Interviewer-Observed Paradata in Mixed-Mode and Innovative Data Collection

Kunz, Tanja; Daikeler, Jessica; Ackermann-Piek, Daniela

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Mitglied der

Leibniz-Gemeinschaft

gesis
Leibniz-Institut
für Sozialwissenschaften

Research Note

Interviewer-Observed Paradata in Mixed-Mode and Innovative Data Collection

Tanja Kunz ⬤ and **Jessica Daikeler** ⬤
GESIS – Leibniz-Institute for the Social Sciences, Germany

Daniela Ackermann-Piek
SRH Fernhochschule – The Mobile University, Germany

Abstract

In this research note, we address the potentials of using interviewer-observed paradata, typically collected during face-to-face-only interviews, in mixed-mode and innovative data collection methods that involve an interviewer at some stage (e.g., during the initial contact or during the interview). To this end, we first provide a systematic overview of the types and purposes of the interviewer-observed paradata most commonly collected in face-to-face interviews—contact form data, interviewer observations, and interviewer evaluations—using the methodology of evidence mapping. Based on selected studies, we illustrate the main purposes of interviewer-observed paradata we identified—including fieldwork management, propensity modeling, nonresponse bias analysis, substantive analysis, and survey data quality assessment. Based on this, we discuss the possible use of interviewer-observed paradata in mixed-mode and innovative data collection methods. We conclude with thoughts on new types of interviewer-observed paradata and the potential of combining paradata from different survey modes.

Keywords

evidence map, contact history information, interviewer observations, interviewer evaluations, face-to-face interview, CAPI-plus, video interview, knock-to-nudge

Introduction

Face-to-face interviewing has long been considered the 'gold standard' among data collection methods in the market and social research, which is mainly due to the long-time higher response rates, the better reaching of hard-to-reach target groups, and thus the less biased and more

Corresponding author:
Tanja Kunz, GESIS – Leibniz-Institute for the Social Sciences, B6, 4-5, Mannheim 68159, Germany.
Email: tanja.kunz@gesis.org

representative survey data (Schober, 2018; Villar & Fitzgerald, 2017). Another advantage of face-to-face interviewing is the unique opportunity for the interviewer to collect additional data, so-called interviewer-observed paradata, about the respondents and nonrespondents, their living environment, and the interview situation itself. These paradata allow researchers and practitioners to learn more about improving fieldwork processes and ensure a high quality of survey data (Groves & Heeringa, 2006; Kirchner et al., 2017; Kreuter, 2013).

In recent years, and further reinforced by the COVID-19 pandemic, there have been increasing calls in the market and social research to switch from face-to-face-only interviewing to mixed-mode designs (Luijkx et al., 2021; Wolf et al., 2021), or other innovative survey data collection methods (Conrad et al., 2022; Endres et al., 2022; Jeannis et al., 2013; Schober, 2018; West et al., 2022). While the idea of mixed-mode data collection and its benefits are not new (de Leeuw, 2005; 2018; Dillman, 2005; Scherpenzeel, 2017), they have become even more important in the post-pandemic era (Cleary et al., 2021; Kuenzi et al., 2022; Kantar Public, 2021). In addition, innovative methods that involve an interviewer in some way, such as knock-to-nudge contact strategies or remote video interviewing, gained prominence (Cornick et al., 2022; West et al., 2022).

Mixed-mode and innovative data collection methods allow for rapidly adapting fieldwork processes to changing conditions and more flexible responses to unforeseen events (Cornick et al., 2022; SHARE-ERIC, 2022). Moreover, they enable the collection of rich interviewer-observed paradata at each step interviewers are actively involved. Even though many survey researchers and practitioners are already familiar with the common interviewer-observed paradata from face-to-face-only interviews, little is known about the meaningful use of these paradata in mixed-mode and innovative data collection methods. Therefore, this research note provides a systematic overview of the most common types and purposes of interviewer-observed paradata in face-to-face-only interviews. Based on this, we discuss their potential uses in mixed-mode and innovative data collection methods and provide initial suggestions for academic research and practice.

Interviewer-Observed Paradata in Face-to-Face-Only Interviewing

We systematically searched for previous empirical studies dealing with interviewer-observed paradata in face-to-face interviews and compiled them using the evidence-mapping methodology (Saran & White, 2018; Snilstveit et al., 2013). We included 102 articles and coded the types and purposes of interviewer-observed paradata (details of the search, screening, and coding process in the Supplementary Appendix).

