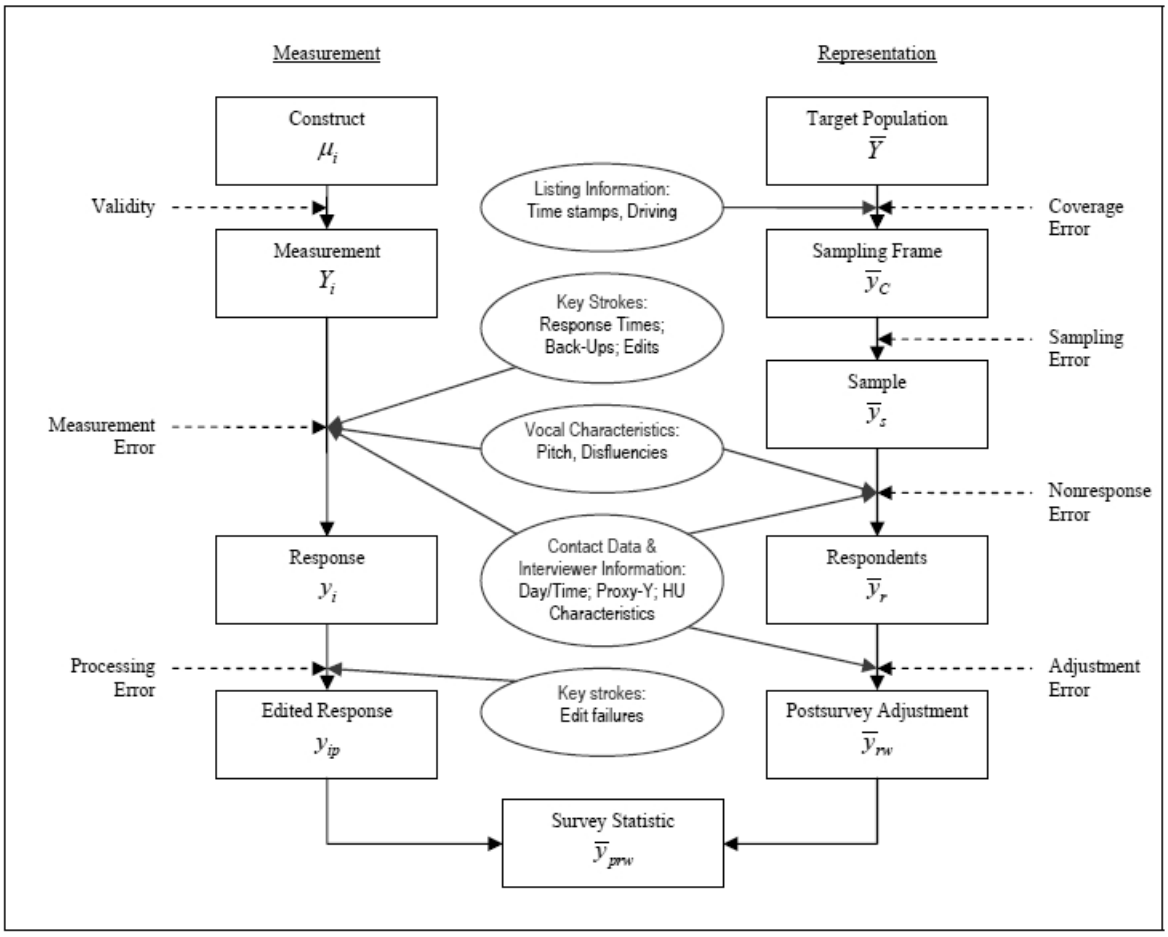
After the questionnaire has been completed, the interviewer may also have to answer a series of questions about the interviewing process, such as the perceived cooperativeness of the respondent. It is also common practice on many surveys for the interviewer to ask the respondent for their consent to be contacted again; either for quality control purposes or for possible future surveys. Other forms of consent include respondents' permission to have their survey data linked to administrative records. And, particularly on longitudinal studies, respondents are sometimes asked to provide their contact details (e.g. landline telephone number, mobile number, email address) as well as the contact details of a stable contact to aid tracing in the case of the respondent moving. The provision of consents and more detailed contact information can be seen as indicators of respondent cooperativeness, which is likely to be related to future requests to take part in follow-up surveys. Studies have shown that those who refuse to give consent for data linkage are more likely to have concerns about privacy and data confidentiality (Armstrong et al, 2008) and less likely to provide information on income and wealth (Jenkins et al, 2006; Olson, 1999; Woolf et al, 2000).

Finally, paradata items can also be collected after the data collection has been completed while the data are being processed; e.g. information about coding, editing, imputation.

