Participation time in self-administered panel surveys: measurements and consequences

Inauguraldissertation

zur Erlangung des akademischen Grades eines

Doktors der Sozialwissenschaften der

Universität Mannheim

vorgelegt von

Isabella Luise Minderop

Dekan der Fakultät für Sozialwissenschaften
Prof. Dr. Michael Diehl

Betreuer:innen
PD Dr. Tobias Gummer
Prof. Dr. Bella Struminskaya

Gutachter:innen
PD Dr. Tobias Gummer
Prof. Dr. Bella Struminskaya
Prof. Dr. Florian Keusch

Tag der Disputation: 14.09.2023

Acknowledgements

I would like to express my sincere gratitude to all who have contributed to this dissertation.

First and foremost, I would like to thank my supervisors, Tobias Gummer and Bella Struminskaya, whose guidance, expertise, and encouragement throughout my research have been invaluable. I appreciate the time and effort you have invested in my success. Your constructive criticism has challenged me to refine my ideas and strengthen my arguments. Your insights have been instrumental in shaping this work, and I am truly grateful for your mentorship and will always value the lessons I have learned under your guidance. I further want to thank Florian Keusch for reviewing this dissertation.

I also want to express my gratitude to GESIS, where I received financial and academic support. Especially, I am grateful to my former colleagues at GESIS — Bernd Weiß, Steffen Pötzschke, Ines Schaurer, Kai Weyandt, Ingo Konradt, Jan-Philipp Kolb, Tobias Heycke, Anna Hebel, Joachim Piepenburg, Kai Böge, Gabi Wahlig, Mirjan Kummerow, Maikel Schwerdtfeger, Sven Stadtmüller, David Bretschi, and Katharina Schmidt — who not only assessed and processed the data on which this dissertation is based but also offered invaluable encouragement and feedback that helped me to improve my research and stay motivated throughout the process. Thank you for always creating a supportive and collaborative research and work environment that made completing this dissertation possible. I would particularly like to acknowledge Bernd Weiß, who generously shared his expertise and always encouraged me to put my dissertation first. Your contribution benefitted this dissertation greatly, and I am immensely grateful for your support. Thank you, Anna-Carolina Haensch, Verena Ortmanns, and Fabienne Krämer, for your feedback and the many hours we spent discussing our research projects.

I am deeply grateful for my family, Raphaela, Ralph, Gernot, and Severin. You supported me whenever I needed support, pushed me whenever I needed pushing, and distracted me whenever I needed distraction.

Table of contents

List of tables

List of figures

1. Introduction

Surveys have long been an essential tool for discovering and monitoring changes in opinions and behaviors in society (e.g., Groves et al., 2009). Data collected in surveys can be used for many purposes, such as informing policy making, public opinion research, advertising, and market research (Alwin, 2007). Longitudinal surveys are especially relevant, since they have several advantages over cross-sectional surveys (Lynn, 2009): First, multiple data collections usually enable the collection of more data than a single data collection, which additionally can be more accurate when the data collection is timelier after an event in question compared to a longer retrospective recall. Second, maintaining an existing sample of panelists often is more cost-effective than recruiting new respondents for each survey. Third, multiple data collections enable researchers to quickly identify conspicuous data that may have resulted from error. Fourth, multiple data collections enable researchers to analyze changes within society over time. In contrast to other longitudinal data of which the respondents may vary, panel data are unique in that the same respondents participate over multiple occasions of the data collection. This repeated data collection enables the analysis of not only gross change but also change on the individual level. Thus, we not only can observe the change, but also have an opportunity to observe the underlying mechanisms responsible for the change intra-individually. Fifth, when researchers can address changes at the individual level, they can further calculate measures of stability that indicate whether a change that occurred has not reversed or whether the opposite is the case, and respondents who change their behavior frequently change it again. Sixth, in analyzing changes over time, unobserved individual effects that remain stable, such as sex, can be controlled for when using panel data. Seventh, multiple data collection occasions enable the identification of causality by revealing a temporal sequence.

However, researchers can only benefit from these advantages if the data derived from panel surveys is correct. The data should reflect the opinions and behaviors of respondents (Groves et al., 2009). If data is missing or deviates from the “true” values, it may lead to false conclusions, and researchers may assume relationships that actually do not exist. Researchers can control most error sources, for instance, by choosing the right population of interest or drawing a representative sample (Groves, 2005). However, one crucial issue concerning data quality is only under limited control by researchers: the cooperation of panelists during the whole participation process. Thus, over the past decades, obtaining respondents’ cooperation to participate and report their opinions and behavior has been a relevant topic in survey methodology (e.g., Glaser, 2012; Olson, 2013). The present dissertation adds to the body of literature on respondent cooperation by introducing one particularly interesting potential predictor of panelist cooperation — participation time, which is the time respondents take to return their self-administered survey. For online surveys, participation time is the time it takes a respondent to submit the last question they answer, no matter whether this also is the last question of the questionnaire. For mail questionnaires, participation time is the time it takes the return envelope to be delivered. In this chapter, I first introduce the process that panelists undergo while participating. In this process, two important decisions determine their cooperation: whether or not to participate and how many cognitive resources they will put into their participation. I then summarize the current literature and indicate the gaps left to fill. I conclude with an outline of the following chapters.

1.1. The participation process

The participation process for one wave can be described in three stages: survey receipt, survey response, and survey return. In Stage 1, *survey receipt*, panelists receive the survey from the survey agency. At this stage, they must decide whether or not to participate. If panelists choose not to participate, they non-respond to this wave. Repeated wave nonresponse may lead to panel attrition, which is the problem that occurs when panelists no longer participate in future survey waves. Multiple definitions exist for panel attrition (Groves, 2005). Panel attrition may include one or multiple times of wave nonresponse as well as permanent drop-out from a sample. It is a central problem for longitudinal surveys for multiple reasons. First, when panelists attrite, they will not provide data in the future, which is a great loss because next to a loss of statistical power, the time series of previously collected data is ended. The worth of such time series increases with the number of data collections. Second, attrited panelists need to be replaced with new panelists, and recruiting them is more costly and less efficient than maintaining an existing sample. Third, when the panelists who attrite from a sample differ from those who remain in the sample, the sample might not represent the whole population of interest anymore, which introduces bias to the collected data. Therefore, researchers have engaged in finding determinants for attrition. One of the determinants which they successfully used for this purpose was previous participations (Lugtig, 2014; Roßmann & Gummer, 2016). However, previous participations provide much more detailed information when not only participation or nonparticipation are taken into account, but also participation time. Respondents who return their survey earlier may be more motivated to participate and therefore may be more likely to remain in a sample. In contrast, later respondents may be more likely to become nonrespondents and attritors in future waves. If panelists choose to participate, they begin responding to the survey and proceed to Stage 2.

In Stage 2, the *survey response*, respondents must decide how many cognitive resources to put into their response. This decision also includes whether they want to finish responding to the survey or break off at some point. If respondents choose to invest few resources in their survey response, these responses may have a lower data quality. Data quality is low when the answer given by a respondent does not reflect their true answer (Groves et al., 2009). Thus, data quality may be lower when, for instance, respondents do not read the question carefully and mistake the task required of them or when they know what is expected of them but do not think about the correct answer long enough. Data quality also is reduced when respondents know the correct response but prefer to avoid reporting it or when they skip the mental process and tick a random response. Such distortions can falsify analysis results, which may lead to false conclusions and reactions to these conclusions. Thus, interest is considerable for finding explanations for differences in reported data quality and for improving data quality in future studies. One of the arguments used to explain differences in data quality is that motivation to participate in a survey influences data quality. Moreover, it has been argued that motivation to participate also influences participation time—motivated respondents participate early, whereas respondents with lesser motivation postpone their participation. If motivation to participate influences both participation time and data quality, the date of participation might serve as an indicator of motivation. If respondents choose to invest many resources in their survey response, their data quality is likely to be high. Either way, respondents then proceed to Stage 3.

In Stage 3, the *survey return*, mail respondents must decide whether or not they want to return their responded survey to the survey infrastructure. If they choose not to return the survey, even though they have put in the effort to respond to it, they have non-responded to this wave in the eyes of the survey agency. As in Stage 1, repeated wave nonresponse may lead to panel attrition. If the mail respondents choose to return the survey, the participation process for this wave is completed in terms of their required actions. Ideally, the survey agency will receive the response and invite the panelist to the next survey wave, and the whole process will start again. If non-responding panelists continuously refrain from participating, this may lead to panel attrition, depending on the guidelines of the survey agency. At Stage 3, mode differences between the online and mail respondents become the most obvious. This stage does not exist for online respondents because the survey agency receives the data they enter without any further necessary actions. If online respondents stop participating in the middle of a survey, the survey agency will receive the data gathered up to this point. In contrast, response and return are two different tasks for mail respondents, since they can think about returning the survey before they actually do. Further, survey return means taking it to a post box, which is quite an effortful task. However, it is impossible for a survey agency to differentiate whether wave nonresponse by mail respondents happened at Stage 3 or Stage 1. Thus, since it is unlikely that mail respondents would exert the effort to participate and then fail to return the survey, this stage is ignored in the future chapters of this dissertation.

The participation process takes a certain amount of time that may differ for respondents, depending on whether they respond right away, after some time, or not at all. This time is *participation time*, which is the time the whole participation process takes, so participation time increases when respondents think about whether to participate, postpone their participation, or interrupt their participation. The reasons for these actions — such as low motivation to participate, a lack of trust in data security, or little available time — could also be related to panel attrition or data quality (Olson, 2013). Thus, respondents with a low motivation to participate could think longer about whether to participate, whereas respondents with a higher motivation would not question their participation. Low-motivated respondents could be eager to finish the survey fast, whereas highly motivated respondents would be expected to aim at providing good answers. At the next survey invitation, low-motivated respondents may decide against participation and non-respond to this survey and possibly any following ones. Hence, lower data quality potentially could be seen as a first stage of panel attrition. An indication of that possibility would be the finding of Loosveldt et al. (2019) who used item nonresponse as a predictor of unit nonresponse in a panel survey and found that item nonresponse in one wave was related to unit nonresponse in the next wave.

Participation time is not measured by questioning the respondent; rather, the survey infrastructure assesses it as paradata. Therefore, participation time has four advantages over other predictors of panel attrition or data quality. First, the assessment of participation time does not increase the survey burden for respondents. Second, participation time is unlikely to be falsified by respondents because they usually are unaware of the assessment. Third, participation time is easily accessible for almost every online survey, so data users can access it even when the data were collected for purposes other than participation time analyses. Fourth, participation time is comparable across multiple surveys when researchers operationalize it in the same manner for all surveys.

1.2. Research gap

The advantages that participation time offers compared to other predictors of panel attrition and data quality sound very promising. However, we need more resilient information about how well participation time can predict panel attrition and data quality. Regarding data quality, we can draw on several studies about how participation time is related to data quality indicators. Some studies have found significantly more item nonresponse among late respondents (Fricker & Tourangeau, 2010; Friedman et al., 2003; Kunz, 2010; Tancreto & Bentley, 2005), but many others have failed to find a statistically significant relationship (Diaz de Rada, 2005; Green, 1991; Helasoja, 2002; Schoenman et al., 2003; Sobal & Ferentz, 1989). Friedman et al. (2003) observed a higher use of non-substantive categories such as “don’t know” or “does not apply.” Green (1991) reported that late respondents are less likely to respond to open questions. De Leeuw & Hox (1988) did not find a higher likelihood of straightlining among later respondents compared to earlier respondents. When Kreuter et al. (2014) compared survey responses to administrative records, they found that late respondents were more likely to give incorrect responses when asked about their receipt of welfare benefits, but not to other questions. Kaminska et al. (2010) found that the relationship between participation time and the reporting of welfare benefits could be explained by cognitive ability. Other studies found that later respondents were less likely to provide matching information (Armenakis & Lett, 1982; Fricker & Tourangeau, 2010; Preisendörfer & Wolter, 2014; Skarbek-Kozietulska et al., 2012). A considerable drawback of the existing literature is that studies have focused on one or few data quality indicators and did not compare the relationships of participation time with several data quality indicators. Multiple studies with different data quality indicators are not necessarily comparable because the design of multiple surveys may be different, and the survey design may significantly impact participation time and data quality. Thus, we cannot make conclusive statements about whether participation time is related to data quality in general or only to selective data quality indicators. A better understanding of the relationship between participation time and data quality might help to predict data quality and possibly increase it through targeted interventions. Chapter 2 targets this research gap.

Longitudinal studies using participation time to examine panel attrition are rare. Lugtig (2014) and Roßmann & Gummer (2016) have shown that response or nonresponse to previous panel waves may be related to future panel participation, but they did not address the participation time. Cohen et al. (2000) found a relationship between participation time and future panel attrition, but their study was restricted to two waves. Therefore, information is lacking as to whether participation time can be a tool for identifying future panel attrition. This research gap is addressed in Chapters 3 and 4.

The current research about participation time involves two other challenges. First, numerous possibilities exist for operationalizing participation time. For example, participation time can be operationalized metrically (e.g., Gummer & Struminskaya, 2020; Skarbek-Kozietulska et al., 2012), binarily (e.g., Voigt et al., 2003), or categorically (e.g., Kreuter et al., 2014). Participation time may refer to actual time in days, or it can be measured as a cumulative sample size indicating the proportion of respondents in a wave who participated prior to a selected respondent, for example, 25%. When different operationalizations of participation time are used, analysis results are not necessarily comparable. This lack of comparability decreases the value of these results for the social sciences because other researchers cannot necessarily classify the results. Chapter 3 focuses on different operationalizations.

Second, until now, participation time has been analyzed only cross-sectionally. However, researchers can use longitudinal data to calculate how habitual a participation time is and use this finding as a possible indicator for other variables. Respondents who employ a specific participation time only once may behave differently than respondents who employ a regular participation time behavior. One explanation for this behavior is that the decision of whether to participate can be made automatically or reflectively (Esser, 2011; Kroneberg, 2014). Chapter 4 addresses whether differences exist in the likelihood of attrition between those respondents who employ more and less habitual participation time. The following sections introduce the dissertation chapters in more detail.