Figure 1 shows in the rows the main types of interviewer-observed paradata in face-to-face studies, namely contact form data, interviewer observations, and interviewer evaluations, and their subtypes (see Table A4 in the Supplementary Appendix for a complete list of paradata types coded in our studies, including examples). Columns list the five primary purposes of interviewer-observed paradata that we identified based on our studies, including fieldwork management, propensity modeling, nonresponse bias analysis, substantive analysis, and survey data quality assessment. The size of the circles corresponds to the frequency with which the paradata occurred as (in)dependent variables in the analyses of the studies. The paradata types most often used for specific purposes are highlighted in light gray and are briefly described below based on selected studies.
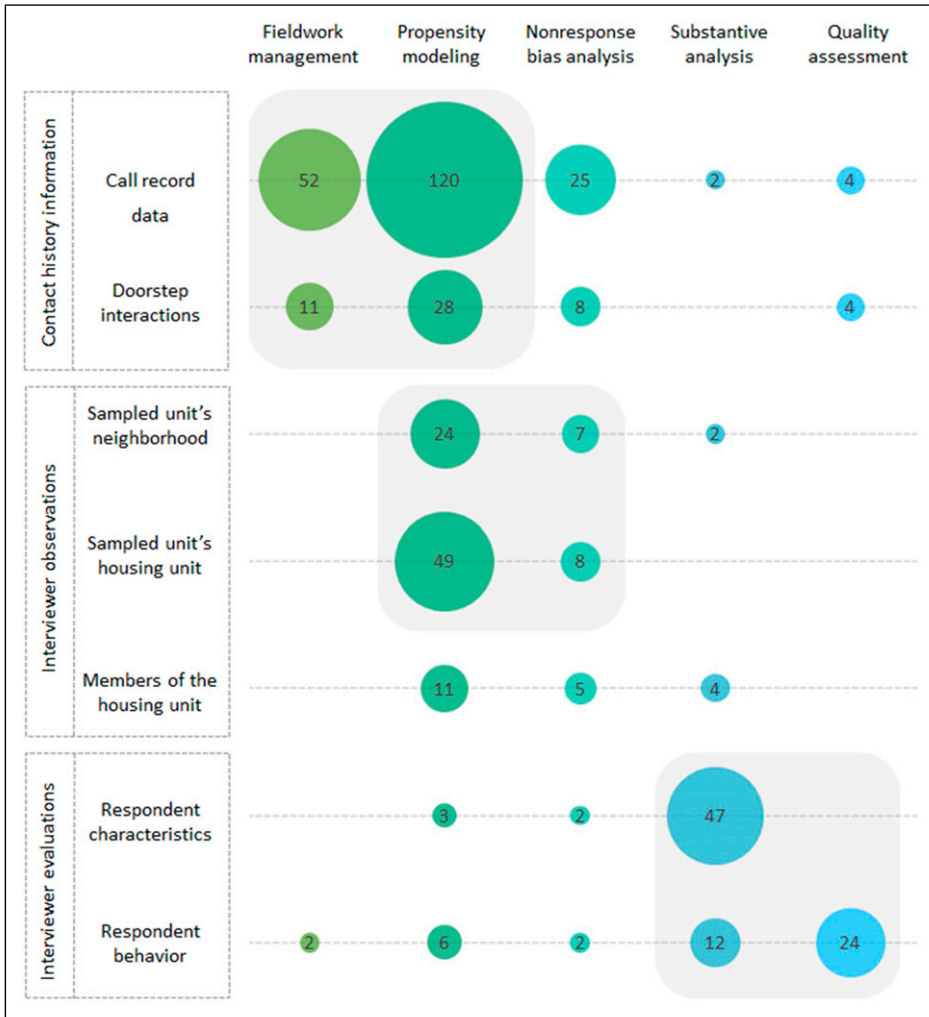
Figure 1. Evidence map on main types and purposes of interviewer-observed paradata in face-to-face interviewing.

Types of Interviewer-Observed Paradata

Contact history information is collected during an interview's recruitment and contact phase; they include *call record data*, usually available for each contact attempt (Durrant et al., 2011). Examples are the time and mode of contact, the outcome of each contact attempt, or the reasons for noncontact. *Doorstep interactions* are available for each contact attempt that results in contact with a household member and include the household members' initial reactions and concerns about participating in the interview (Loosveldt & Joye, 2016). ***Interviewer observations*** are usually collected only once at the first contact attempt for all sample units, including noncontacts and refusals (Durrant et al., 2011). They include observations on the *neighborhood* (e.g., the condition of houses in the area) and the *housing unit of the sample unit* (e.g., physical barriers to accessing the house). Interviewer

observations also refer to *the members of a housing unit* (e.g., single or couple household); this information is collected only for successfully contacted households (Olson, 2013). **Interviewer evaluations** are usually recorded after the completion of the interview and are therefore only available for interviewed respondents (Kirchner et al., 2017); they relate to characteristics of the respondent (e.g., sociodemographic status) and characteristics of the interview situation (e.g., respondents' engagement during the interview).

Purposes of Interviewer-Observed Paradata

Fieldwork management generally aims to improve contact and cooperation during the field phase to increase response rates and sample representativeness while reducing survey costs. In particular, *call record data* are used in fieldwork management to optimize, tailor, or even initiate targeted solutions for making successful contact (e.g., prioritizing or stopping calls in case of unsuccessful call sequences, increasing the number of active interviewers in the field) (Durrant et al., 2019; Kennickell, 2017; Konicki & Adams, 2016; Purdon et al., 1999; Safir & Tan, 2009; Vandenplas et al., 2017; Zelenak & Davis, 2013). *Doorstep interactions* are used as part of the fieldwork management to gain insights into reasons for nonparticipation, such as lack of time (e.g., "too busy") or privacy concerns (e.g., "don't trust surveys") (Bates & Piani, 2005; Vercruyssen et al., 2011). This information can be used, among others, to adjust contact timing or guide interviewer training regarding refusal conversion.