3.9 In conclusion

All the paradata items described in this chapter have the potential for measuring survey error (see figure 2 for a diagram showing the relationship between a selection of paradata items and total survey error). Further methodological research is required to examine the relationship between these paradata items and different components of survey error, and identify the key paradata items to be collected and the best ways to use them.

Figure 3.1 Total Survey Error components and paradata for their assessment



Source: Kreuter and Casas-Cordero, 2010

4 Access and Dissemination

4.1 Introduction

This review has demonstrated the potential value of paradata to measure and improve data quality as well as save costs. Paradata are therefore of interest not only to data collection organisations but also survey clients, survey analysts and survey methodologists. However, paradata are not usually released with the publicly available survey data and access to paradata is problematic for a variety of reasons including the cost involved in preparing datasets, the need for advice on how to manage and interpret the paradata, concerns about interviewer and respondent confidentiality, and concerns about damaging the commercial interests of the data collection agency.

4.2 Complex data

The range of paradata described in chapter 3 tends to come from disparate operational systems. On the whole, these systems have been developed for survey management purposes and they are not necessarily suitable for producing useful data for survey clients, survey analysts and survey methodologists.

The format and structure of the datasets vary and can be quite complex. They can also be very large and messy, particularly when paradata from different systems are being combined. Consequently much time can be spent accessing and manipulating the paradata to produce datasets that are useful for the data users. This issue can be addressed to some extent with the development of warehouses which can be designed to integrate paradata from various sources into a single and consistent structure. Statistics Canada has developed a warehouse and other data collection agencies are known to be moving in this direction. However, such warehouses tend to be work in progress, initially incorporating a selection of key paradata items for operational decision making with the option of increasing the range of paradata items over time. Consequently it is likely that significant resources will continue to be required for the production of bespoke datasets, particularly for methodological research.

Resources are not only required for the production of datasets but also for the provision of information to provide insight into the data. Information about fieldwork practices, classifications and measures of accuracy are necessary to assist interpretation and appropriate utilisation of the results. Given that research into and use of paradata is still in its infancy, much of this information is not readily available.

It has been suggested that paradata are relatively low cost for researchers because the paradata are either automatically generated by-products of the computerised systems or are being routinely collected as part of the operational running of surveys. Indeed the costs of paradata are low in comparison with the collection of new data. Nonetheless, the costs associated with the production of bespoke datasets for clients, analysts and methodologists should not be under-estimated. These costs, however, are likely to

decrease over time with the development of paradata warehouses, identification of key paradata items, improved knowledge about the properties of paradata items and the production of metadata.

4.3 Confidentiality concerns

Social surveys in the UK rely on the voluntary cooperation of the public. Their cooperation is sought on the understanding that their personal information and identities will be protected. Ensuring actual and perceived confidentiality of respondents' data is therefore essentail, including the associated paradata.

Various paradata sources include information that could disclose respondents' identities; e.g. address details, interviewer remarks, audio recordings. Consequently such data cannot be released without thorough processing and the removal of any information that could possibly be used either on its own or in combination with other information to identify respondents. The removal of address details is relatively straightforward. However, finding and removing disclosive information in semi-structured text files (e.g. audit trails) and audio recordings, while maintaining the usability of these data sources for research is problematic and time consuming.

The linking of paradata from multiple sources to survey data, administrative records, and other sources of information creates an extra challenge for safeguarding the identity of respondents. Furthermore, the often required inclusion of geographic data increases the risk that individuals can be identified.

In addition to the protection of the identities of survey respondents, data collection agencies also have a responsibility to protect the identities of their interviewers.

A number of techniques exist to control the risk of statistical disclosure which should be considered when releasing paradata to users; such as

- Rounding of values;
- Recoding of categories;
- Adding noise to the data;
- Micro aggregation of the paradata: the paradata dataset is divided into small groups and average values are computed;
- Data rank swapping: after ranking data in an ascending order, each value is randomly replaced by a close neighbour.

In addition to restricting the paradata being released, it is also possible to provide users access to the paradata in a data enclave. A data enclave is a supervised data-use facility where users are closely monitored to prevent unauthorised use or copying of the paradata.

4.4 Commercial sensitivity

Survey data collection agencies operate in a competitive market. Consequently these agencies tend to be cautious about the release of information that may prejudice their commercial interests. Paradata are particularly sensitive in this respect given that these data can be used to reveal details about their fieldwork processes.