1.3. Outline of the dissertation

The analyses of this dissertation are based on GESIS Panel data. The GESIS Panel is a German probability-based mixed-mode open panel (Bosnjak et al., 2018) that contains around 5,000 panelists who are at least 18 years old. The sample was initially recruited in 2013 and consists of a random sample drawn from municipal population registers. It was refreshed in 2016, 2018, and 2021. Panelists from all cohorts first participated in a recruitment interview and were invited subsequently to complete a self-administered welcome survey. The respondents of this welcome survey were regarded as regular panelists. The American Association for Public Opinion Research (AAPOR) Response Rate 1 in the face-to-face recruitment interview was 36% for the initial recruitment cohort, 33% for the 2016 refreshment cohort, and 31% for the

2018 refreshment cohort. In the initial recruitment cohort, 80% of the recruited respondents participated in the welcome survey; in the 2016 refreshment cohort, 81% participated in the welcome survey; and in the 2018 recruitment, 78% of the recruited respondents participated in the welcome survey. Every two months, all panelists are invited to complete a GESIS Panel survey. With wavily completion rates of around and above 90%, almost all panelists participate: around 65% participate online, and the others receive a paper-and-pencil questionnaire together with a postal invitation. All panelists receive a 5-EUR prepaid incentive. Online respondents additionally receive an email with a web link to the survey and email reminders both one and two weeks after the field start. The GESIS Panel is open to researchers from all fields to submit questionnaire proposals, which leads to highly diverse survey topics. The surveys usually take around 20 to 30 minutes to complete and cover at least four topics. Panel attrition in the GESIS Panel can have two causes: first, panelists can decide to attrite themselves and stop their participation; second, panelists who have not participated in three subsequent survey waves are removed from the sample by the GESIS Panel.

This dissertation consists of three studies that explored different aspects of participation time. The chapters can be read independently and may be repetitive in some aspects, such as the description of the data.

1.3.1. Chapter 2: Examining the relationship between participation time and multiple data quality indicators

Chapter 2 addresses the relationship between participation time and data quality. Participation time has been argued to be a good predictor of data quality because the motivation to participate is reflected in both participation time and data quality (Olson, 2013). Although previous studies have found relationships between participation time and data quality indicators, they mainly focused on one indicator and, thus, did not compare multiple data quality indicators with each other. This comparison is crucial, because a precise definition of how data quality can be operationalized is lacking. Usually, researchers operationalize data quality using the indicators available in their data. Since the availability of indicators in data can differ between datasets, it needs to be determined whether the operationalizations of data quality are comparable. Researchers with many similar items might use a straightlining indicator as an operationalization for data quality, whereas researchers with many open questions might operationalize data quality by investigating the length of the open responses. When deriving statements about data quality from these singular indicators, researchers assume that data quality indicators measure the same underlying construct. However, without comparing these data quality indicators to each other, we do not know whether this assumption is valid. Thus, it is not surprising that the literature on the relationship between participation time and data quality is mixed (Friedman et al., 2003; Helasoja, 2002; Kaminska et al., 2010; Kreuter et al., 2014; Preisendörfer & Wolter, 2014). In a comparative study, I address this research gap by investigating the relationship between participation time and 16 data quality indicators. The research goal is two-tiered and addresses the following questions: Is there a relationship between participation time and data quality? If a relationship is present, does it include every data quality indicator or only a subset of indicators? Based on the framework of satisficing, I argue that respondents who return the survey later may be less motivated and, therefore, more likely to satisfice. I investigate the following 16 dependent variables: (i) item nonresponse, (ii) the use of the "don't know" category, (iii) motivated misreporting, (iv) acquiescence, (v) choosing the first response category, (vi) choosing the middle response category, (vii) choosing the last response category, (viii) choosing the first, middle, or last response category, (ix) responding to open questions, (x) the length of responses to open questions, (xi) non-differentiation measured by the probability of differentiation, (xii) non-differentiation measured by the coefficient of variation, (xiii) straightlining, (xiv) the duration of the survey for online respondents, (xv) anchoring in looping questions, and (xvi) inconsistency between responses. Depending on the distribution of the data quality indicators, I estimate logistic, poisson, zero-inflated poisson, negative binomial, or zero-inflated negative binomial regressions.

1.3.2. Chapter 3: Predicting panel attrition using multiple operationalisations of response time

Chapter 3 addresses two problems: whether we can use participation time to predict attrition and how to measure participation time. Keeping panelists in a sample is a fundamental challenge for panel surveys because a sample that remains the same over multiple survey waves is essential for investigating longitudinal questions. A loss of panelists may be related to nonresponse error, higher variance, and higher survey costs (due to the recruitment of new panelists). Thus, panel attrition is a severe problem, although it may be predicted using simple means, namely participation time. We assume that respondents who participate earlier are higher motivated and provide more actual data (Bollinger & David, 2001) and that this theoretical expectation may be transferred to panel attrition. Thus, we expect higher motivated respondents to remain in a panel. We also assume that respondent characteristics could be related to participation time and their likelihood to provide accurate data (Olson, 2013).

Similarly, respondent characteristics could be expected to be related to participation time and the likelihood of attrition. Thus, we expect late respondents to be more likely to attrite. The second aim of Chapter 3 was to evaluate the operationalization of participation time. Vast possibilities exist to operationalize participation time, but the results from studies drawing on different operationalizations are not necessarily comparable. For example, participation time can be operationalized as a metric variable (e.g., Gummer & Struminskaya, 2020; Skarbek-Kozietulska et al., 2012), a binary variable (e.g., Voigt et al., 2003), or a categorical variable (e.g., Kreuter et al., 2014). Moreover, participation time can refer to actual time in days, but it also can be measured as a cumulative sample size (e.g., 25%) that indicates the proportion of respondents in a wave who participated before a given respondent. I operationalized participation time in 13 ways as (i) a metric measurement of the days until the survey return, (ii) whether respondents participated on the first day, (iii) in the first week, (iv) in the first two weeks, (v) before or after the first reminder, (vi) or before or after the second reminder. I further operationalized participation time as (vii) the number of contacts measured (viii), how many respondents participated prior to one respondent, (ix) whether the respondent was among the first 5% of respondents, (x) whether the respondent was among the first 10%, (xi) whether they were among the first 50%, (xii) among the last 10%, or (xiii) among the last 5%. I estimated random-effects logistic panel regressions to investigate which operationalizations of participation time are related to panel attrition and to observe whether the different operationalizations predict attrition similarly well or whether some operationalizations perform better than others.

1.3.3. Chapter 4: Now, later, or never? Using response-time patterns to predict panel attrition

Chapter 4 also investigates the relationship between participation time and panel attrition but changes the focus from measuring participation time to the reasons for this relationship, and also introduces habitual participation time. Habitual participation time can be defined as how regularly panelists participate with a similar participation time, in other words, whether they usually respond at the same time or switch between earlier and later responses. This chapter proposes a dual process of decision-making during the participation process. Some panelists act automatically without considering a different behavior than they usually engage in, whereas others reflect on how they should behave. Thus, panelists with a fixed participation time habit may be more likely to act automatically without considering nonresponse (compare Esser, 2011; Kroneberg, 2014; Lugtig, 2014). Panelists who lack habitual behavior could be more likely to reflect on their decision to participate. Chapter 4 also proposes a model of the mechanisms that impact the relationship between participation time and panel attrition. I argue that available time and the evaluation of the previous survey are related to both participation time and panel attrition. To test this hypothesis, I estimated three logistic regression models: one that uses participation time and the participation time habit to explain the variance in attrition, one that uses available time and the evaluation of the previous survey to explain the variance in attrition, and one full model with all the predictors. By using this approach, I disentangled whether the relationship between participation time and panel attrition exists because participation time and panel attrition are both related to available time and the evaluation of the previous survey or whether different mechanisms are responsible for this relationship. In addition, I estimated how well all three models predicted attrition to determine whether participation time could be a proxy for available time and survey evaluation.

1.4. References

Alwin, D. F. (2007). Margins of error: A study of reliability in survey measurement (Bd. 547). John Wiley & Sons.

Armenakis, A. A., & Lett, W. L. (1982). Sponsorship and follow-up effects on response quality of mail surveys. *Journal of Business Research*, *10*(2), 251–262. https://doi.org/10.1016/0148-29638290031-5

Bollinger, C. R., & David, M. H. (2001). Estimation with response error and nonresponse: Food-stamp participation in the SIPP. *Journal of Business & Economic Statistics*, *19*(2), 129–141. https://doi.org/10.1021/jf0483366

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review*, *36*(1), 103–115. https://doi.org/10.1177/0894439317697949

Cohen, S. B., Machlin, S. R., & Branscome, J. M. (2000). Patterns of survey attrition and reluctant response in the 1996 medical expenditure panel survey. *Health Services and Outcomes Research Methodology*, *1*(2), 131–148. https://doi.org/10.1023/A:1012543121850

De Leeuw, E. D., & Hox, J. J. (1988). Artifacts in Mail Surveys: The Influence of Dillman's Total Design Method on the Quality of the Responses. In W. E. Saris and I. N. Gallhofer (Eds.) *Sociometric Research. Vol. 2: Data Analysis.* (pp. 61–73). Springer. https://doi.org/10.1007/978-1-349-19054-6_3

Diaz de Rada, V. (2005). The Effect of Follow-up Mailings on The Response Rate and Response Quality in Mail Surveys. *Quality & Quantity*, *39*(1), 1–18. https://doi.org/10.1007/s11135-004-5950-5

Esser, H. (2011). Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften? In G. Albert (Ed.) *Kölner Zeitschrift für Soziologie und Sozialpsychologie/Sonderhefte*, *50* (pp. 45–62). VS Verlag für Sozialwissenschaften.

Fricker, S., & Tourangeau, R. (2010). Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys. *Public Opinion Quarterly*, *74*(5), 934–955. https://doi.org/10.1093/poq/nfq064

Friedman, E. M., Clusen, N. A., & Hartzell, M. (2003). Better Late? Characteristics of Late Respondents to a Health Care Survey. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 992–98.

Glaser, P. (2012). Respondents cooperation: Demographic profile of survey respondents and its implication. *Handbook of survey methodology for the social sciences*, 195-207. https://doi.org/10.1007/978-1-4614-3876-2_12

Green, K. E. (1991). Reluctant Respondents: Differences between Early, Late, and Nonresponders to a Mail Survey. *The Journal of Experimental Education*, *59*(3), 268–276. https://doi.org/10.1080/00220973.1991.10806566

Groves, R. M. (2005). *Survey errors and survey costs*. John Wiley & Sons.

Groves, R. M., Floyd, J. F., Couper, M. P., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed). John Wiley & Sons.

Gummer, T., & Struminskaya, B. (2020). Early and Late Participation during the Field Period: Response Timing in a Mixed-Mode Probability-Based Panel Survey. *Sociological Methods & Research*, *0*(0), 1-24. https://doi.org/10.1177/0049124120914921

Helasoja, V. (2002). Late response and item nonresponse in the Finbalt Health Monitor Survey. *The European Journal of Public Health*, *12*(2), 117–123. https://doi.org/10.1093/eurpub/12.2.117

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opinion Quarterly*, *74*(5), 956–984. https://doi.org/10.1093/poq/nfq062

Kreuter, F., Müller, G., & Trappmann, M. (2014). A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents. *Sociological Methods & Research*, *43*(3), 452–464. https://doi.org/10.1177/0049124113508094

Kroneberg, C. (2014). Frames, scripts, and variable rationality: An integrative theory of action. In G. Manzo (Ed.) *Analytical Sociology: Actions and networks* (pp. 97-123). John Wiley & Sons. https://doi.org/10.1002/9781118762707.ch04

Kunz, F. (2010). Follow-Ups in Mail Surveys. Empirical Findings on the Effects on the Response Rate, the Responses and the Sample Composition. *methods, data, analyses*, 4(2), 127-155. https://doi.org/10.12758/mda.2010.009.

Lemay, M. (2009). *Understanding the mechanism of panel attrition*. University of Maryland, College Park.

Loosveldt, G., & Billiet, J. (2002). Item nonresponse as a predictor of unit nonresponse in a panel survey. *Journal of Official Statistics, 18*(4), 545-557.

Lugtig, P. (2014). Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers. *Sociological Methods & Research*, *43*(4), 699–723. https://doi.org/10.1177/0049124113520305

Lynn, P. (2009). *Methodology of longitudinal surveys*. John Wiley & Sons.

Olson, K. (2013). Do non-response follow-ups improve or reduce data quality?: A review of the existing literature: *Do Non-response Follow-ups Improve or Reduce Data Quality? Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 129–145. https://doi.org/10.1111/j.1467-985X.2012.01042.x

Preisendörfer, P., & Wolter, F. (2014). Who Is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys. *Public Opinion Quarterly*, *78*(1), 126–146. https://doi.org/10.1093/poq/nft079

Roßmann, J., & Gummer, T. (2016). Using Paradata to Predict and Correct for Panel Attrition. *Social Science Computer Review*, *34*(3), 312–332. https://doi.org/10.1177/0894439315587258

Schoenman, J. A., Berk, M. L., Feldman, J. J., & Singer, A. (2003). Impact Of Differential Response Rates On The Quality Of Data Collected In The CTS Physician Survey. *Evaluation & the Health Professions*, *26*(1), 23–42. https://doi.org/10.1177/0163278702250077

Skarbek-Kozietulska, A., Preisendörfer, P., & Wolter, F. (2012). Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. *Zeitschrift für Soziologie*, *41*(1), 5-23. https://doi.org/10.1515/zfsoz-2012-0103

Sobal, J., & Ferentz, K. S. (1989). Comparing physicians’ responses to the first and second mailings of a questionnaire. *Evaluation & the Health Professions*, *12*(3), 329–339. https://doi.org/10.1177/016327878901200030

Tancreto, J. G., & Bentley, M. (2005). Determining the effectiveness of multiple nonresponse followup contact attempts on response and data quality. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 3626-3632.

Voigt, L. F., Koepsell, T. D., & Daling, J. R. (2003). Characteristics of telephone survey respondents according to willingness to participate. *American Journal of Epidemiology*, *157*(1), 66–73. https://doi.org/10.1093/aje/kwf185

2. Examining the relationship between participation time and multiple data quality indicators

2.1. Abstract

Assessing and verifying data quality is highly relevant for survey infrastructures. A particularly interesting predictor of data quality is participation time – that is, the time it takes a respondent to return a self-administered survey. Participation time has been shown to be related to data quality, but data quality is usually measured using single indicators only. In the present study, we estimated the relationships between participation time and 16 data quality indicators, including item nonresponse, responses to specific categories, motivated misreporting, acquiescence, open responses, nondifferentiation, anchoring in looping questions, response inconsistency, and survey duration. We found statistically significant relationships between participation time and eight of our data quality indicators. Moreover, these relationships were found to vary in terms of their functional form. We thus advise using multiple data quality indicators when investigating data quality.