Propensity modeling can provide helpful information for effective fieldwork management by predicting the likelihood of respondent contact and cooperation in a survey. Additional information about potential (non)respondents in advance or in the early stages of contacting can help determine the best timing and strategies for contact. Based on our studies, it is mostly *call record data* used for this purpose. Findings show, for example, that the likelihood of contact is highest for the first call and calls made in the evening or on weekends, while it decreases with the number of calls made previously (Blom, 2012; Durrant et al., 2011). Concerning the likelihood of cooperation, findings are more mixed; some studies indicate higher cooperation with later contact attempts (Durrant et al., 2013), while others show that more contact attempts mean lower cooperation (Groves & Heeringa, 2006; Kreuter & Kohler, 2009; West & Groves, 2013). *Doorstep interactions* reveal higher refusal rates among households that express concerns more frequently; also, refusal rates differ by type of concern (Bates et al., 2008; Bates & Piani, 2005; Vercruyssen et al., 2011; West & Groves, 2013). Identifying the types of concerns that cause interviewers major problems in obtaining cooperation can help develop appropriate interviewer training and other strategies to address these concerns successfully (e.g., transferring cases to more experienced interviewers). *Interviewer observations* are also commonly used for propensity modeling. Findings show that observations on the neighborhood and housing of sampled units help predict contact (e.g., contact is less likely in areas where the interviewer would feel unsafe after dark or for houses in poor condition) (Blom, 2012; Durrant et al., 2011; Durrant & Steele, 2009; Steele & Durrant, 2011), but that it seems to be very context-dependent, which observations are appropriate for predicting cooperation (Blom et al., 2011; Casas-Cordero, 2010; Durrant et al., 2013; Durrant et al., 2017; Groves & Heeringa, 2006; Krueger & West, 2014; Vercruyssen & Loosveldt, 2017; West, 2013; West & Groves, 2013).

Nonresponse bias analysis generally aims to assess the consequences of nonparticipation of sampled cases for survey estimates and to adjust for possible nonresponse bias due to systematic differences between respondents and nonrespondents (Groves, 2006). *Call record data* and *doorstep interactions* are helpful for nonresponse bias assessment. For example, respondents with

multiple contact attempts (hard-to-reach) or respondents who express initial concerns (hard-to-persuade) serve as proxies for true nonrespondents to examine the extent of bias (Boniface et al., 2017; Lee et al., 2018; Lynn & Clarke, 2002). These paradata are also used to examine the effect of increased fieldwork efforts (e.g., repeated contact attempts, refusal conversion) on reducing nonresponse bias (Lynn & Clarke, 2002; Moore et al., 2018). In contrast, call record data and doorstep interactions proved to be of little use in improving the quality of nonresponse adjustments (Biemer et al., 2013; Maitland et al., 2009; Peytchev & Olson, 2007; Wagner et al., 2014)—either due to low correlations between paradata-derived indicators and key survey variables (Hanly et al., 2016; Kreuter & Kohler, 2009; Peytchev & Olson, 2007; Wagner et al., 2014) or due to under-reporting of contact attempts by interviewers (Biemer et al., 2013). *Interviewer observations* about the housing unit and its members (e.g., type of household, presence of children, receipt of un-employment benefits) also do not have substantial utility in nonresponse adjustment, as they do not predict response outcomes and key survey variables well (Kreuter et al., 2010; West et al., 2014).

Substantive analyses that address content issues of all kinds can benefit from *interviewer evaluations* as proxy information about respondents, such as their socioeconomic status (Davis et al., 1999; Møller, 1992), health status (Haug & Folmar, 1986; Prigerson et al., 1997; Sakshaug et al., 2010), political knowledge (Ansolabehere et al., 2008; Gay, 2014; McCann & Lawson, 2006; Morris et al., 2003; Treier & Hillygus, 2009; Winter, 2010), or parent-child interaction (Ellis et al., 2003; Fryer & Levitt, 2013; Leventhal & Brooks-Gunn, 2005; Morris et al., 2003; Zaslow et al., 1995). However, the usefulness of interviewer evaluations as proxy information seems highly context-dependent. In some cases, interviewer evaluations are interchangeable with self-reports (e.g., language proficiency) (Tubergen & Kalmijn, 2005; Winter, 2010); in other cases, they do not measure the same thing as self-reports (e.g., neighborhood and home conditions) and thus should be used as supplemental information rather than a substitute (Ansolabehere et al., 2008; Lee & Waite, 2018).

Survey data quality assessment relies primarily on *interviewer evaluations* of respondents' behavior and their engagement and understanding during the interview to detect potential measurement problems and poor data quality (Bricker, 2014; Mellinger et al., 1982; Weissman et al., 1996; West et al., 2018). In this context, Perales and Baffour (2018) found that interviewer evaluations of respondent behavior during the interview are good predictors of data quality because they point in the same direction as results based on objective indicators of survey engagement (e.g., panel dropout and item nonresponse).

Interviewer-Observed Paradata in Mixed-Mode and Innovative Data Collection Methods

First, we briefly describe three data collection methods with interviewer participation that have gained prominence in the market and social research during the COVID-19 pandemic, including CAPI-plus, video interviewing, and knock-to-nudge (Cornick et al., 2022). Second, we discuss the use of interviewer-observed paradata for these three methods. These uses are anecdotal and do not claim to be exhaustive.

CAPI-plus, as a type of sequential mixed mode, means that computer-assisted personal in-terviewing (CAPI) is the default mode of data collection. If respondents decline to participate in a face-to-face interview, they are offered an alternative mode, often telephone interviews or self-administered web surveys. *Computer-assisted video interviewing (CAVI)* is a form of remote interviewing in which the interviewer and respondent communicate via video call. The video interview usually does not take place during the initial contact, but appointments are made for a

specific time slot (Schober et al., 2020). **Knock-to-nudge (KtN)** is a contact method in which face-to-face interviewers visit sampled households and ask respondents at the doorstep to participate in a non-face-to-face survey. An appointment is made for the survey, which is conducted later, usually by telephone, video interview, or mixed mode (Cornick et al., 2022; Kastberg & Siegler, 2022).