However, there is no reason to believe that it is the intention of survey clients, analysts and methodologists to prejudice the commercial interests of data collection agencies. On the whole they are motivated to use paradata to pursue their aim of optimising data quality within cost constraints – an aim that is not at odds with the commercial interests of the agencies. Nonetheless, differences in perspective and culture may deter data collection agencies from sharing their paradata with academic researchers. Data collection agencies may also be reluctant to share paradata with survey clients for fear of attempts to micromanage data collection without a thorough understanding of the practice of collecting survey data.

This situation could be improved through collaboration between data collection agencies and those wishing to use paradata. Shared objectives (e.g. optimising data quality within cost constraints) are more likely to become evident when collaborating. And those wanting to use the paradata will become more aware of the commercial sensitivities involved in the release of paradata. Data collection agencies will gain access to specialist skills and resources that may not exist within the organisation or which may not be cost effective to develop in-house. And finally, collaboration will ease access to commercially sensitive data on the understanding that all parties, including the data collection agency, will be involved in the research and the dissemination of results.

4.5 In conclusion

To sum up, the access and dissemination of paradata for bona fide research can be improved by:

- The development of paradata warehouses with metadata;
- The inclusion of money in research budgets for data collection agencies to produce bespoke paradata sets and to provide advice on how to analyse and interpret the results;
- The use of statistical disclosure control techniques and/or the use of data enclaves to safeguard the identities of respondents and interviewers;
- Closer collaboration between data producers and data users to fully exploit the potential of using paradata to manage data quality and costs.

5 Conclusion

This review has demonstrated the potential value of paradata to obtain a clearer understanding of the causes of survey error, to improve the design and management of data collection so that survey error is minimised within cost constraints, and to improve statistical adjustments for survey error. However, there are barriers that need to be overcome before we can exploit the paradata for these and other purposes.

The need for tools and techniques to harness and manage the vast amounts of paradata that could be made available has been highlighted by Couper (1998). Data collection agencies have introduced computerised systems that link the various stages of the survey process, thus capturing large quantities of paradata on a continuous basis. But the format and structure of the paradata can be quite complex and messy. A number of data collection agencies are now following the lead of Statistics Canada to develop paradata warehouses to make it easier to use their paradata by integrating the paradata from various sources into a single and consistent structure.

The use of paradata is still in its infancy and it is still unclear which paradata items are key to collect and the best ways to use them. Also, little is known about the quality of paradata and how this may affect the utility of the data. Further methodological research is required to examine the relationship between paradata and the different components of survey error.

Given the complex structure of paradata, complex models are likely to be required to fully exploit the richness of the data. Reporting back from an expert workshop, Lynn and Nicolaas (2010) suggest that techniques such as data mining, regression trees and random forests may have something to offer. Also, parallel modelling of a substantive process of interest and the data collection process could be investigated.

Finally, access to paradata is problematic for a variety of reasons including the cost involved in preparing datasets, the need for advice on how to manage and interpret the paradata, concerns about interviewer and respondent confidentiality, and concerns about damaging the commercial interests of the data collection agency. Collaboration between data producers and data users should help in overcoming these barriers.

References

Armstrong, V., Barnett, J., Cooper, H., Monkman, M., Moran-Ellis, J. And Shepherd, R. (2008) "Public Perspectives on the Governance of Biomedical Research: A qualitative study in a deliberative context", edited by Wellcome Trust: University of Surrey.

Bassili, J.N., and Scott, B.S. (1996), "Response Latency as a Signal to Question Problems in Survey Research." Public Opinion Quarterly, 60: 390–399.

Biemer, P.P., Chen, P. And Wang, K. (2010), "Using Level of Effort Paradata in Nonresponse Adjustments with Application to Field Surveys", Paper presented at "Recent Advances in the Use of Paradata in Social Survey Research", a one-day research symposium on 9 December 2010 at the Royal Statistical Society.

Blom, A. (2009) "Nonresponse Adjustments in the European Social Survey: The Added Value of Process Weights", Paper presented at "The Application of Paradata in Social Surveys", a one-day seminar on Tuesday 25 August 2009, London School of Economics

Campanelli, P./Sturgis, P. and Purdon, S. (1997): Can you hear me knocking: An investigation into the impact of interviewers on survey response rates. Technical report, The Survey Methods Centre at SCPR, London.

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. (1997). Survey Introductions and Data Quality. Public Opinion Quarterly, 61, 317–338.

Couper, M. (1998): Measuring survey quality in a CASIC environment. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Couper, M. P., & Lyberg, L. E. (2005). The use of paradata in survey research. Paper presented at the 54th Session of the International Statistical Institute, Sydney, Australia.