2.2. Introduction

Good data quality is necessary for drawing correct conclusions from survey data (Dillman et al., 2016; Groves et al., 2009). Research findings can have implications that lead to changes within society, but if data is biased, the implications drawn from data analyses may be incorrect. Hence, maintaining good data quality is essential in all data collections. Researchers therefore take great steps to ensure that the quality of their data is as good as possible, and they assess data quality using data quality indicators, such as item nonresponse and straightlining.

As the quality of data is a highly important topic, many researchers have predicted data quality using a large number of data quality predictors, such as demographic characteristics (Reimers et al., 2022) and survey design features (Millar & Dillman, 2012). In addition to the vast body of research on predictors that must be assessed during surveys, previous studies have used participation time as a data quality predictor (Friedman et al., 2003; Helasoja, 2002; Kaminska et al., 2010; Kunz, 2010; Preisendörfer & Wolter, 2014; Skarbek-Kozietulska et al., 2012). Participation time describes the time it takes a respondent to return a self-administered survey to the survey agency. For web interviews, participation time is the time it takes a respondent to submit the last question they answer, regardless as to whether this question is also the last question of the questionnaire. For mail questionnaires, participation time is the time it takes for the return envelope to be delivered. Participation time has some advantages over other predictors of data quality: First, respondents do not directly report their participation time, and they are therefore less likely to falsify it. Second, participation time is routinely available to web surveys and is easily operationalized. Third, as a form of paradata, participation time neither requires additional survey time nor increases the survey burden for respondents. Fourth, participation time is comparable across surveys, whereas content variables often differ in their wording or in the scales they use. Overall, predicting data quality using participation time as a proxy does not require data quality indicators to be measured and additionally confers many advantages.

However, previous studies on participation time as a predictor of data quality have mostly been limited to single data quality indicators. Indeed, a distinct analysis of the relationship between participation time and multiple data quality indicators remains lacking. Without multiple indicators contrasted against one another, it is not possible to make clear determinations as to whether participation time is related to data quality in general or only to selective indicators of data quality. Hence, without comparing the relationships between participation time and data quality indicators, we cannot determine with certainty whether participation time can be used as a proxy for data quality.

In the present study, we investigated the relationship between participation time and data quality using multiple data quality indicators – namely (i) item nonresponse, (ii) the use of the "don't know" category, (iii) motivated misreporting, (iv) acquiescence, (v) choosing the first category, (vi) choosing the middle category, (vii) choosing the last category, (viii) choosing the first, middle, or last category, (ix) responding to open questions, (x) the length of responses to open questions, (xi) non-differentiation measured by the probability of differentiation, (xii) non-differentiation measured by the coefficient of variation, (xiii) straightlining, (xiv) the duration of the survey for online respondents, (xv) anchoring in looping questions, and (xvi) inconsistency between responses. In so doing, we aimed to answer two research questions: (1) Is there a relationship between participation time and data quality? (2) If a relationship is present, can it be found for every data quality indicator, or only for a subset of indicators?

2.3. Background

In this section, we provide theoretical argumentation as to why participation time can be associated with our data quality indicators, and we subsequently present findings from previous studies about participation time and the respective data quality indicators. In so doing, we draw on the theoretical framework of satisficing (Krosnick, 1991), which is frequently used to explain data quality in surveys.

Tourangeau et al. (2000) argue that a respondent proceeds through four stages of cognitive processing when answering a survey question: (i) question comprehension, (ii) information retrieval, (iii) judgment, and (iv) response. Question comprehension refers to reading or listening to a survey question and interpreting its meaning. With information retrieval, a respondent thinks about the information that is relevant for answering the question. In this step, the respondent finds a response to the question. In Tourangeau et al.'s (2000) stage of judgement, the respondent evaluates their response and asks themself whether they can report their answer or whether this answer is truly correct. In the final step, the respondent must mentally adapt the information to the given response format. If all steps are carefully fulfilled, the respondent should report the optimal response (Krosnick, 1991).

However, respondents do not necessarily proceed in this order. They may repeat or even skip certain steps. Thus, when respondents are not motivated to report the correct answer or do not find the correct response immediately, they may not put as much effort into the response process. Failing to handle every step with due diligence – so-called "satisficing" – may lead to poor data quality. Satisficing can occur in a number of ways: Respondents may proceed through the steps less diligently than they would if they wished to respond optimally, or they may even completely skip some or all of the steps.

There are three main reasons for satisficing: a lack of cognitive ability, an excess of task difficulty, or a lack of motivation (Krosnick, 1991). The lack of motivation is also a reason for late participation (Minderop & Weiß, 2022). Hence, participation time and satisficing may be associated with each other. Kaminska et al. (2010) built one satisficing indicator from multiple forms of satisficing and investigated the relationship between this indicator and participation time, but the authors did not find evidence that greater participation time leads to a greater likelihood of satisficing. Contrary to Kaminska et al. (2010), we investigated the relationships between participation time on the one hand and the individual data quality indicators on the other hand. We hypothesized that respondents who return a survey late should be less engaged in the survey, which should increase their probability of satisficing (Krosnick, 1991). In the following sections, we introduce the data quality indicators used in this study and point out their relationship to participation time as found in previous studies.

2.3.1. Item nonresponse

Item nonresponse means that respondents do not answer a given question. This behavior may result from several processes (De Leeuw et al., 2003; Dillman, 2002; Krosnick, 1999) and can be seen as a form of satisficing. When respondents lack motivation, they may decide to participate later and engage in greater item nonresponse. Empirical studies have found mixed results regarding the possible relationship between participation time and item nonresponse. Many studies have found significantly more item nonresponse among late respondents (Armenakis & Lett, 1982; Fricker & Tourangeau, 2010; Preisendörfer & Wolter, 2014; Skarbek-Kozietulska et al., 2012)(Fricker & Tourangeau, 2010; Friedman et al., 2003; Kunz, 2010; Tancreto & Bentley, 2005), but many others have failed to find a statistically significant relationship at all (Diaz de Rada, 2005; Green, 1991; Helasoja, 2002; Schoenman et al., 2003; Sobal & Ferentz, 1989). Overall, we expected that respondents who return a survey later should engage in item nonresponse with more items.

2.3.2. Specific Categories

In contrast to item nonresponse, choosing specific categories means that respondents report answers to questions, but the reported answers might be incorrect. For instance, respondents might choose the “don’t know” category – even though this category does not correspond to their true opinion – simply because it is easier to report any response than to engage in the cognitive processing required to produce an answer. Other categories for which this behavior is possible are middle and extreme categories. When investigating nonsubstantive categories such as “don’t know” or “does not apply,” Friedman et al. (2003) observed significantly more such responses among late respondents in two out of three surveys. We expected that late respondents should be more likely than earlier respondents to provide responses that indicate satisficing.

2.3.3. Motivated misreporting

During survey participation, respondents may learn that if they respond to a question in a specific way, researchers may ask additional questions or give the respondents the opportunity to skip further questions. Respondents can then use this knowledge to shorten the survey duration and respond with a false answer in order to avoid further questions. In so doing, the respondents know the correct answer to the question but prefer to revise their answer in anticipation of the potentially forthcoming tasks. We expected this behavior to be more common among late respondents because they were presumed to be less motivated to participate than earlier respondents.

2.3.4. Acquiescence

Acquiescence is the tendency to agree with statements, regardless of their content (Krosnick, 1991). Agreement may be cognitively easier to handle than disagreement. Acquiescence means that there is a discrepancy between the true and the reported answer, and it is therefore a form of satisficing. Research on the relationship between participation time and acquiescence remains lacking. Nevertheless, we assumed that respondents who participate late should be less motivated than other respondents and should hence be more likely to satisfice and use acquiescence.

2.3.5. Open responses

Two features of open responses are often analyzed when it comes to data quality: (i) whether respondents actually respond to open questions and (ii) if so, how much information they provide. Not responding to open questions is an indicator of lower data quality, whereas long responses reflect higher data quality due to the greater effort required. Respondents should ideally take the time to respond in as much detail as necessary, as is usually done in longer responses. Shorter responses do not always reflect the respondent's complete opinion to the researchers. Green (1991) reported that late respondents are less likely to respond to open questions. As we expected late respondents to be less motivated, we posited that they should be more likely to not respond to open questions and to provide shorter answers.

2.3.6. Non-differentiation

Non-differentiation means that respondents tend to use the same or similar response categories in their responses to multiple items (Krosnick & Alwin, 1988). Due to the low variability in responses, researchers have reason to assume that the given answers may be incorrect. Typically, this could be due to the indication that respondents have applied response patterns in grid questions, such as switching between two categories. Non-differentiation results when respondents fail to invest the necessary mental resources in their responses to a survey. An extreme form of non-differentiation is straightlining, in which respondents only use one response category. Although the empirical reality indicates that late respondents are not more likely to straightline than earlier respondents (De Leeuw & Hox, 1988), in the light of the theoretical background, we expected greater rates of nondifferentiation and straightlining among respondents with longer participation times.

2.3.7. Anchoring in looping questions

Surveys often include questions that are related to one another, which may cause some respondents to use the response to one of the questions to produce a response to a different question. Anchoring is the phenomenon that occurs when one value influences the estimation of a different value (Kahneman et al., 1937). This could, for instance, be the case when a question is asked in multiple contexts. One example would be a question about how thoroughly a respondent listens to multiple people. A respondent may remember their previous response (i.e., how much they listen to their mother) and – consciously or subconsciously – report the same response to the related question (i.e., how much they listen to their brother). When previous responses are repeated in questions with a similar context, we speak of anchoring in looping questions. Anchoring in looping questions in this context may result from the failed effort to differentiate between two distinct yet related questions. Instead of evaluating the new question, respondents may repeat the response to the previous question. We expected that late respondents should be more likely to repeat their response from the previous question because they should be more eager to finish the survey than to correctly respond to the remaining questions.

2.3.8. Inconsistent responses

It is possible for respondents to overreport or underreport behavior, for instance, if this behavior is desirable or undesirable (Preisendörfer & Wolter, 2014). Respondents consciously or subconsciously evaluate the risks and benefits of truthful responses and decide whether to respond truthfully depending on the evaluation outcome (Tourangeau et al., 2000). Empirical studies have found evidence that late respondents are more likely to respond less consistently. Indeed, Skarbek-Kozietulska et al. (2012) found that the likelihood of matching administrative data that were linked to the survey was lower among late respondents than among respondents who had returned their survey earlier. Moreover, both Fricker & Tourangeau (2010) and Armenakis & Lett (1982) found that responses from later respondents were less likely to be validated with responses to a second survey. Furthermore, Kreuter et al. (2014) and Kaminska et al. (2010) found that the last respondents are significantly more likely to underreport having received welfare benefits. However, late respondents were not found to provide less consistent responses to other questions than earlier respondents. We expected that late respondents might provide more inconsistent responses because they might feel that they have been pushed to participate by repeated reminders.

2.3.9. Duration

In the following section, we focus on interview duration. Respondents who rush through a survey are less likely to provide correct responses to the questions (Draisma & Dijkstra, 2004; Zhang & Conrad, 2014). Speeding is argued to be a result of omitted steps in the cognitive processing of the response process. When respondents fail to execute the entire response process, responses are more likely to deviate from each respondent's true response. However, respondents who take particularly long may be distracted or may engage in multiple tasks, which inhibits them from focusing their attention on the survey. Stocké (2004) argues that when respondents require a long time to respond to a question, they may be more likely to lie – an assumption that draws on the idea that respondents execute the response process multiple times until they find a response that they are comfortable reporting. Respondents who have already delayed their participation could be argued to also delay returning the survey and to have a particularly long survey duration. Alternatively, respondents who participate late could be argued to be more likely to speed through the survey because they feel guilty for their delay and want to compensate for it. We are aware that these expectations contradict each other. Due to lacking research on this topic, it was not possible to assess which mechanism should be stronger.

Overall, we expected to find lower data quality for respondents with greater participation time. In the light of our research questions, this meant that we did not expect to find differences in the relationships between participation time and data quality.

2.4. Data & Method

The analyses used here were based on data from the GESIS Panel, which is a German probability-based mixed-mode open panel (Bosnjak et al., 2018) that contains ca. 5,000 panelists who are at least 18 years old. The sample was initially recruited in 2013 and consisted of a random sample drawn from municipal population registers. It was refreshed in 2016, 2018, and 2020. Respondents from both cohorts participated in a recruitment interview and were subsequently invited to complete a self-administered welcome survey. After completing this welcome survey, the participants were regarded as regular panelists. American Association for Public Opinion Research (AAPOR) Response Rate 1 in the face-to-face recruitment interview was 36% for the initial recruitment cohort and 33% for the refreshment cohort. For the initial recruitment cohort, 80% of the recruited respondents participated in the welcome survey, and for the refreshment cohort, 81% participated in the welcome survey. Every two months, all panelists were invited to complete a GESIS Panel survey. Almost all 5,000 panelists participated: Around 65% of respondents participated online, and the other participants received a paper-and-pencil questionnaire together with a postal invitation. All respondents received a 5-EUR prepaid incentive. Online respondents additionally received an email with a web link to the survey and email reminders both one and two weeks after the field start. The GESIS Panel is open to researchers from all fields to submit questionnaire proposals, which leads to highly diverse survey topics. The surveys usually take ca. 20 to 30 minutes to complete and cover at least 4 different topics.

For the present study, we used the data published in the March 2020 GESIS Panel Version 36 (GESIS, 2020), which contained 38 waves (2013–2019) and the first three cohorts. With one small exception made in the survey evaluation, we restricted ourselves to the GESIS Panel wave "gb," which was conducted from April to May 2019. This wave contained data quality indicators that were not assessed in other waves. Below, we present the variables used in this analysis (see overview in Table 2.1).

Table 2.1. Variable description.

	Minimum	Maximum	Mean	Std. Dev.	n
Participation timing	0	55	9.38	9.54	4497
Participation timing, squared	0	3025	179.07	377.46	4497
Data quality indicators					
Item nonresponse	0	29	1.68	3.48	4497
Don't know category	0	30	1.47	2.70	4497
Motivated misreporting	0	1	0.17	0.37	3831
Acquiescence	0	4	1.24	1.16	4497
First category	2	104	36.75	8.81	4497
Middle category	2	66	20.88	9.36	4497
Last category	2	69	19.51	7.58	4497
One of first/middle/last category	2	142	77.14	12.42	4497
Open response	0	1	0.10	0.30	4494
Length open response	0	820	11.80	57.23	4494
Coefficient of Variation	0	1	0.32	0.09	4497
Probability of differentiation	2	80	50.87	9.00	4497
Straightlining	0	7	0.28	0.74	4497
Anchoring in loop questions	0	24	13.42	5.26	2925
Inconsistent responses	0	14.19	0.94	1.38	4497
Duration	2	61.02	21.89	9.39	3181
Control variables					
Low education	0	1	0.25	0.43	4497
Medium education	0	1	0.34	0.48	4497
High education	0	1	0.41	0.49	4497
Evaluation previous survey: interesting	1	5	3.87	0.88	4497
Evaluation previous survey: diverse	1	5	3.99	0.86	4497
Evaluation previous survey: important	1	5	3.54	0.88	4497
Evaluation previous survey: long	1	5	2.28	0.98	4497
Evaluation previous survey: difficult	1	5	1.90	0.83	4497
Evaluation previous survey: too personal	1	5	2.35	1.07	4497

Note: n= Number of cases, Std. Dev.= Standard Deviation

Table 2.1, continued. Variable description.