Challenges in Contact and Cooperation

A major objective of mixed-mode and innovative data collection methods is to improve contact and cooperation to increase response rates and sample representativeness. For example, offering an alternative non-face-to-face mode in CAPI-plus can make the survey attractive to those concerned about face-to-face interaction or who want to avoid an interviewer in their home (Cornick et al., 2022). Similarly, face-to-face recruitment for a non-face-to-face survey through KtN can increase response rates. However, it also affects the distribution of respondent characteristics (e.g., younger, unmarried, living in larger households and the most deprived areas), presumably due to different likelihoods of respondents being at home and responding to the interviewer's knock on the door (Kastberg & Siegler, 2022). In addition, KtN requires comprehensive call scheduling due to the postponement of the interview. Concerning CAVI, not all respondents have access to an Internet-enabled device with a camera and microphone. Even if the technical requirements are met, not all respondents are ready for and comfortable with a video interview. Like KtN, CAVI involves comprehensive scheduling (Endres et al., 2022; Schober et al., 2020). Respondents' varying ability and willingness to participate and the more complex call scheduling, particularly for CAVI and KtN, underscore the importance of tailored fieldwork management, propensity modeling, and nonresponse bias analysis.

In all three data collection methods presented, contact history information can be usefully applied to fieldwork management and propensity modeling to better understand the mechanisms of successful contact and cooperation and develop an effective call scheduling and recruitment strategy, ultimately increasing response rates and sample representativeness. For example, call record data help optimize contact timing and prioritize cases most difficult in CAPI-plus and KtN to reach at home or those most likely to refuse in face-to-face mode. When different modes are combined, call sequence outcomes can improve recruitment strategy by tailoring the timing of mode switching (e.g., after how many contact attempts in CAPI mode, it is advisable to switch to another mode) and the number and type of reminders (e.g., call reminders, postal reminders, or email follow-ups). We also encourage gathering interviewer observations on the sampled unit's neighborhood and housing unit in CAPI-plus and KtN during (initial) face-to-face contact. As they have proven helpful for propensity modeling in face-to-face-only studies, they are promising for deriving tailored treatments before or early in the field phase in CAPI-plus and KtN (e.g., assigning cases to the appropriate mode). In addition, we recommend paying particular attention to doorstep concerns. It is crucial to understand respondents' concerns and barriers to data collection methods that are new and unfamiliar to many respondents. KtN and CAVI may involve concerns other than those from face-to-face-only interviews (e.g., unwillingness to provide a phone number during KtN, inadequate technical equipment, or discomfort with using video in CAVI). Only when we know the specific concerns can appropriate strategies be developed to encourage respondents to participate (e.g., sending experienced interviewers specially trained in refusal conversion, conducting brief doorstep training on the use of video). In addition, contact history information and interviewer observations on all sampled cases, including nonrespondents, help assess the extent of nonresponse and the consequences of nonparticipation for sample composition and survey estimates. For example, interviewer observations of (non)respondents' sociodemographic characteristics (e.g., age,

ethnicity, language spoken) or household type and composition (e.g., single-person household, presence of children) may explain why some respondents are more likely to refuse in CAVI than others or to prefer one mode over another in CAPI-plus and KtN. These paradata can also provide insight into how switches in survey mode and increased fieldwork effort counteract nonresponse (bias). Particularly in mixed-mode data collection, the success of a measure (e.g., number of reminders, amount of incentives) may vary by survey mode, so measures should be tailored to the mode (e.g., different number of reminders or incentives depending on the mode in CAPI-plus or KtN).

Challenges in Data Quality

As with all data collection methods, a challenge with mixed-mode and innovative methods is ensuring the quality of the survey data. Mixing modes results in survey data being collected under very different conditions (e.g., interviewer presence or absence, verbal or visual presentation of question stimuli, differing question formats); thus, mode effects can affect data quality and survey estimates (Conrad et al., 2022; de Leeuw & Hox, 2015; Endres et al., 2022; Lugtig et al., 2011; West et al., 2022). Moreover, when relatively new data collection methods are used that are unfamiliar to both interviewers and respondents, such as CAVI, little is known about the problems that may occur during the interview, such as interrupted speech and frozen or distorted video (Conrad et al., 2022), and about the impact of the new interview situation and the problems encountered on response behavior and data quality. These technical and other issues make it even more essential to take a closer look at the conditions under which the survey data are collected and to evaluate their quality thoroughly.

One advantage of CAPI-plus (when CAPI mode is selected) and CAVI is that interviewers and respondents can usually see each other, and interviewers can thus perceive respondents' attributes, facial expressions, and nonverbal cues. The visual interview-respondent interaction allows for an extensive collection of interviewer evaluations of respondent characteristics that can be used as proxy information for substantive analyses. Most importantly, we recommend the collection of detailed interviewer evaluations of the interview situation and respondent behavior to enable an informed survey data quality assessment. Especially in CAVI mode, new and unexpected interactions and problems may occur, which should be documented through comprehensive interviewer evaluations (e.g., screen sharing not working, technically related interruptions, acoustically related difficulty understanding questions, distractions from incoming emails and notifications on the respondent's device) to identify low-quality data and explain differences in data quality between survey modes. In addition, interviewer evaluations can help identify groups of respondents for whom CAVI is particularly problematic (e.g., less technically savvy respondents, elderly) and for whom another mode is preferable. Due to the lack of immediate proximity between interviewer and respondent, interviewers should be specifically trained to collect interviewer evaluations in CAVI mode so that they know exactly what to look for in the interview situation and how to interpret respondents' (non)verbal behaviors appropriately.