Draisma, S., and Dijkstra, W. (2004), "Response Latency and (Para)Linguistic Expressions as Indicators of Response Error." In S. Presser, J. Rogthgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer (Eds.), Methods for Testing and Evaluating Survey Questionnaires. Hoboken, NJ: John Wiley and Sons, pp. 131-147.

Durrant, G., D'Arrigo, J.and Steele, F. (2009): Using field process data to predict best times of contact conditioning on household and interviewer influences. Technical report, Southampton Statistical Science Research Institute.

Groves, R.M. and Couper, M.P. (1996). Contact-Level Influences on Cooperation in Faceto-
Face Surveys. Journal of Official Statistics, 12, 63–83.

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.

Jenkins, S.P., Cappellari, L., Lynn, P., Jäckle, A. and Sala, E. (2006) "Patterns of consent: evidence from a general household survey." Journal of the Royal Statistical Society, Series A 169: 701-722.

Lehtonen, R (1996). Interviewer attitudes and unit nonresponse in two different interview schemes. In. S. Laaksonen (eds). International perspectives on nonresponse: Proceedings of the sixth international workshop on household survey nonresponse. Helsinki: Statistics Finland

Lynn, P. and Nicolaas, G. (2010) “Making Good Use of Survey Paradata”. Survey Practice, April: www.surveypractice.org.

Kreuter, F. and Casas-Cordero, C. (2010) “Paradata”, Working Paper Series of the German Council for Social and Economic Data (RatSWD), Working Paper No. 136.

McGee, Alice (2007). CARI on NatCen: Using a combination of Computer Assisted Recorded Interviewing (CARI) and Behavior Coding to measure data quality in the English Longitudinal Study of Ageing (ELSA). NatCen Newsletter, Vol 25: Spring 2007.

Morton-Williams, J. and Young, P. (1987). Obtaining the Survey Interview – An Analysis of Tape Recorded Doorstep Introductions. Journal of the Market Research Society, 29, 35–54.

Olson, J.A. (1999) "Linkages With Data From Social Security Administrative Records in the Health and Retirement Study." Social Security Bulletin 62.

Scheuren, F. (2005): Paradata from concept to completion. In: Proceedings of the Statistics Canada Symposium. Methodological Challenges for Future Information Needs.

Sinibaldi, J., Jäckle, A., Tipping, S. and Lynn, P. (2009) “Interviewer Characteristics, their Doorstep Behaviour, and Survey Co-operation”, Paper presented at the American Association of Public Opinion Research Conference, May 14-17, 2009.

Sinibaldi, J., Eckman, S., West, B., Casas-Cordero, C., McCulloch, S.K. and Kreuter, F. (2010)
“Assessment of the Quality of Interviewer Observations”, Paper presented at “Recent Advances in the Use of Paradata in Social Survey Research”, a one-day research symposium on 9 December 2010 at the Royal Statistical Society.

Stussman, B.J., Taylor, B.L., and Riddick, H. (2004). Partials and Break-offs in the National Health Interview Survey, 2002. Proceedings of the 2003 Federal Committee on Statistical Methodology Conference, 21–25.

Swires-Hennessy, E. and Drake, M. (1992). The Optimum Time at which to Conduct Interviews. Journal of the Market Research Society 34, 61-72.

Wang, K. and Biemer, P. (2010). The accuracy of interview paradata: results from a field study, Proceedings of the American Association of Public Opinion Research Annual Conference, Chicago, IL.

Woolf, S.H., Rothemich, S.F., Johnson, R.E. and Marsland, D.W. (2000) "Selection Bias From Requiring Patients to Give Consent to Examine Data for Health Services Research." Archives of Family Medicine 9: 1111-1118.

Yan, T., and Tourangeau, R. (2008), “Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times.” Applied Cognitive Psychology, 22 (1): 51-68.

Appendix A Overview of Paradata

	Modes	Available for all sampled cases	Available for all contacted cases	Available for all refusals	Available for all productive cases
Call record data	CAPI, CATI	✓	✓	✓	✓
Interviewer observations	CAPI, CATI	✓	✓	✓	✓
Interviewer characteristics	CAPI, CATI	✓	✓	✓	✓
Doorstep interaction	CAPI		✓	✓	✓
Reasons for refusal	CAPI, CATI			✓	
Audit files	CAPI, CATI				✓
Journal files	Web				✓
Audio Recordings	CAPI, CATI		✓	✓	✓
Silent Monitoring	CATI		✓	✓	✓
Respondent's assessment of data quality	CAPI, CATI, web				✓
Interviewer's assessment of data quality	CAPI, CATI				✓
Respondent consents	CAPI, CATI, web				✓
Respondent contact information	CAPI, CATI, web				✓