	Minimum	Maximum	Mean	Std. Dev.	n
Cohort=1	0	1	0.64	0.48	4497
Cohort=2	0	1	0.18	0.38	4497
Cohort=3	0	1	0.18	0.38	4497

Note: n= Number of cases, Std. Dev.= Standard Deviation

Item nonresponse was measured as the number of questions among all closed questions to which a respondent had not responded. The survey consisted of 187 questions in total. The variable was metric and ranged from 0 to 187.

When investigating responses to *specific categories*, we focused on the first, middle, last, and "don't know" categories. We counted how often respondents had used the first, middle, last, or "don't know" category. The variables were metric and ranged from 0 to 187 (first and last category), from 0 to 141 (middle category), and from 0 to 32 ("don't know" category). It was also possible for respondents to switch between the first, middle, and last category. Therefore, we built an indicator that summed up the use of all of these categories. This indicator was also metric and ranged from 0 to 187.

We assessed *motivated misreporting* by asking respondents to provide information in a more burdensome question format with 5 scale points and 8 items.[1] In total, respondents were asked questions about their listening style three times, with slight alterations to the question context each time. First, they were asked to describe how they listen to other people in general (first context) using a burdensome question. Subsequently, they were asked who they talk about politics with the most (second context). If they indicated that they talk about politics with someone, the burdensome question was repeated by asking respondents how they listen to the person they talked about politics with the most. Respondents were next asked who they talk about politics with the second-most (third context), and the burdensome question was repeated by asking respondents how they listen to the person they talk about politics with the second-most.

After having responded to the burdensome question twice, respondents may have anticipated having to answer the burdensome question in the third context and may therefore have falsely

[1] The question text is provided in the appendix.

indicated that they do not talk about politics with a second person. The variable was coded binarily: 0 means that the respondent indicated talking about politics with a second person, and 1 means that the respondent indicated not talking about politics with a second person. An indication that the respondent does not talk about politics with a second person was measured as motivated misreporting.

In order to measure *acquiescence*, we assessed one matrix in which respondents were asked how much they agreed with four statements on a scale ranging from 1 to 7, with 1 representing complete disagreement and 7 representing complete agreement. For each respondent, we counted how often they indicated agreement (6) or complete agreement (7). The variable ranged from 0 to 4.

One *open question* in the survey required a text entry. In this question, the survey agency asks for remarks on the survey. From this question, we derived two data quality indicators: The first indicated whether respondents had responded to this specific question. This indicator was binary. No response was coded as 0, and a response was coded as 1. As a second data quality indicator, we used the *length of the open response*. We counted the length of the response in characters. There were a few outliers in this variable. We used the 99% percentile as the upper threshold and re-coded higher values to this percentile. Hence, this variable was metric and ranged from 0 to 820.

We used two *non-differentiation* indicators: the *coefficient of variation* and the *probability of differentiation*. Both indicators were calculated using 7 question sets. We used all question sets with at least 5 scale points and at least 5 items. The probability of differentiation indicates the number of scale points that respondents used in a set of questions and usually ranges from 0 to almost 1, with 0 indicating that only one response category had been used and 1 indicating that all values had been used (Krosnick & Alwin, 1988; Roßmann et al., 2018). In that sense, a probability of differentiation of 0 would mean that a respondent had been completely non-differentiated because they had not varied their response categories. This indicator is particularly suited for detecting patterns in a response scheme, such as varying between two categories. For our analyses, we calculated the probability of differentiation for each of the 7 question sets and calculated the mean out of the seven probabilities for each respondent. We then multiplied the mean probability of differentiation by 100 for future analysis. The coefficient of variation measured the distance among the scale points that the respondents had used. For each respondent, we calculated the mean and standard deviation of the grid questions and divided the standard deviation by the mean (Roßmann et al., 2018). The coefficient of variation could be 0 or greater, with 0 indicating non-differentiation and larger values indicating higher variation in responses. This indicator measured whether respondents had often used neighboring categories.

Similarly, the *straightlining* indicator was measured using 7 matrix questions, all of which had at least 5 scale points and at least 5 items. Straightlining in this context meant not varying in the response category at all among a question set. We assessed the number of times respondents practiced straightlining. Hence, straightlining ranged from 0 to 7.

We measured *anchoring in looping questions* using the repeated question set described in the section on motivated misreporting. The question set contains eight questions and was repeated three times, with a slightly different context each time.[2] Hence, ideally, the responses should have been different. We counted the times the responses were exactly the same. The minimum of this variable was 0, and the maximum was 24.

We further measured *inconsistency* in responses by assessing six questions that were asked twice within the wave: once at the beginning, and once at the end of the survey. The second assessment was altered only in the response scale. There were three experimental groups with different scale pairs to account for scale effects.[3] Three types of scales were used: First, a fully verbalized horizontal 7-point scale; second, an endpoint-labeled horizontal 11-point scale; and third, an open text field in which respondents could respond with numbers between 0 and 100. Group one received three questions in the 7-point scale, which were repeated using the open question, and three questions using the open response, which were repeated using the 11-point scale. Group two received three questions with the open response, which were repeated using the 11-point scale, and three questions using the 11-point scale, which were repeated using the 7-point scale. Group three received three questions with the 11-point scale, which were repeated using the 7-point scale, and three questions using the 7-point scale, which were repeated using the open response. We standardized the responses and calculated the differences between the first and second responses to the same questions. Smaller deviations indicated that the responses were more similar, while higher deviations indicated that the responses were more different. For every respondent, we summed the differences of all six questions and multiplied them by

---

[2] The question text is provided in the appendix.
[3] The experimental groups were conceptualized in a different experiment that went beyond the bounds of the present analyses.

100 for the data analysis. The variable was metric, the minimum was 0, and the maximum was 1,419.

We measured *duration* in minutes, which was automatically collected by the survey software. This duration could be 0 or greater, with higher values indicating longer responses. We used the 1st percentile as the lower threshold and the 99th percentile as the upper threshold (Gummer & Roßmann, 2015). Lower values were re-coded to the 1st percentile threshold, while higher values were re-coded to the 99th percentile threshold. The minimum was 2, and the maximum was 61. This indicator was available for online respondents only.

For the operationalization of *participation time*, we calculated the days it took respondents to return the survey. For every respondent, we subtracted the date of the field start from the date of the return of the survey. For web respondents, the return date was measured by the survey software. For mail participants, we used the date when the survey arrived at the survey agency (in Germany, an individual's specific location within the country has no effect on the amount of time that it takes for postal mail to arrive). We further multiplied participation time with itself and created *participation time squared* in order to allow for nonlinear effects in the analyses.

Table 2.1 provides the range, the mean, and the standard deviation of the variables used in the analysis.

In order to analyze the relationship between participation time and data quality, we estimated 16 models. We used logistic, Poisson, and negative-binomial regression models depending on the distribution of the dependent variable. For binary dependent variables, we computed logistic regressions. For count variables, we tested the dispersion of the dependent variable. We computed negative binomial regressions to overdispersed dependent variables, and we computed Poisson regressions when the dependent variable was not overdispersed. Wherever the share of zeros was greater than 30%, we computed zero-inflated models. The assumptions of the respective models (Roback & Legler, 2021; Wooldridge, 2010; Yang & Berdine, 2015; Zeileis et al., 2008) were met. An overview of the used models is provided in Table 2. We allowed for a nonlinear effect of participation time by including a nonlinear effect, which was important because data quality may have been best or worst among respondents who had returned their survey around the middle of the field duration, especially because the GESIS Panel field duration is particularly long. Allowing for a nonlinear effect could have improved the fit to the data. Respondents who returned their surveys around the middle of the GESIS

Panel field duration returned the survey at a time that may be considered representative of early or late participation in other studies (i.e., after 15 days, whereas field duration in other studies may be 20 days). Mail respondents were only analyzed in 15 models because the duration of the survey was not available for them. For most of the models, we had 4,497 respondents, which was the same as the number of complete cases of the respective GESIS Panel wave. Motivated misreporting, anchoring in looping questions, and duration could only be analyzed with a reduced number of respondents. Motivated misreporting and anchoring in looping questions were assessed in multiple filters, which means that not all respondents were confronted with the relevant questions for us to be able to assess these indicators. Duration could only be measured for the online mode respondents; hence, mail-mode respondents are missing in this analysis.

In our regression models, we controlled for cohort, participation mode, education, and evaluation of the previous survey wave. The evaluation of the previous survey was measured in six categories (i.e., interesting, diverse, important for science, long, difficult, and overly personal). Controlling for the cohort accounted for previous survey experience – that is, the number of surveys to which a respondent had been invited. Controlling for the mode was important with regard to participation time because the participation process differed by mode (Bosnjak et al., 2018). Both variables were also expected to be related to data quality. Education was controlled for because it might have been related to unobserved measures, such as understanding of the survey questions or interest in the topic. Such unobserved measures could have been related to participation time as well as to data quality. We expected the evaluation of the previous wave to be associated with participation time as well as with data quality.

We further plotted the predicted values for all data quality indicators across different participation times. This graph allowed us to see whether the data quality indicators behaved similarly over participation times – that is, whether early respondents had provided similar data quality for all indicators or whether we could find differences in the functional forms of the relationships between participation time and the data quality indicators.

2.5. Results

Tables 2.2 presents the 16 dependent variables. For each dependent variable, we summarize the model that we used to explain the variance in this variable, the average marginal linear and nonlinear effects of participation time, the significance level and standard errors of these linear

Table 2.2. Regression models and AMEs of participation timing on 16 data quality indicators.

Dependent variable	Item nonresponse	Don't know category	Motivated misreporting	Acquiescence
Regression Model	Zero inflated poisson	Zero inflated poisson	Logistic	Zero infl. neg. bin.
Linear effect	0.00625 (0.01002)	0.02651** (0.00920)	0.00659*** (0.00181)	-0.02050*** (0.00528)
Nonlinear effect	-0.00005 (0.00026)	-0.00077** (0.00025)	-0.00015** (0.00005)	0.00029** (0.00011)
n	4497	4497	3831	4497

Note: *: p<0.05, **: p<0.01, ***: p<0.001, AME= Average marginal effect, Zero infl. neg. bin.= Zero inflated negative binomial, standard errors in parentheses

Table 2.2, continued. Regression models and AMEs of participation timing on 16 data quality indicators.

Dependent variable	First category	Middle category	Last category	First, middle or last category
Regression Model	Negative binomial	Negative binomial	Negative binomial	Negative binomial
Linear effect	-0.02336 (0.03653)	0.02479 (0.03098)	-0.06697 (0.03564)	-0.07179 (0.05163)
Nonlinear effect	-0.00069 (0.00092)	-0.00062 (0.00078)	0.00200* (0.00089)	0.00088 (0.00130)
n	4497	4497	4497	4497

Note: *: p<0.05, **: p<0.01, ***: p<0.001, AME= Average marginal effect, Zero infl. neg. bin.= Zero inflated negative binomial, standard errors in parentheses

Table 2.2, continued. Regression models and AMEs of participation timing on 16 data quality indicators.

Dependent variable	Open response	Legth open response	Coefficient of variation	Probability of differentiation
Regression Model	Logistic	Zero infl. neg. bin.	Negative binomial	Poisson
Linear effect	-0.00294** (0.00113)	-0.42163* (0.1812)	-0.00035 (0.00241)	0.01764 (0.03047)
Nonlinear effect	0.00006* (0.00003)	0.01088** (0.00353)	0.00001 (0.00006)	-0.00016 (0.00077)
n	4494	4494	4497	4497

Note: *: p<0.05, **: p<0.01, ***: p<0.001, AME= Average marginal effect, Zero infl. neg. bin.= Zero inflated negative binomial, standard errors in parentheses

Table 2.2, continued. Regression models and AMEs of participation timing on 16 data quality indicators.

Dependent variable	Straightlining	Anchoring in loop questions	Inconsistent responses	Duration
Regression Model	Zero infl. neg. bin.	Negative binomial	Negative binomial	Negative binomial
Linear effect	0.00228 (0.00407)	-0.06104* (0.02783)	-0.01073 (0.00631)	0.10147* (0.04061)
Nonlinear effect	-0.00005 (0.00012)	0.00125 (0.00069)	0.00013 (0.00016)	-0.00077 (0.00116)
n	4497	2925	4497	3181

Note: *: p<0.05, **: p<0.01, ***: p<0.001, AME= Average marginal effect, Zero infl. neg. bin.= Zero inflated negative binomial, standard errors in parentheses

and nonlinear effects, and the model's number of cases. For the zero-inflated model, we focus on the count model part, which is the mechanism of interest to us. The full analysis models are provided in the appendix.

Overall, the effects are fairly small, though some effects are statistically significant. We found an effect of participation time on the use of the "don't know" category: Late respondents were more likely to use the "don't know" category than earlier respondents at first but were less likely to do so later. We additionally found that late respondents were more likely to use motivated misreporting at first but were less likely to use motivated misreporting later. Participation time was also related to acquiescence: Late respondents were less likely to use acquiescence at first but were more likely than earlier respondents to use acquiescence later. Late respondents were additionally more likely to use the last category than were earlier respondents. Moreover, late respondents were less likely to respond to open questions at first but were more likely to respond to open questions later. Participation time was related to the length of open responses: At first, late respondents were less likely to provide longer responses than were earlier respondents, but those who responded even later were more likely to provide longer responses. We also found a negative effect of participation time on anchoring in looping questions: Late respondents were less prone to this behavior. Late respondents additionally had a longer survey duration than earlier respondents.

We did not find relationships between participation time and the use of the first category; the use of the middle category; the use of one of the first, middle, or last categories; the coefficient of variation; the probability of differentiation; straightlining; or inconsistencies between measurements even though we had expected to find that late respondents would perform poorer on these data quality indicators.