Conclusions and Considerations for Future Research

The range of interviewer-observed paradata in face-to-face interviewing is diverse, as are their purposes, as we have shown through a systematic overview of the previous literature. Moreover, we found that the usefulness of interviewer-observed paradata is often highly dependent on the interview context. Using CAPI-plus, CAVI, and KtN as examples, we have discussed the applicability

of interviewer-observed paradata, typically collected in face-to-face-only interviews, in mixed-mode and innovative data collection methods. We have shown that it is necessary to develop modified and new interviewer-observed paradata tailored to the specific needs of a data collection method to realize its full potential. Modified and new paradata require additional interviewer training and a thorough assessment of the quality and applicability of these paradata in the context of mixed-mode and innovative data collection methods, as the collection conditions may differ significantly from those of face-to-face-only interviews.

A worthwhile endeavor from our perspective is to combine interviewer-observed paradata with paradata from other survey modes. Mixed-mode and innovative methods that involve web-based data collection can profit from web paradata (e.g., response times, questionnaire navigation, and device information) that can be used to better understand question-answer processing on the part of respondents and to assess survey data quality (for a comprehensive overview of web paradata and their uses, see, for example, Callegaro, 2013; Kunz & Hadler, 2020; McClain et al., 2019). For example, like interviewer evaluations, response time data can indicate whether respondents have comprehension problems with individual questions or how much effort they put into answering them. These automatically collected web paradata can substitute for at least some interviewer evaluations and allow for the economical collection of paradata by saving interviewer time to record interviewer-observed paradata and increasing standardization by eliminating interviewer variability in the collection of these paradata. Or they can be collected supplementally to compare interviewer evaluations and web paradata to assess their quality per se and to decide what type of paradata will be most useful in future data collection.

Survey researchers and practitioners have recognized in the wake of the COVID-19 pandemic that future survey data collection will likely include multiple modes and different approaches to best meet respondents' needs. It is therefore necessary to further develop the practice of paradata collection and use and adapt it to the new data collection conditions, particularly mixed-mode settings. We would like to stimulate future research to provide evidence-based insights into how paradata from different survey modes can be usefully supplemented and combined to improve the efficiency of data collection and the quality of survey data.

ORCID iDs

Tanja Kunz ⓘ https://orcid.org/0000-0001-8460-2583
Jessica Daikeler ⓘ https://orcid.org/0000-0002-4879-8344

References

Ansolabehere, S., Rodden, J., & Snyder, J. M. (2008). The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting. *American Political Science Review, 102*(2), 215–232. https://doi.org/10.1017/S0003055408080210

Bates, N., Dahlhamer, J., & Singer, E. (2008). Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics*, *24*(4), 591–612.

Bates, N., & Piani, A. (2005). Participation in the National Health Interview Survey: Exploring reasons for reluctance using contact history process data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference*. National Center for Education Statistics. https://nces.ed.gov/FCSM/pdf/2005FCSM_Bates_Piani_VIB.pdf

Biemer, P. P., Chen, P., & Wang, K. (2013). Using level-of-effort paradata in non-response adjustments with application to field surveys. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *176*(1), 147–168. https://doi.org/10.1111/j.1467-985X.2012.01058.x

Blom, A. G. (2012). Explaining cross-country differences in survey contact rates: Application of decomposition methods. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *175*(1), 217–242. https://doi.org/10.1111/j.1467-985x.2011.01006.x

Blom, A. G., de Leeuw, E. D., & Hox, J. J. (2011). Interviewer effects on nonresponse in the European Social Survey. *Journal of Official Statistics*, *27*(2), 359–377.

Boniface, S., Scholes, S., Shelton, N., & Connor, J. (2017). Assessment of non-response bias in estimates of alcohol consumption: Applying the continuum of resistance model in a general population survey in England. *PLoS ONE*, *12*(1), 1–12. https://doi.org/10.1371/journal.pone.0170892

Bricker, J. (2014). *Survey incentives, survey effort, and survey costs*. FEDS Working Paper No. 2014-74. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2503478

Callegaro, M. (2013). Paradata in web surveys. In F. Kreuter (Ed), *Improving surveys with paradata. Analytic Uses of process information* (pp. 261–279). Wiley & Sons.

Casas-Cordero, C. (2010). *Neighborhood characteristics and participation in household surveys (Doctoral Thesis)*. https://drum.lib.umd.edu/handle/1903/11255

Cleary, A., Fisher, S., Green, A., Jakeways, S., Lloyd, R., & McTiernan, L. (2021). *Embracing mixed mode. More than a plan B*. Ipsos Views. https://www.ipsos.com/sites/default/files/ct/publication/documents/2021-02/20210211-embracingmixedmodes_final.pdf

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2022). Video in survey interviews: Effects on data quality and respondent experience. methods, data, analyses, online first. https://doi.org/10.12758/mda.2022.13

Cornick, P., d'Ardenne, J., Maslovskaya, O., Mesplie-Cowan, S., Nicolaas, G., & Smith, P. (2022). *Review of options for the National Survey for Wales*. Welsh Government. https://www.gov.wales/sites/default/files/statistics-and-research/2022-04/review-of-options-for-the-national-survey-for-wales.pdf

Davis, C. L., Aguilar, E. E., & Speer, J. G. (1999). Associations and activism: Mobilization of urban informal workers in Costa Rica and Nicaragua. *Journal of Interamerican Studies and World Affairs*, *41*(3), 35–66. https://doi.org/10.2307/166158

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, *21*(2), 233–255.

de Leeuw, E. D. (2018). Mixed-Mode: Past, present, and future. *Survey Research Methods*, *12*(2). https://doi.org/10.18148/srm/2018.v12i2.7402

de Leeuw, E. D., & Hox, J. J. (2015). Survey mode and mode effects. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds), *Improving survey methods. Lessons from recent research* (pp. 22–34). Routledge.

Dillman, D. A. (2005). Mixed-mode surveys. In S. J. Best & B. Radcliff (Eds.), *Polling America: An encyclopedia of public opinion* (pp. 149–153). Greenwood Press.

Durrant, G. B., D'Arrigo, J., & Steele, F. (2011). Using paradata to predict best times of contact, conditioning on household and interviewer influences. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *174*(4), 1029–1049. https://doi.org/10.1111/j.1467-985X.2011.00715.x

Durrant, G. B., D'Arrigo, J., & 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: Statistics in Society*, *176*(1), 251–269. https://doi.org/10.1111/j.1467-985X.2012.01073.x

Durrant, G. B., Maslovskaya, O., & Smith, P. W. F. (2017). Using prior wave information and paradata: Can they help to predict response outcomes and call sequence length in a longitudinal study? *Journal of Official Statistics*, *33*(3), 801–833. https://doi.org/10.1515/jos-2017-0037

Durrant, G. B., Maslovskaya, O., & Smith, P. W. F. (2019). Investigating call record data using sequence analysis to inform adaptive survey designs. *International Journal of Social Research Methodology*, *22*(1), 37–54. https://doi.org/10.1080/13645579.2018.1490981

Durrant, G. B., & 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: Statistics in Society*, *172*(2), 361–381. https://doi.org/10.1111/j.1467-985X.2008.00565.x

Ellis, B. J., Bates, J. E., Dodge, K. A., Fergusson, D. M., Horwood, L. J., Pettit, G. S., & Woodward, L. (2003). Does father absence place daughters at special risk for early sexual activity and teenage pregnancy? *Child Development*, *74*(3), 801–821. https://doi.org/10.1111/1467-8624.00569

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2022). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods*, *11*, 144–159. https://doi.org/10.1017/psrm.2022.30

Fryer, R. G., & Levitt, S. D. (2013). Testing for racial differences in the mental ability of young children. *American Economic Review*, *103*(2), 981–1005. https://doi.org/10.1257/aer.103.2.981

Gay, C. (2014). Knowledge matters: Policy cross-pressures and black partisanship. *Political Behavior*, *36*(1), 99–124. https://doi.org/10.1007/s11109-013-9227-3

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675. https://doi.org/10.1093/poq/nfl033

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *169*(3), 439–457. https://doi.org/10.1111/j.1467-985X.2006.00423.x

Hanly, M., Clarke, P., & Steele, F. (2016). Sequence analysis of call record data: Exploring the role of different cost settings. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *179*(3), 793–808. https://doi.org/10.1111/rssa.12143

Haug, M. R., & Folmar, S. J. (1986). Longevity, gender, and life quality. *Journal of Health and Social Behavior*, *27*(4), 332–345. https://doi.org/10.2307/2136948

Jeannis, M., Terry, T., Heman-Ackah, R., & Price, M. (2013). Video interviewing: An exploration of the feasibility as a mode of survey application. *Survey Practice*, *6*(1), 1–5. https://doi.org/10.29115/SP-2013-0001

Kantar Public. (2021). *The future of sequential mixed mode in survey design*. https://www.kantarpublic.com/inspiration/thought-leadership/future-of-sequential-mixed-mode-survey-design

Kastberg, S., & Siegler, V. (2022). *Impact of COVID-19 on ONS social survey data collection*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/impactofcovid19ononssocialsurveydatacollection

Kennickell, A. B. (2017). Getting to the top: Reaching wealthy respondents in the SCF. *Statistical Journal of the IAOS*, *33*(1), 113–123. https://doi.org/10.3233/SJI-160295

Kirchner, A., Olson, K., & Smyth, J. D. (2017). Do interviewer postsurvey evaluations of respondents' engagement measure who respondents are or what they do? A behavior coding study. *Public Opinion Quarterly*, *81*(4), 817–846. https://doi.org/10.1093/poq/nfx026

Konicki, S., & Adams, T. (2016). Adaptive design research for the 2020 Census. *Statistical Journal of the IAOS*, *32*(2), 167–176. https://doi.org/10.3233/SJI-161007

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

Kreuter, F., & Kohler, U. (2009). Analyzing contact sequences in call record data. Potential and limitations of sequence indicators for nonresponse adjustments in the European Social Survey. *Journal of Official Statistics*, *25*(2), 203–226.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T. M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R. M., Raghunathan, T. E., & Raghunathan, T. E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *173*(2), 389–407. https://doi.org/10.1111/j.1467-985X.2009.00621.x

Krueger, B. S., & West, B. T. (2014). Assessing the potential of paradata and other auxiliary data for non-response adjustments. *Public Opinion Quarterly*, *78*(4), 795–831. https://doi.org/10.1093/poq/nfu040

Kuenzi, M., Robert, C., & Cooper, A. (2022). Pandemic pivot: The shift from CAPI to CATI, single-device vs. multi-device approaches, and the future of data collection modes. AAPOR 2022.