We visualized the predicted values of each model in Figure 2.1. Each plot shows the participation time in days on the x-axis and the predictions of the respective data quality indicators on the y-axis. The line represents the mean predictions per day. The shaded area around the line depicts the confidence intervals of the prediction.

Figure 2.1. Functional forms of the relationships of participation timing with data quality indicators

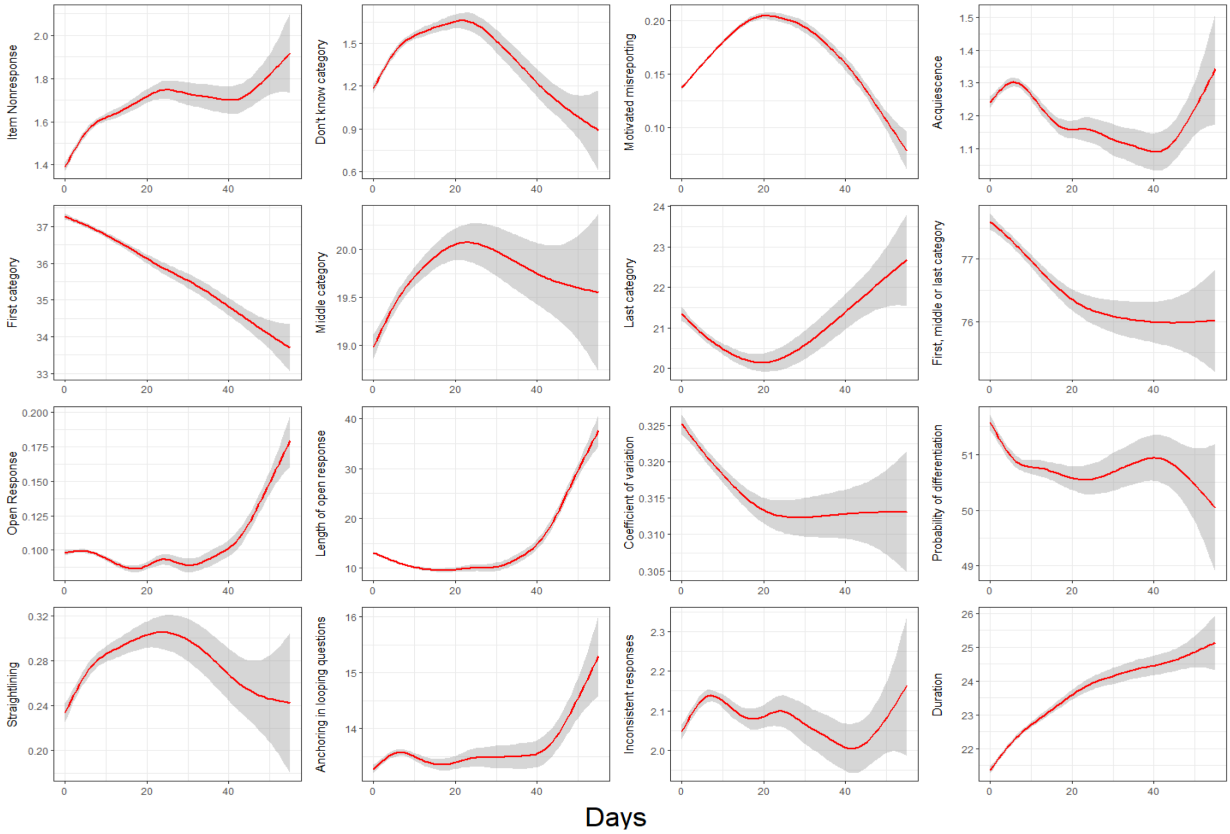


Figure 2.1 shows that the predictions of the “don’t know” category strongly increased until 20 days after the field start, at which point they decreased again. The decrease in the predictions of the “don’t know” category among the last respondents was not in line with our expectations. Similarly, the predictions of motivated misreporting slightly increased until 20 days after the field start, at which point motivated misreporting was predicted to decrease again. The latest respondents were predicted to use motivated misreporting less than the earliest respondents. We expected the first increase in motivated misreporting among late respondents, but the later decrease was not in line with our expectations. The predictions of acquiescence first increased

for a short time and later decreased over time. Around 40 days after the field start, the predictions of acquiescence increased again. We did not expect a nonlinear relationship, but the last outcome was in line with our expectations. The predictions of last category use decreased at first but increased again starting from 20 days after the field start. This finding was also in line with our expectations.

The predictions of responses to open questions displayed a different pattern: For the first respondents, the predictions of responding to the open questions remained stable. Around 30 days after the field start, the predictions strongly increased, which went against our expectations. A similar pattern was observed for the length of open responses: The predictions of the length of open responses remained stable until around 30 days after the field start, at which point they increased almost exponentially. This finding was not in line with our expectations. The predictions of anchoring in looping questions first mainly remained stable over time, but 40 days after the field start, the predictions of anchoring in looping questions strongly increased. Our expectation was that the increase would start earlier. The predictions of survey duration monotonously increased with increasing participation time. The functional form slightly flattened beginning 25 days after the field start. We expected this finding; however, the opposite was also expected.

Overall, we found that participation time was associated with data quality for some indicators, but not for all. Interestingly, the effects found did not all point in the same direction. Some indicators showed that late respondents had provided data of lesser quality. For other significant effects, we found that the latest respondents had provided data of better quality than had earlier respondents. We found differences not only in the directions of the findings, but also in the overall predictions based on participation time. The functional forms of the data quality predictions varied tremendously in terms of shape. We thus concluded that the multiple data quality indicators were not necessarily comparable with one another.

2.6. Conclusion & Discussion

The present study investigated the relationship between participation time and multiple data quality indicators. It was crucial to us to discover whether these relationships behave similarly or whether they are different. We found that the relationships between participation time and data quality indicators are indeed diverse. Participation time had a statistically significant relationship with some – albeit not all – data quality indicators. The latest respondents were less likely to use the “don’t know” category and to engage in motivated misreporting, but they were more likely to provide open responses, to provide longer open responses, to use acquiescence, to use the last category, to have higher anchoring in looping questions, and to have a longer survey duration. We further found a variety in the functional forms of the associations between participation time and data quality.

This variety in the functional forms means that the assumption that late respondents always provide poor data quality was not supported in our results. One reason for the discovered variance may be that the data quality indicators that were used here had not measured the same latent concept. There are two main reasons as to why we may not have measured the same latent concept with all data quality indicators: First, data quality may have been much more diverse than initially expected, and respondents who had provided poor data quality on one indicator may have provided good data quality on another indicator. Second, the data quality indicators we used may not all have measured data quality due to problems in both operationalization and conceptualization. In some cases, indicators of poor data quality may have measured not only poor data quality, but also usual behavior, for instance, when looking at first or last category responses.

The present study has several practical implications for survey methodologists and survey practitioners: First, this study revealed that participation time is related to several data quality indicators. Hence, participation time can partially be used as a predictor of data quality. This finding is critical given the easy usability of participation time as paradata. However, we also found that participation time is not related to all of the data quality indicators used in this study. For survey practitioners, we recommend collecting and evaluating respondents’ participation time. Survey methodologists could look into the reasons for these findings in greater detail in future studies. Second, we compared multiple data quality indicators and showed that their relationships with participation time were not necessarily similar. We thus concluded that data quality indicators cannot be used interchangeably. We therefore advise researchers who investigate data quality to investigate multiple indicators instead of just one in order to avoid incorrect conclusions. Third, we provided a deeper understanding of the group of late respondents. We revealed that they do not necessarily provide data of poor quality – as has been generally assumed – and that the relationship between participation time and data quality is in fact much more diverse. The finding that late respondents may even provide data of higher quality on some indicators has not yet been shown and suggests a new way of dealing with late respondents: Instead of generally aiming to prevent late response, practitioners should first consider which data quality indicators are important to them. Fourth, we showed that the relationship between participation time and data quality is not always linear. In fact, respondents with medium participation time were predicted to provide particularly poor data quality on some indicators. We thus advise researchers and practitioners to allow for nonlinearity when investigating participation time.

The present study is not without limitations. First, the study could only offer an overview of the relationship between participation time and data quality, and it lacked an in-depth analysis of the data quality indicators. Second, it was not possible for us to compare the models because they had different dependent variables. Future studies could thus thematize the fit of different models. Third, there are certainly more data quality indicators that should be examined in this context, and there are surely better ways to operationalize the data quality indicators that we used. However, using secondary data, we had to operationalize the data quality indicators from the data that were available to us, though it would have been desirable to have had more measurements of some data quality indicators, such as acquiescence. Fourth, our measures of data quality were mostly indirect. In some cases, our measures of low data quality could also reflect the true opinion of respondents. In the future, researchers should collect longitudinal primary data on data quality indicators and examine the causal relationship between participation time and data quality. Fifth, due to the long GESIS Panel field period, the number of cases with very high participation time was fairly low. We therefore found vast confidence intervals among later participants. We decided to keep all panelists in the analysis; however, it could have been wise to truncate the data or to use them with a shorter participation time.

2.7. References

Armenakis, A. A., & Lett, W. L. (1982). Sponsorship and follow-up effects on response quality of mail surveys. *Journal of Business Research*, *10*(2), 251–262. https://doi.org/10.1016/0148- 29638290031-5

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review*, *36*(1), 103–115. https://doi.org/10.1177/0894439317697949

De Leeuw, E. D., & Hox, J. J. (1988). Artifacts in Mail Surveys: The Influence of Dillman's Total Design Method on the Quality of the Responses. In W. E. Saris and I. N. Gallhofer (Eds.) *Sociometric Research. Vol. 2: Data Analysis.* (pp. 61–73). Springer. https://doi.org/10.1007/978-1-349-19054-6_3

De Leeuw, E., Hox, J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of official statistics*, *19*(2), 153–176.

Diaz de Rada, V. (2005). The Effect of Follow-up Mailings on The Response Rate and Response Quality in Mail Surveys. *Quality & Quantity*, *39*(1), 1–18. https://doi.org/10.1007/s11135-004-5950-5

Dillman, D., Eltinge, J., Groves, R., and Little, R. (2002). Survey nonresponse in design, data collection and analysis. In R. Groves, D. Dillman, J. Eltinge, and R. Little (Eds.), *Survey Nonresponse* (pp. 3–26). Wiley.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2016). *Internet, phone, mail and mixed-mode surveys: The tailored design method*, John Wiley & Sons.

Draisma, S., & Dijkstra, W. (2004). Response Latency and (Para)Linguistic Expressions as Indicators of Response Error. In S. Presser, J.M. Rothgeb, M.P. Couper, J,T. Lessler, E. Martin, J. Martin, E. Singer (Eds.) *Methods for Testing and Evaluating Survey Questionnaires* (pp. 131–147). John Wiley & Sons, Ltd. https://doi.org/10.1002/0471654728.ch7

Fricker, S., & Tourangeau, R. (2010). Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys. *Public Opinion Quarterly*, *74*(5), 934–955. https://doi.org/10.1093/poq/nfq064

Friedman, E. M., Clusen, N. A., & Hartzell, M. (2003). Better Late? Characteristics of Late Respondents to a Health Care Survey. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 992–98.

GESIS . (2020). *GESIS Panel - Extended Edition*. GESIS, Cologne. ZA5664 Datenfile Version 36.0.0. https://doi.org/10. 4232/1.13476

Green, K. E. (1991). Reluctant Respondents: Differences between Early, Late, and Nonresponders to a Mail Survey. *The Journal of Experimental Education*, *59*(3), 268–276. https://doi.org/10.1080/00220973.1991.10806566

Groves, R. M., Floyd, J. F., Couper, M. P., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed). John Wiley & Sons.

Gummer, T., & Roßmann, J. (2015). Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review*, *33*(2), 217–234. https://doi.org/10.1177/0894439314533479

Helasoja, V. (2002). Late response and item nonresponse in the Finbalt Health Monitor Survey. *The European Journal of Public Health*, *12*(2), 117–123. https://doi.org/10.1093/eurpub/12.2.117

Kahneman, D, Slovic, P. &. Tversky, A. (1937). *Judgement under Uncertainty: Heuristics and biases*. 37. Cambridge University Press.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opinion Quarterly*, *74*(5), 956–984. https://doi.org/10.1093/poq/nfq062

Kreuter, F., Müller, G., & Trappmann, M. (2014). A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents. *Sociological Methods & Research*, *43*(3), 452–464. https://doi.org/10.1177/0049124113508094

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Krosnick, J. A. (1999). Survey research. *Annual review of psychology*, *50*(1), 537–567. https://doi.org/10.1146/annurev.psych.50.1.537

Krosnick, J. A., & Alwin, D. F. (1988). A Test of the Form-Resistant Correlation Hypothesis: Ratings, Rankings, and the Measurement of Values. *Public Opinion Quarterly*, *52*(4), 526-538. https://doi.org/10.1086/269128

Kunz, F. (2010). Follow-Ups in Mail Surveys. Empirical Findings on the Effects on the Response Rate, the Responses and the Sample Composition. *methods, data, analyses*, 4(2), 127-155. https://doi.org/10.12758/mda.2010.009.

Millar, M. M., & Dillman, D. A. (2012). Do Mail and Internet Surveys Produce Different Item Nonresponse Rates? An Experiment Using Random Mode Assignment. *Survey Practice*, *5*(2), 1-7. https://doi.org/10.29115/SP-2012-0011.

Minderop, I. L., & Weiß, B. (2022). Now, later, or never? Using response-time patterns to predict panel attrition. *International Journal of Social Research Methodology*, *0*(0), 1–14. https://doi.org/10.1080/13645579.2022.2091259

Preisendörfer, P., & Wolter, F. (2014). Who Is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys. *Public Opinion Quarterly*, *78*(1), 126–146. https://doi.org/10.1093/poq/nft079

Reimers, J. A., Turner, R. C., Crawford, B. L., Jozkowski, K. N., Lo, W.-J., & Keiffer, E. A. (2022). Demographic comparisons on data quality measures in web-based surveys. *Personality and Individual Differences*, *193*, 0-0. https://doi.org/10.1016/j.paid.2022.111612

Roback, P., & Legler, J. (2021). *Beyond multiple linear regression: Applied generalized linear models and multilevel models in R*. CRC.

Roßmann, J., Gummer, T., & Silber, H. (2018). Mitigating Satisficing in Cognitively Demanding Grid Questions: Evidence from Two Web-Based Experiments. *Journal of Survey Statistics and Methodology*, *6*(3), 376–400. https://doi.org/10.1093/jssam/smx020

Schoenman, J. A., Berk, M. L., Feldman, J. J., & Singer, A. (2003). Impact Of Differential Response Rates On The Quality Of Data Collected In The CTS Physician Survey. *Evaluation & the Health Professions*, *26*(1), 23–42. https://doi.org/10.1177/0163278702250077

Skarbek-Kozietulska, A., Preisendörfer, P., & Wolter, F. (2012). Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. *Zeitschrift für Soziologie, 41*(1), 5-23. https://doi.org/10.1515/zfsoz-2012-0103

Sobal, J., & Ferentz, K. S. (1989). Comparing physicians' responses to the first and second mailings of a questionnaire. *Evaluation & the Health Professions, 12*(3), 329–339. https://doi.org/10.1177/016327878901200030

Stocké, V. (2004). Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie*, *33*(4), 303–320. https://doi.org/10.1515/zfsoz-2004-0403

Tancreto, J. G., & Bentley, M. (2005). Determining the effectiveness of multiple nonresponse followup contact attempts on response and data quality. In *Proceedings of the Survey Research Methods Section of the American Statistical Association,* 3626-3632.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yang, S., & Berdine, G. (2015). The negative binomial regression. *The Southwest Respiratory and Critical Care Chronicles*, *3*(10), 50–53.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of statistical software*, *27*(8), 1–25. https://doi.org/10.18637/jss.v027.i08

Zhang, C., & Conrad, F. G. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127–135. https://doi.org/10.18148/srm/2014.v8i2.5453

2.8. Appendix

Table A.2.1. Coefficients and standard errors of the full models.

	Item nonresponse	Don't know category	Motivated misreporting	Acquiescence
Participation timing	0.012*** (0.004)	0.002 (0.004)	0.046*** (0.013)	-0.003 (0.005)
Participation timing, squared	-0.000*** (0.000)	-0.003*** (0.000)	-0.001*** (0.000)	0.000 (0.000)
Medium education	-0.144*** (0.034)	-0.276*** (0.036)	-0.005 (0.133)	-0.100** (0.044)
High education	-0.105*** (0.035)	-0.359*** (0.038)	0.037 (0.129)	-0.382*** (0.046)
Evaluation previous survey: interesting	-0.076*** (0.023)	-0.004 (0.024)	-0.073 (0.080)	0.024 (0.030)
Evaluation previous survey: diverse	-0.015 (0.023)	-0.091*** (0.024)	0.109 (0.079)	-0.064** (0.030)
Evaluation previous survey: important	0.050*** (0.017)	0.017 (0.019)	-0.177*** (0.059)	0.003 (0.022)
Evaluation previous survey: long	0.021 (0.016)	-0.014 (0.017)	0.045 (0.053)	-0.034 (0.021)
Evaluation previous survey: difficult	0.017 (0.019)	0.132*** (0.018)	0.066 (0.063)	-0.014 (0.024)
Evaluation previous survey: too personal	0.020 (0.014)	0.031** (0.015)	-0.001 (0.047)	0.012 (0.017)
Cohort=2	0.072** (0.032)	-0.193*** (0.037)	-0.065 (0.110)	-0.004 (0.043)
Cohort=3	0.098*** (0.032)	-0.213*** (0.038)	-0.207* (0.111)	-0.022 (0.043)
Mode=offline	0.205*** (0.033)	0.174*** (0.035)	-0.113 (0.119)	0.116*** (0.044)
Constant	1.109*** (0.096)	1.229*** (0.103)	-1.553*** (0.346)	0.812*** (0.126)
Observations	4,497	4,497	3,831	4,497
Regression	Zero inflated poisson	Zero inflated poisson	Logistic	Zero inflated neg. bin.
Log Likelihood	-9,424.543	-8,168.694	-1,701.602	-6,401.817
Akaike Inf. Crit.	18.905.09	14,352.82	3,431.204	12,861.63

Note: *p<0.1; **p<0.05; ***p<0.01, neg.bin= negative binomial, standard errors in parentheses

Table A.2.1, continued. Coefficients and standard errors of the full models.

	First category	Middle category	Last category	First, middle or last category
Participation timing	-0.001 (0.001)	0.001 (0.002)	-0.003* (0.002)	-0.001 (0.001)
Participation timing, squared	-0.000 (0.000)	-0.000 (0.000)	0.000** (0.000)	0.000 (0.000)
Medium education	-0.016 (0.010)	-0.007 (0.016)	-0.017 (0.018)	0.014** (0.007)
High education	-0.044*** (0.010)	-0.080*** (0.016)	-0.041** (0.018)	-0.031*** (0.007)
Evaluation previous survey: interesting	-0.012* (0.006)	-0.019* (0.010)	0.019* (0.011)	-0.005 (0.004)
Evaluation previous survey: diverse	0.001 (0.006)	-0.039*** (0.010)	0.033*** (0.011)	-0.001 (0.004)
Evaluation previous survey: important	-0.001 (0.005)	-0.029*** (0.007)	0.066*** (0.008)	0.011*** (0.003)
Evaluation previous survey: long	-0.008* (0.004)	0.009 (0.007)	-0.001 (0.008)	-0.002 (0.003)
Evaluation previous survey: difficult	-0.030*** (0.005)	0.050*** (0.008)	-0.098*** (0.009)	-0.028*** (0.003)
Evaluation previous survey: too personal	-0.015*** (0.004)	0.021*** (0.006)	-0.007 (0.007)	-0.005* (0.003)
Cohort = 2	0.011 (0.009)	-0.042*** (0.014)	0.057*** (0.016)	0.009 (0.006)
Cohort = 3	0.002 (0.009)	-0.046*** (0.014)	0.077*** (0.015)	0.009 (0.006)
Mode = offline	-0.022** (0.009)	0.017 (0.014)	0.007 (0.017)	-0.004 (0.006)
Constant	3.796*** (0.027)	3.182*** (0.042)	2.750*** (0.049)	4.417*** (0.018)
Observations	4,497	4,497	4,497	4,497
Regression	Negative binomial	Negative binomial	Negative binomial	Negative binomial
Log Likelihood	-15,949.170	-15,024.980	-15,831.140	-17,532.380
Akaike Inf. Crit.	31,926.350	30,077.960	31,690.280	35,092.760

Note: *p<0.1; **p<0.05; ***p<0.01, standard errors in parentheses

Table A.2.1, continued. Coefficients and standard errors of the full models.

	Open response	Length of open response	Coefficient of variation	Probability of differentiation
Participation timing	-0.038***	-0.014	-0.001	0.0003
	(0.015)	(0.016)	(0.008)	(0.001)
Participation timing, squared	0.001**	0.001*	0.00002	-0.00000
	(0.0003)	(0.0003)	(0.0002)	(0.00001)
Medium education	-0.071	0.290*	-0.017	0.017***
	(0.144)	(0.160)	(0.077)	(0.006)
High education	-0.031	0.546***	0.022	0.072***
	(0.143)	(0.161)	(0.076)	(0.006)
Evaluation previous survey: interesting	-0.051	0.115	0.001	0.008**
	(0.092)	(0.100)	(0.049)	(0.004)
Evaluation previous survey: diverse	0.056	0.028	0.033	0.029***
	(0.092)	(0.091)	(0.048)	(0.004)
Evaluation previous survey: important	0.064	-0.066	0.005	-0.008***
	(0.068)	(0.070)	(0.036)	(0.003)
Evaluation previous survey: long	0.132**	-0.002	-0.013	0.002
	(0.061)	(0.064)	(0.033)	(0.003)
Evaluation previous survey: difficult	0.046	0.024	-0.044	-0.012***
	(0.072)	(0.071)	(0.039)	(0.003)
Evaluation previous survey: too personal	0.032	-0.123**	-0.026	-0.001
	(0.055)	(0.061)	(0.029)	(0.002)
Cohort = 2	0.137	0.204	0.018	0.010*
	(0.129)	(0.138)	(0.066)	(0.005)
Cohort = 3	0.363***	0.246*	0.016	0.011**
	(0.121)	(0.128)	(0.065)	(0.005)
Mode = offline	0.511***	-0.085	-0.003	-0.015***
	(0.132)	(0.150)	(0.070)	(0.006)
Constant	-2.995***	5.137***	-1.123***	3.787***
	(0.399)	(0.403)	(0.209)	(0.017)
Observations	4,494	4,494	4,497	4,497
Regression	Logistic	Zero inflated neg. bin.	Negative binomial	Poisson
Log Likelihood	-1,405.578	-3,880.376	-2,594.940	-16,292.470
Akaike Inf. Crit.	2,839.156	7,818.75	5,217.880	32,612.940

Note: *p<0.1; **p<0.05; ***p<0.01, standard errors in parentheses

Table A.2.1, continued. Coefficients and standard errors of the full models.

	Straightlining	Memory effects	Inconsistent responses	Duration
Participation timing	0.009 (0.016)	-0.005** (0.002)	-0.005* (0.003)	0.005** (0.002)
Participation timing, squared	-0.0002 (0.0004)	0.0001* (0.0001)	0.0001 (0.0001)	-0.00004 (0.0001)
Medium education	0.027 (0.116)	-0.019 (0.021)	-0.203*** (0.028)	-0.084*** (0.022)
High education	-0.600*** (0.131)	-0.142*** (0.021)	-0.316*** (0.028)	-0.139*** (0.021)
Evaluation previous survey: interesting	0.044 (0.092)	-0.021 (0.013)	0.0003 (0.019)	0.021* (0.012)
Evaluation previous survey: diverse	-0.412*** (0.108)	-0.034** (0.013)	-0.032* (0.019)	-0.011 (0.012)
Evaluation previous survey: important	0.034 (0.086)	0.030*** (0.010)	0.021 (0.014)	0.011 (0.009)
Evaluation previous survey: long	0.109* (0.059)	-0.026*** (0.009)	-0.017 (0.013)	0.003 (0.008)
Evaluation previous survey: difficult	0.178** (0.072)	-0.006 (0.011)	0.014 (0.015)	0.016* (0.010)
Evaluation previous survey: too personal	-0.138** (0.058)	0.003 (0.008)	0.002 (0.011)	-0.001 (0.007)
Cohort = 2	0.009 (0.112)	0.021 (0.018)	-0.045* (0.026)	0.060*** (0.017)
Cohort = 3	-0.016 (0.129)	0.009 (0.018)	-0.001 (0.026)	0.126*** (0.017)
Mode = offline	0.036 (0.119)	0.065*** (0.019)	0.088*** (0.027)	
Constant	0.007 (0.307)	2.843*** (0.057)	1.027*** (0.080)	3.007*** (0.055)
Observations	4,497	2,925	4,497	3,181
Regression	Zero inflated neg. bin	Negative binomial	Negative binomial	Negative binomial
Log Likelihood	-2,765.626	-8,910.142	-7,160.668	-11,120.670
Akaike Inf. Crit.	5589.251	17,848.280	14,349.340	22,267.340

Note: *p<0.1; **p<0.05; ***p<0.01, standard errors in parentheses

Figure A.2.1. Questionnaire for anchoring in looping questions and motivated misreporting, page 1.

GesellschaftsMonitor
gesis

Nachfolgend finden Sie mehrere Aussagen, die beschreiben, welche Art von Zuhörer/Zuhörerin Sie sind.

(22) Bitte geben Sie für jede Aussage an, inwieweit diese auf Sie selbst zutrifft oder nicht zutrifft.
Bitte denken Sie dabei nicht an eine bestimmte Situation, sondern an Ihre allgemeine Art anderen Menschen im Gespräch zuzuhören, also daran wie Sie üblicherweise und in den meisten Situationen zuhören.

	trifft überhaupt nicht zu	trifft eher nicht zu	trifft teils zu, teils nicht zu	trifft eher zu	trifft voll und ganz zu
	1	2	3	4	5
Ich höre zu, um die Gefühle und die Stimmung der sprechenden Person zu verstehen.	O	O	O	O	O
Ich höre bis zum Ende zu, was eine Person zu sagen hat, und bilde mir erst dann eine Meinung.	O	O	O	O	O
Mir fallen oft unmittelbar Fehler auf, in dem was andere Personen sagen.	O	O	O	O	O
Wenn ich jemandem zuhöre, geht es mir immer auch darum herauszubekommen, wie er/sie sich fühlt.	O	O	O	O	O
Ich warte, bis alle Punkte genannt wurden, bevor ich mir ein Urteil oder eine Meinung bilde.	O	O	O	O	O
Ich werde ungeduldig, wenn Personen langatmig und ausschweifend erzählen.	O	O	O	O	O
Mir fällt es schwer Personen zuzuhören, die zu lange brauchen, um ihre Gedanken mitzuteilen.	O	O	O	O	O
Ich merke häufig, wenn andere Personen etwas Unlogisches erzählen.	O	O	O	O	O

Figure A.2.2. Questionnaire for anchoring in looping questions and motivated misreporting, page 2.

GesellschaftsMonitor
gesis

Nun wüssten wir gerne etwas über die beiden Personen, mit denen Sie sich in letzter Zeit am meisten über die Parteien und die Politik unterhalten haben. Dies können z.B. Familienmitglieder, Freunde/innen oder Arbeitskollegen/innen sein.

(23) Wenn Sie zunächst einmal an diejenige Person denken, mit der Sie sich in den letzten 7 Tagen am häufigsten über die Parteien und die Politik unterhalten haben: In welcher Beziehung steht diese Person zu Ihnen?

- O Ehepartner/in, Partner/in
- O Eltern
- O Kind
- O Andere/r Verwandte/r
- O Freund/in
- O Arbeitskollege/in
- O Nachbar/in
- O Vereins- oder Verbandskollege/in
- O Keines der Genannten/anderer Kontakt

- O *Trifft nicht zu / mit niemandem in der letzten Woche über Parteien und Politik unterhalten* → ***Bitte weiter mit Frage (33)!***

(24) An wie vielen der letzten 7 Tage haben Sie sich speziell mit dieser Person über die Parteien und die Politik unterhalten?

- O an 0 Tagen
- O an 1 Tag
- O an 2 Tagen
- O an 3 Tagen
- O an 4 Tagen
- O an 5 Tagen
- O an 6 Tagen
- O an 7 Tagen

Figure A.2.3. Questionnaire for anchoring in looping questions and motivated misreporting, page 3.

GesellschaftsMonitor
gesis

(25) Nachfolgend finden Sie mehrere Aussagen, die beschreiben, auf welche Art Sie speziell dieser Person im Gespräch zugehört haben. Bitte geben Sie für jede Aussage an, inwieweit diese zutrifft oder nicht zutrifft.