Kunz, T., & Hadler, P. (2020). *Web paradata in survey research*. GESIS - Leibniz Institute for the Social Sciences (GESIS Survey Guidelines). https://doi.org/10.15465/gesis-sg_en_037

Lee, H., & Waite, L. J. (2018). Cognition in context: The role of objective and subjective measures of neighborhood and household in cognitive functioning in later life. *The Gerontologist*, *58*(1), 159–169. https://doi.org/10.1093/geront/gnx050

Lee, S., Fredriksen-Goldsen, K. I., McClain, C., Kim, H.-J., & Suzer-Gurtekin, Z. T. (2018). Are sexual minorities less likely to participate in surveys? An examination of proxy nonresponse measures and associated biases with sexual orientation in a population-based health survey. *Field Methods*, *30*(3), 208–224. https://doi.org/10.1177/1525822x18777736

Leventhal, T., & Brooks-Gunn, J. (2005). Neighborhood and gender effects on family processes: Results from the moving to opportunity program. *Family Relations*, *54*(5), 633–643. https://doi.org/10.1111/j.1741-3729.2005.00347.x

Loosveldt, G., & Joye, D. (2016). Defining and assessing survey climate. In C. Wolf, D. Joye, T. W. Smith, & Y.-c. Fu (Eds), *The Sage Handbook of survey methodology* (pp. 67–76). https://doi.org/10.4135/9781473957893

Lugtig, P., Lensvelt-Mulders, G. J. L. M., Frerichs, R., & Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. *International Journal of Market Research*, *53*(5), 669–686. https://doi.org/10.2501/ijmr-53-5-669-686

Luijkx, R., Jónsdóttir, G. A., Gummer, T., Ernst Stähli, M., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S. Þ., Hjaltason, Á. B., Joye, D., Lomazzi, V., Maineri, A. M., Milbert, P., Ochsner, M., Pollien, A., Sapin, M., Solanes, I., & Wolf, C. (2021). The European Values Study 2017: On the way to the future using mixed-modes. *European Sociological Review*, *37*(2), 330–346. https://doi.org/10.1093/esr/jcaa049

Lynn, P., & Clarke, P. (2002). Separating refusal bias and non-contact bias: Evidence from UK national surveys. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *51*(3), 319–333. https://doi.org/10.1111/1467-9884.00321

Maitland, A., Casas-Cordero, C., & Kreuter, F. (2009). An evaluation of nonresponse bias using paradata from a health survey. *JSM Proceedings of the Government Statistics Section* (pp. 370–378). American Statistical Association.

McCann, J. A., & Lawson, C. (2006). Presidential campaigns and the knowledge gap in three transitional democracies. *Political Research Quarterly*, *59*(1), 13–22. https://doi.org/10.1177/106591290605900102

McClain, C. A., Couper, M. P., Hupp, A. L., Keusch, F., Peterson, G., Piskorowski, A. D., & West, B. T. (2019). A typology of web survey paradata for assessing total survey error. *Social Science Computer Review*, *37*(2), 196–213. https://doi.org/10.1177/0894439318759670

Mellinger, G. D., Huffine, C. L., & Balter, M. B. (1982). Assessing comprehension in a survey of public reactions to complex issues. *Public Opinion Quarterly*, *46*(1), 97–109. https://doi.org/10.1086/268702

Møller, V. (1992). Spare time use and perceived well-being among black South African youth. *Social Indicators Research*, *26*(4), 309–351. https://doi.org/10.1007/BF00347894

Moore, J. C., Durrant, G. B., & Smith, P. W. F. (2018). Data set representativeness during data collection in three UK social surveys: Generalizability and the effects of auxiliary covariate choice. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *181*(1), 229–248. https://doi.org/10.1111/rssa.12256

Morris, P., Bloom, D., Kemple, J., & Hendra, R. (2003). The effects of a time-limited welfare program on children: The moderating role of parents' risk of welfare dependency. *Child Development*, *74*(3), 851–874. https://doi.org/10.1111/1467-8624.00572

Olson, K. (2013). Paradata for nonresponse adjustment. *The Annals of the American Academy of Political and Social Science*, *645*(1), 142–170. https://doi.org/10.1177/0002716212459475

Perales, F., & Baffour, B. (2018). Respondent mental health, mental disorders and survey interview outcomes. *Survey Research Methods*, *12*(2), 161–176. https://doi.org/10.18148/srm/2018.v12i2.7225

Peytchev, A., & Olson, K. (2007). Using interviewer observations to improve nonresponse adjustments: NES 2004. In *JSM Proceedings of the Survey Research Methods Section* (pp. 3364–3371). American Statistical Association.

Prigerson, H. G., Bierhals, A. J., Kasl, S. V., Reynolds, C. F., Shear, M. K., Day, N., Beery, L. C., Newsom, J. T., & Jacobs, S. (1997). Traumatic grief as a risk factor for mental and physical morbidity. *The American Journal of Psychiatry*, *154*(5), 616–623. https://doi.org/10.1176/ajp.154.5.616

Purdon, S., Campanelli, P., & Sturgis, P. (1999). Interviewers' calling strategies on face-to-face interview surveys. *Journal of Official Statistics*, *15*(2), 199–216.

Safir, A., & Tan, L. (2009). Using contact attempt history data to determine the optimal number of contact attempts. In *JSM Proceedings of the Survey Research Methods Section* (pp. 5970–5984). American Statistical Association.

Sakshaug, J. W., Couper, M. P., & Ofstedal, M. B. (2010). Characteristics of physical measurement consent in a population-based survey of older adults. *Medical Care*, *48*(1), 64–71. https://doi.org/10.1097/MLR.0b013e3181adcbd3

Saran, A., & White, H. (2018). Evidence and gap maps: A comparison of different approaches. *Campbell Systematic Reviews*, *14*(1), 1–38. https://doi.org/10.4073/cmdp.2018.2

Scherpenzeel, A. (2017). Mixing online panel data collection with innovative methods. In S. Eifler & F. Faulbaum (Eds), *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung* (pp. 27–49). Springer VS.

Schober, M. F. (2018). The future of face-to-face interviewing. *Quality Assurance in Education*, *26*(2), 290–302. https://doi.org/10.1108/QAE-06-2017-0033

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice*, *13*(1), 1–11. https://doi.org/10.29115/SP-2020-0014

SHARE-ERIC (2022). *SHARE Annual Activity Report 2021 – 2022*. SHARE-ERIC. https://share-eric.eu/fileadmin/user_upload/Other_Publications/Annual_Activity_Reports/SHARE-ERIC_AnnualActivityReport2021-22.pdf

Snilstveit, B., Vojtkova, M., Bhavsar, A., & Gaarder, M. (2013). *Evidence gap maps: A tool for promoting evidence-informed policy and prioritizing future research*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2367606

Steele, F., & Durrant, G. B. (2011). Alternative approaches to multilevel modelling of survey non-contact and refusal. *International Statistical Review*, *79*(1), 70–91. https://doi.org/10.1111/j.1751-5823.2011.00133.x

Treier, S., & Hillygus, D. S. (2009). The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly*, *73*(4), 679–703. https://doi.org/10.1093/poq/nfp067

Vandenplas, C., Loosveldt, G., & Beullens, K. (2017). Fieldwork monitoring for the European Social Survey: An illustration with Belgium and the Czech Republic in round 7. *Journal of Official Statistics*, *33*(3), 659–686. https://doi.org/10.1515/jos-2017-0031

van Tubergen, F., & Kalmijn, M. (2005). Destination-language proficiency in cross-national perspective: A study of immigrant groups in nine Western countries. *American Journal of Sociology*, *110*(5), 1412–1457. https://doi.org/10.1086/428931

Vercruyssen, A., & Loosveldt, G. (2017). Using Google Street View to validate interviewer observations and predict nonresponse: A Belgian case study. *Survey Research Methods*, *11*(3), 345–360. https://doi.org/10.18148/srm/2017.v11i3.6301

Vercruyssen, A., van de Putte, B., & Stoop, I. A. L. (2011). Are they really too busy for survey participation? The evolution of busyness and busyness claims in Flanders. *Journal of Official Statistics*, *27*(4), 619–632.

Villar, A., & Fitzgerald, R. (2017). Using mixed modes in survey research: Evidence from six experiments in the ESS. In M. J. Breen (Ed.), *Values and Identities in Europe: Evidence from the European Social Survey* (pp. 273–310). Routledge.

Wagner, J., Valliant, R., Hubbard, F., & Jiang, L. C. (2014). Level-of-effort paradata and nonresponse adjustment models for a national face-to-face survey. *Journal of Survey Statistics and Methodology*, *2*(4), 410–432. https://doi.org/10.1093/jssam/smu012

Weissman, J. S., Levin, K., Chasan-Taber, S., Massagli, M. P., Seage, G. R., & Scampini, L. (1996). The validity of self-reported health-care utilization by AIDS patients. *AIDS*, *10*(7), 775–783. https://doi.org/10.1097/00002030-199606001-00013

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: Statistics in Society*, *176*(1), 211–225. https://doi.org/10.1111/j.1467-985X.2012.01038.x

West, B. T., Conrad, F. G., Kreuter, F., & Mittereder, F. (2018). Can conversational interviewing improve survey response quality without increasing interviewer effects? *Journal of the Royal Statistical Society Series A: Statistics in Society*, *181*(1), 181–203. https://doi.org/10.1111/rssa.12255

West, B. T., & Groves, R. M. (2013). A propensity-adjusted interviewer performance indicator. *Public Opinion Quarterly*, *77*(1), 352–374. https://doi.org/10.1093/poq/nft002

West, B. T., Kreuter, F., & Trappmann, M. (2014). Is the collection of interviewer observations worthwhile in an economic panel survey? New evidence from the German Labor Market and Social Security (PASS) Study. *Journal of Survey Statistics and Methodology*, *2*(2), 159–181. https://doi.org/10.1093/jssam/smu002

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology*, *10*(2), 317–336. https://doi.org/10.1093/jssam/smab040

Winter, N. J. G. (2010). Masculine republicans and feminine democrats: Gender and Americans' explicit and implicit images of the political parties. *Political Behavior*, *32*(4), 587–618. https://doi.org/10.1007/s11109-010-9131-z

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting general social surveys as self-administered mixed-mode surveys. *Public Opinion Quarterly*, *85*(2), 623–648. https://doi.org/10.1093/poq/nfab039

Zaslow, M. J., Berlin, L. J., Brooks-Gunn, J., Coiro, M. J., Spiker, D., Moore, K. A., & Brown, B. (1995). *Differentiating among measures of parenting behavior in two studies of mothers and their preschoolers: The role of informant and context*. https://cms.childtrends.org/wp-content/uploads/1995/01/1995-07DifferentiatingAmongMeasuresofParentingBehaviorinTwoStudiesofMothersandTheirPreschoolers.pdf

Zelenak, M. F., & Davis, M. C. (2013). *Impact of multiple contacts by computer-assisted telephone interview and computer-assisted personal interview on final interview outcome in the American Community Survey*. U.S. Census Bureau Working Papers. https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Zelenak_01.pdf