Bitte denken Sie dabei an die Situationen in den letzten 7 Tagen zurück, in denen Sie sich mit dieser Person über die Parteien und die Politik unterhalten haben.

	trifft überhaupt nicht zu 1	trifft eher nicht zu 2	trifft teils zu, teils nicht zu 3	trifft eher zu 4	trifft voll und ganz zu 5
Ich hörte zu, um die Gefühle und die Stimmung dieser Person zu verstehen.	O	O	O	O	O
Ich hörte bis zum Ende zu, was diese Person zu sagen hatte, und bildete mir erst dann eine Meinung.	O	O	O	O	O
Mir fielen oft unmittelbar Fehler auf, in dem was diese Person sagte.	O	O	O	O	O
Als ich dieser Person zuhörte, ging es mir auch darum herauszubekommen, wie er/sie sich fühlt.	O	O	O	O	O
Ich wartete, bis alle Punkte genannt wurden, bevor ich mir ein Urteil oder eine Meinung bildete.	O	O	O	O	O
Ich wurde ungeduldig, wenn diese Person langatmig und ausschweifend erzählte.	O	O	O	O	O
Mir fiel es schwer dieser Person zuzuhören, wenn sie zu lange brauchte, um ihre Gedanken mitzuteilen.	O	O	O	O	O
Ich merkte häufig, wenn diese Person etwas Unlogisches erzählte.	O	O	O	O	O

(26) Wie gut kennt sich diese Person Ihrer Meinung nach mit Politik aus?

O Sehr gut
O Gut
O Weniger gut
O Gar nicht

Figure A.2.4. Questionnaire for anchoring in looping questions and motivated misreporting, page 4.

GesellschaftsMonitor
gesis

(27) Wenn Sie sich mit dieser Person über politische Fragen unterhalten, wie oft sind Sie dann unterschiedlicher Meinung?

- O Immer
- O Oft
- O Selten
- O Nie

(28) Wenn Sie nun einmal an die Person denken, mit der Sie sich in den letzten 7 Tagen <u>am zweitmeisten</u> über die Parteien und die Politik unterhalten haben: In welcher Beziehung steht diese Person zu Ihnen?

- O Ehepartner/in, Partner/in
- O Eltern
- O Kind
- O Andere/r Verwandte/r
- O Freund/in
- O Arbeitskollege/in
- O Nachbar/in
- O Vereins- oder Verbandskollege/in
- O Keines der Genannten/anderer Kontakt

- O *Trifft nicht zu / mit niemand Weiterem in der letzten Woche über Parteien und Politik unterhalten* → ***Bitte weiter mit Frage (33)!***

(29) An wie vielen der letzten 7 Tage haben Sie sich speziell mit dieser Person über die Parteien und die Politik unterhalten?

- O an 0 Tagen
- O an 1 Tag
- O an 2 Tagen
- O an 3 Tagen
- O an 4 Tagen
- O an 5 Tagen
- O an 6 Tagen
- O an 7 Tagen

Figure A.2.5. Questionnaire for anchoring in looping questions and motivated misreporting, page 5.

GesellschaftsMonitor
gesis

(30) Nachfolgend finden Sie mehrere Aussagen, die beschreiben, auf welche Art Sie speziell dieser Person im Gespräch zugehört haben. Bitte geben Sie für jede Aussage an, inwieweit diese zutrifft oder nicht zutrifft.

Bitte denken Sie dabei an die Situationen in den letzten 7 Tagen zurück, in denen Sie sich mit dieser Person über die Parteien und die Politik unterhalten haben.

	trifft überhaupt nicht zu 1	trifft eher nicht zu 2	trifft teils zu, teils nicht zu 3	trifft eher zu 4	trifft voll und ganz zu 5
Ich hörte zu, um die Gefühle und die Stimmung dieser Person zu verstehen.	O	O	O	O	O
Ich hörte bis zum Ende zu, was diese Person zu sagen hatte, und bildete mir erst dann eine Meinung.	O	O	O	O	O
Mir fielen oft unmittelbar Fehler auf, in dem was diese Person sagte.	O	O	O	O	O
Als ich dieser Person zuhörte, ging es mir auch darum herauszubekommen, wie er/sie sich fühlt.	O	O	O	O	O
Ich wartete, bis alle Punkte genannt wurden, bevor ich mir ein Urteil oder eine Meinung bildete.	O	O	O	O	O
Ich wurde ungeduldig, wenn diese Person langatmig und ausschweifend erzählte.	O	O	O	O	O
Mir fiel es schwer dieser Person zuzuhören, wenn sie zu lange brauchte, um ihre Gedanken mitzuteilen.	O	O	O	O	O
Ich merkte häufig, wenn diese Person etwas Unlogisches erzählte.	O	O	O	O	O

(31) Wie gut kennt sich diese Person Ihrer Meinung nach mit Politik aus?

O Sehr gut
O Gut
O Weniger gut
O Gar nicht

Figure A.2.6. Questionnaire for anchoring in looping questions and motivated misreporting, page 6.

GesellschaftsMonitor
gesis

(32) Wenn Sie sich mit dieser Person über politische Fragen unterhalten, wie oft sind Sie dann unterschiedlicher Meinung?

- O Immer
- O Oft
- O Selten
- O Nie

3. Predicting panel attrition using multiple operationalisations of response time[4]

3.1. Abstract

Panel attrition is a major problem for panel survey infrastructures. When panelists attrit from a panel survey, the infrastructure is faced with (i) the costs of recruiting new respondents, (ii) a broken timeline of existing data, and (iii) potential nonresponse bias. Previous studies have shown that panel attrition can be predicted using respondents' response time. However, response time has been operationalised in multiple ways, such as (i) the number of days it takes respondents to participate, (ii) the number of contact attempts made by the data collection organisation, and (iii) the proportion of respondents who have participated prior to a given respondent. Due to the different operationalisations of response time, it is challenging to identify the best measurement to use for predicting panel attrition. In the present study, we used data from the GESIS Panel – which is a German probability-based mixed-mode (i.e., web and mail) panel survey – to compare different operationalisations of response time using multiple logistic random-effects models. We found both that the different operationalisations have similar relationships to attrition and that our models correctly predict a similar amount of attrition.

3.2. Introduction

Keeping respondents from attriting is a fundamental challenge for panel surveys and is particularly important for three main reasons: First, respondents who attrit might be different from respondents who stay in a panel survey. When a particular group of respondents stops participating, the survey data can become biased if those who stay and those who have attrited differ on key outcome variables (Bethlehem, 2002; Groves, 2006; Lynn & Lugtig, 2017). Second, recruiting new respondents is costly, for example, with regard to sampling or securing the cooperation of new respondents (Groves, 2005). The expenses incurred when recruiting new respondents include the additional work that researchers conduct in planning and

[4] This chapter has been published in the journal "Survey Methods: Insights from the Field". In the published article, the term "response time" has been used to describe what has previously been called "participation time" in this dissertation. In the published article, I used British English spelling. In this chapter, I maintain this choice for reasons of similarity to the published article.

implementing the recruitment, drawing a sample, contacting and incentivising potential respondents, and monitoring new panelists. Measures for keeping respondents in the sample can be more cost-effective than is recruiting new panelists. Third, the value of panelists' data increases with every panel wave in which the same respondents participate because more information provides greater potential for analysis. In other words, the longer a respondent remains in a panel, the longer the period of time for which respondent change can be analysed.

As panel attrition is a highly relevant problem, it is important to learn more about the determinants of this attrition. Among all possible correlates of attrition, response time is particularly interesting because as a form of paradata, it is available for almost every survey and is easy to measure. Paradata are data that are collected as by-products of collecting survey data (Kreuter, 2013). For self-administered surveys, response time describes the time it takes a respondent to return a survey to the survey agency. More specifically, for web surveys, response time is the time a respondent takes to submit the last answer to a question. For mail surveys, response time is the time it takes the respondent to return the questionnaire. Finally, for computer-administered surveys, response time is usually available, does not increase the response burden, and is not prone to measurement error (Roßmann & Gummer, 2016). Existing studies on the association between response time and panel attrition have found evidence that late response time may be associated with a higher probability of attriting (Cohen et al., 2000). Thus, by using response time, online panel data collection organisations could potentially identify future attrition before participants stop participating. This knowledge would enable panels to intervene prior to a potential unit nonresponse or panel attrition and thus also to prevent data loss.

Although predicting attrition based on response time could enable panel infrastructures to intervene before respondents attrit, research on response time is limited. Despite some exceptions (e.g., Cohen et al., 2000), little evidence exists as to how response time is related to panel attrition. This dearth of information is aggravated by the fact that response time can be measured in several ways, and we currently lack a common operationalisation of response time (Kennickell, 1999). For example, response time can be operationalised as a metric variable (e.g., Gummer & Struminskaya, 2020; Skarbek-Kozietulska et al., 2012), as a binary variable (e.g., Voigt et al., 2003), or as a categorical variable (e.g., Kreuter et al., 2014). Moreover, response time does not have to refer to the actual time in days; rather, it can also be measured as a cumulative sample size (e.g., 25%) that indicates the proportion of respondents in a wave who participated prior to a given respondent. If different operationalisations of response time are applied in different studies, the results are thus not necessarily comparable.

In the present study, we address this research gap and compare the different methods used to operationalise response time. In so doing, we aim to answer the following two research questions: (i) How are the different operationalisations of response time related to panel attrition? (ii) Do models with different operationalisations predict attrition similarly well, or are there especially good or bad operationalisations?

In order to answer these questions, we first provide an overview as to how response time has been operationalised in previous studies and then present the results of these studies. Second, we operationalise response time in different ways, and third, we predict attrition using multiple operationalisations of response time in separate random-effects logistic panel regressions. In order to find the operationalisation of response time that best predicts attrition, we compare models with respect to their ability to correctly predict attrition. Each model includes one operationalisation of response time. Based on these calculations, we provide recommendations as to which operationalisation of response time should be used in future studies.

3.3. Background

In investigating the reasons for a possible relationship between response time and panel attrition, we drew on theoretical expectations that describe the relationship between response time and measurement quality. One theoretical expectation was that respondents who participate early in the field would generally be more highly motivated than later respondents and would therefore also be more likely to provide accurate data (Bollinger & David, 2001). Similarly, the greater motivation of earlier respondents was thought to potentially be related to their lower likelihood of attrition. A second expectation was that respondent characteristics or values would be connected to both response time and measurement quality (Olson, 2013). Respondents who have less time were assumed to speed through the survey and to perhaps not think about every question with due diligence. Similarly, respondents with less time were assumed to stop participating in surveys entirely and thus to attrit. When it comes to reminders, a third theoretical expectation was that the additional attention that late respondents receive could make them feel pressured to participate. This pressure was thought to potentially lead to a negative attitude towards the survey, which would be reflected in a lower measurement for quality (Brehm & Brehm, 2013). If panelists felt pressured to participate, this pressure was assumed to potentially lead to their attrition. Overall, we expected that later participants would be more likely to attrit from a panel in future waves.

In the following sections, we evaluate possible operationalisations of response time. Table 3.1 summarises examples of the operationalisations of response time taken from the literature.

Table 3.1. Summary of the operationalisations of response time, and examples of references to studies in which these operationalisations have been applied.

Operationalisation	Exemplary references
Days: metric operationalisation	Gummer & Struminskaya, 2020; Kennickell, 1999; Preisendörfer & Wolter, 2014; Schoenman et al., 2003; Skarbek-Kozietulska et al., 2012; Struminskaya & Gummer, 2021; Yan et al., 2004
Days: binary operationalisation	Bates & Creighton, 2000; Cohen et al., 2000; Cohen & Carlson, 1995; Friedman et al., 2003; Kay et al., 2001; Voigt et al., 2003
Number of contacts	Armenakis & Lett, 1982; Cannell & Fowler, 1963; Curtin et al., 2000; de Leeuw & Hox, 1988; Diaz de Rada, 2005; Donald, 1960; Eckland, 1965; Gilbert et al., 1992; Green, 1991; Helasoja, 2002; Kennickell, 1999; Korkeila et al., 2001; Kreuter et al., 2014; Kunz, 2010; Lin & Schaeffer, 1995; Schoenman et al., 2003; Treat & Stackhouse, 2002
Respondents: metric operationalisation	
Respondents: binary operationalisation	Bates & Creighton, 2000

Days: metric operationalisation: Response time can be a metric variable that measures the days until a survey is returned. For this metric operationalisation, each day of delay is an increase in the measurement of response time (e.g., Gummer & Struminskaya, 2020; Kennickell, 1999; Preisendörfer & Wolter, 2014; Skarbek-Kozietulska et al., 2012; Struminskaya & Gummer, 2021). One advantage of this operationalisation is that it offers a richer potential for analysis than does a binary or categorical operationalisation, and it additionally enables a detailed description of respondents' behaviour. However, one disadvantage of this operationalisation is that it does not come with a built-in threshold for distinguishing "late respondents" from those who participated earlier. This lack of a threshold contrasts with binary or categorial operationalisations, which per se go hand in hand with grouping respondents. Based on the theoretical assumptions presented above as well as the extant research both on this operationalisation and on other data-quality indicators, we expected that longer response time would be related to a higher probability of attrition.

Days: binary operationalisation: It is relatively common in research on response time to choose a threshold for separating early and late respondents. In this case, a response before a given day is treated as an early response, and a response after this day is treated as a late response (Friedman et al., 2003). One common threshold is the date on which a reminder is sent (Cohen et al., 2000; Cohen & Carlson, 1995; Kay et al., 2001). A binary operationalisation of response time requires thoughts about when to set a threshold. Example thresholds could be one or two weeks after the field start, or even after the first day. Such an early threshold would separate particularly motivated respondents from other respondents. Therefore, setting a threshold of after the first day would identify respondents who are more motivated and thereby least likely to attrit rather than identifying respondents who have lower motivation than others. One advantage of this operationalisation is that it distinguishes "late respondents" from those who have participated earlier. However, one disadvantage of this operationalisation is that thresholds vary across studies, which means that response time that is operationalised using a threshold is often not comparable. Based on the aforementioned theoretical assumptions as well as the extant research both on this operationalisation and on other data-quality indicators, we expected that the category that is related to longer response time would have a higher attrition rate.

Number of contacts: Response time can be measured using the number of contacts or reminders (e.g., 1 vs. ~3) (Curtin et al., 2000; Gilbert et al., 1992; Green, 1991; Helasoja, 2002; Kennickell, 1999; Kreuter et al., 2014; Kunz, 2010; Lin & Schaeffer, 1995). Usually, returns per day are highest around field start and decrease over time. Additional contacts – such as reminders – typically lead to an increase in returns per day, which is then followed by a decrease. Hence, operationalising response time by using the number of contacts separates respondents who reply after each contact. Advantages of using the number of contacts as the operationalisation of response time include the vast body of available research that has used this operationalisation as well as its design-orientation because contacts are an investment of resources that are used to achieve respondent participation. However, one disadvantage of using the number of contacts as the operationalisation of response time is that we do not know whether respondents who were reminded would have responded without this reminder at a later time. In fact, this operationalisation assumes that this would not be the case. Based on the aforementioned theoretical assumptions we expected that respondents who are contacted more often would be more likely to attrit.

Respondents: metric operationalisation: Response time can be operationalised not only by drawing on the number of days or the number of contacts, but also by taking into account respondents' return order. For example, it would be possible to measure how many other respondents had returned a survey prior to a given respondent's return date for the survey. Such a metric operationalisation would analyse respondents' behaviour relative to the behaviour of other respondents. One advantage of this operationalisation strategy is that it offers the potential to examine respondents who participate early in the field time but later than most other respondents. While these former respondents would be classified as earlier respondents using other operationalisations, they would be classified as later respondents here. As we do not know whether these respondents behave more like early or late respondents, it was important to take this operationalisation into account, which may be especially relevant for surveys with a short field time. One disadvantage of this operationalisation of response time is that not much research has been conducted on it thus far. Based on the aforementioned theoretical assumptions, we expected that respondents who participate later than other respondents would be more likely to attrit.

Respondents: binary operationalisation: It is also possible to distinguish between first or last respondents on the one hand and other respondents on the other hand by using a fixed number, such as the first or last 100 respondents, or a percentage, such as the first or last 10% of respondents, as was done in Bates and Creighton's (2000) study. From an analytical point of view, the advantage of this operationalisation is that it offers the ability to classify respondents into smaller groups from among all respondents who participated on a given day. Moreover, survey practitioners may want to select a group of a precise size for potential targeted procedures. This operationalisation is the only one that enables researchers to choose the size of the group whose members should be regarded as early or late respondents. However, one disadvantage of using this operationalisation is that a respondent's response time depends empirically on other respondents' response time, but this is not actually the case in real life. Drawing on the above-mentioned theoretical background, we expected that respondents who participate among the first group would be less likely to attrit than would later respondents.

3.4. Data and methods

In order to investigate the way in which different operationalisations predict attrition, we used data from the GESIS Panel, which is a probability-based mixed-mode (i.e., web and mail) panel survey that collects data every two to three months (Bosnjak et al., 2018). The GESIS Panel was recruited in 2013 and covers respondents who were sampled from the municipal registries of the general population of Germany. Refreshment samples were drawn in 2016, 2018, and 2021.

Before becoming a regular panelist, respondents participated in a face-to-face recruitment interview. The AAPOR Response Rate 1 in the recruitment interview was 35.5% for the first cohort (i.e., the 2013 recruitment) and 33.2% for the second cohort (i.e., the 2016 refreshment). The third and fourth cohorts (i.e., the 2018 and 2021 refreshments, respectively) were not included in the data that we used for the present study. Respondents to the recruitment interview were invited to a self-administered welcome survey. Among all respondents who had agreed to attend the GESIS Panel, 79.9% of the first cohort and 80.5% of the second cohort participated in the welcome survey. Respondents to this survey became regular panelists. Overall, about 65% of respondents participate in the web mode, and mode switches between the waves are not possible.

Every two months, around 5,000 panelists participate in GESIS Panel surveys. Respondents receive a 5-EUR prepaid incentive with every survey invitation, which is sent via postal mail. Web respondents additionally receive an email with a web link to the survey and up to two email reminders – that is, one reminder each both one and two weeks after the field start. The data used in the present study stem from the dataset of the GESIS Panel Extended Edition, Version 35 (GESIS, 2020), which was published in March 2020 and contains 34 waves (2014–2019).

The dependent variable of panel attrition was operationalised as follows: We defined a case as attrition if a respondent had participated in at least one GESIS Panel wave and had ultimately stopped participating before the fourth wave of 2019. Panel attrition can take two forms: First, respondents can actively de-register, and second, respondents who have not responded in three consecutive waves – despite the fact that their invitations were delivered – are not invited to future waves. For each wave, we used a binary indicator of whether a respondent had attritted as a dependent variable for our analyses. If a respondent had stayed in the panel for a particular wave, the variable was coded as 0. If a respondent had attritted after this wave, the variable was coded as 1. If a respondent had already attritted in one of the previous waves, the variable was coded as a missing value. In the following, we introduce the measurements of response time. A table with examples is provided in the appendix.

Days: metric operationalisation: For the metric operationalisation of the response time in days, we calculated the days it took respondents to participate and then send back their surveys in the respective waves. For every respondent and wave, we subtracted the date of the field start from the date of the return of the survey. For web respondents, the return date was measured by the survey software. For mail participants, we took into account the date when the survey agency received the filled-in questionnaire. It should be noted that the same type of postal mail sent from within Germany takes equally long to be delivered within the country, irrespective of the sending or delivery location.

Days: binary operationalisation: For the operationalisation of response time as a response before or after a specific date, we calculated five variables: (i) response on the first day or later, (ii) response after the first online reminder, (iii) response after the second online reminder, (iv) response in the first week or later, and (v) response in the first two weeks or later. Due to the lack of mail respondents who had participated on the first day, we did not estimate the relationship between this operationalisation and panel attrition for these respondents. We also did not estimate the relationship between response after reminders and panel attrition for mail respondents because these respondents had not received reminders. Instead, for both modes, we estimated the effects of participating both in the first week and in the first two weeks. For each operationalisation except for response on the first day, we coded early response as 0 and late response as 1. Response on the first day was coded as 1 if the respondent had participated on the first day and as 0 if they had participated later.

Number of contacts: In order to calculate the number of contacts, we counted the invitation as the first contact, the first reminder as the second contact, and the second reminder as the third contact. Thus, online respondents could have one, two, or three contacts. Mail respondents did not receive any reminders from the GESIS Panel and thus had only one contact. Therefore, we did not estimate the association between the number of contacts and response time for mail respondents. In practice, web respondents had four contacts because they had received the invitation not only via mail, but also via email. We counted the first two contacts as one contact because the two invitations had been received almost simultaneously. Web respondents could not respond to the first mail contact because the link to the questionnaire had not been included in the invitation letter that had been sent via postal mail and had only been available in the email invitation.

Respondents: metric operationalisation: For the operationalisation of response time as the proportion of panel members who had participated prior to a given respondent, we first separated web and mail respondents. Then, we ordered the two subsamples based on how quickly the respondents had returned the survey, and we assigned numbers to the respondents in this order. For web respondents, we could calculate the exact order by relying on the time stamps of their participation. For the mail respondents, we sorted them by relying on the time stamps provided by the postal service provider. For each subsample, we calculated both how many respondents had participated in the respective waves and how many respondents had returned their survey prior to a given respondent.

Respondents: binary operationalisation: For respondents who had responded prior to – or later than – other respondents, we compared five thresholds: 5%, 10%, 50%, 90%, and 95%. We contrasted the respondents who had participated before a certain threshold with the respondents who had participated after this threshold. We chose the first and last 5% and 10% of respondents as well as an equal division of the sample (i.e., 50%) in order to concentrate not only on late respondents, but also on those who had responded early.

Table 3.2. Description of the variables used in the analysis.

			Web participations (n = 3508, m = 54411)		Mail participations (n =1558, m =19822)	
	Minimum	Maximum	Mean	Std. Dev.	Mean	Std. Dev.
Attrition	0	1	0.01	0.12	0.03	0.16
Days: metric	0	68	7.98	9.69	14.34	11.32
Days binary:						
on first day	0	1	0.31	0.46	0.00	0.03
after first reminder	0	1	0.32	0.46	0.50	0.50
after second reminder	0	1	0.14	0.35	0.29	0.45
seven days after the field start	0	1	0.38	0.49	0.60	0.49
14 days after the field start	0	1	0.19	0.39	0.35	0.48
Number of contacts	1	3	1.61	0.80	2.12	0.78

Note: n = number of individuals, m = number of person*wave cases. Among all respondents, 32.22% of web respondents and 43.10% of mail respondents attrited.

Table 3.2, continued. Description of the variables used in the analysis.

	Minimum	Maximum	Web participations (n = 3508, m = 54411) Mean	Std. Dev	Mail participations (n =1558, m =19822) Mean	Std. Dev
Respondents: metric	0	100	50.11	28.84	50.78	28.67
Respondents binary: First 5%	0	1	0.05	0.22	0.05	0.22
First 10%	0	1	0.10	0.30	0.09	0.29
First 50%	0	1	0.50	0.50	0.50	0.50
Last 10%	0	1	0.11	0.31	0.11	0.31
Last 5%	0	1	0.06	0.23	0.06	0.23
Available time						
Full-time work	0	1	0.53	0.50	0.35	0.48
Partner	0	1	0.75	0.43	0.67	0.47
Children	0	1	0.62	0.49	0.68	0.47
Survey experience						
Difficult	1	5	1.91	0.87	2.04	0.90
Diverse	1	5	3.79	0.87	3.70	0.89
Important	1	5	3.44	0.88	3.48	0.95
Interesting	1	5	3.73	0.87	3.67	0.89
Long	1	5	2.30	1.01	2.23	0.97
Too personal	1	5	2.24	1.05	2.24	1.03
Control						
Mode = Mail	0	1	0.00	0.00	1.00	0.00
Participations	0	32	13.33	9.15	12.46	9.01
High education	0	1	0.57	0.50	0.24	0.43
Medium education	0	1	0.31	0.46	0.38	0.49
Low education	0	1	0.12	0.33	0.39	0.49

Note: n = number of individuals, m = number of person*wave cases. Among all respondents, 32.22% of web respondents and 43.10% of mail respondents attrited.

In Table 3.2, the descriptive statistics – that is, the ranges, means, and standard deviations – of the variables used in the analyses are grouped by survey mode. Generally, mail respondents tended to participate later, which may have been the result of the different response processes. We estimated 22 random-effects logistic panel regression models via the maximum likelihood of panel attrition to the following wave, which was 13 models for the web mode and 9 models for the mail mode. The models differed in their operationalisation of response time. As web and mail respondents were treated differently in the surveying process, their response time also differed. For example, mail respondents had to take the survey to a mailbox in order to send it to the researchers, whereas web respondents did not have to take any additional steps beyond answering the questions. This additional step may have delayed the response process for mail respondents but did not affect web participants. All operationalisations of response time were calculated for each wave in which a respondent had participated. Below, we first examine operationalisations of the number of days (Days: metric operationalisation and Days: binary operationalisation) before turning to the number of contacts (Number of contacts) followed by the two operationalisations for other respondents (Respondents: metric operationalisation and Respondents: binary operationalisation).

Our models can be described as

$$attrition_{i(t+1)} = \mu_i + \beta_1 * participation\ time_{it} + \beta_2 * previous\ participations\ in\ the\ panel_{it} + \beta_3 * working\ full\ time_{it} + \beta_4 * recept\ of\ medium\ education_{it} + \beta_5 * recept\ of\ high\ education_{it} + \beta_6 * having\ a\ partner_{it} + \beta_7 * having\ children_{it} + \beta_8 * previous\ panel\ wave{:}\ difficult_{it} + \beta_9 * previous\ panel\ wave{:}\ diverse_{it} + \beta_{10} * previous\ panel\ wave{:}\ important_{it} + \beta_{11} * previous\ panel\ wave{:} interesting_{it} + \beta_{12} * previous\ panel\ wave{:} long_{it} + \beta_{13} * previous\ panel\ wave{:} too\ personal_{it} + \alpha_i + \varepsilon_{it},$$

where i is the respondent, t is the wave, t+1 is the following wave, μ is an intercept, β_{1-13} represents the regression coefficients, ε_{it} is an error term that is different for each individual in every wave, and $alpha_i$ is an error term that includes a set of random variables. A detailed description of the method can be found in Allison (2009).

We analysed the data on all panelists who had participated in at least one GESIS Panel wave and who had not switched from mail to web mode. As a mode switch had been possible during one web-push event in 2018, during which 272 panelists switched modes, we excluded these panelists from our analyses, which left 3,508 web respondents and 1,558 mail respondents. The random-effects method enabled us to compare respondents who had attritted with those who had not. Contrary to fixed-effects models, respondents who had stayed in the panel were included in the random-effects analysis. This decision was important for the comparison of correctly predicted attrition (CPA) and overall correct predictions (OCP), which we used to compare model accuracy. Another example of using this method can be found in Boehmke & Greenwell (2019). In our models, we controlled for previous participation in the panel by including the following control variables: education, factors that influence available time (i.e., working full time, having a partner, and having children), and the evaluation of the previous panel wave (i.e., as difficult, diverse, important, interesting, long, or overly personal). Previous participation was measured as a metric count of the number of survey waves in which the respondent had already participated. Education was considered in three binary variables (i.e., a low, medium, or high level of education) and reflected each respondent's highest educational degree. Working full time, having a partner, and having children were all considered binary variables, with 0 indicating that the description did not apply to the respondent and 1 indicating that it did apply. Previous survey experience reflected the individual evaluation of the last survey and could range from 1 (i.e., "not at all") to 5 (i.e., "very"). Available time, survey experience, the number of times a respondent had previously participated, and education could all be argued to be related to both survey response time and panel attrition; therefore, we decided to include these items in order to account for confounding effects.

In order to compare how well the different operationalisations of response time predicted attrition, we assessed the correctness of the model's prediction of attrition. Each model estimated the predicted probability that each respondent in each wave (respondent*wave case) would attrit from the panel. This estimation enabled us to compare the calculated prediction of attrition with the actual attrition provided by our data – that is, whether or not a respondent had attritted after the respective wave. Thus, this comparison revealed the percentage of all cases of attrition that had been correctly predicted by each model, with the minimum value being 0 and the maximum being 100. When the group size of respondent*wave cases with a high probability of attrition increased, the number of predicted cases of attrition also increased. However, in this case, many panelists who had stayed in a panel were also predicted to attrit, which led to a high number of false predictions regarding respondents who had stayed. Therefore, we added correctly predicted attrition and correctly predicted staying as a method of examining the overall performance of the models. The sum is our second indicator – that is, overall correct predictions. We multiplied both indicators – that is, correctly predicted attrition and overall correct predictions – and ranked the operationalisations based on this product (i.e., correctly predicted to attrite × overall correct predictions).

The first step for calculating correctly predicted attrition was to estimate the probability by which each respondent would attrit. For each operationalisation, we made this calculation based on our regression models. The second step was to compare the probability of attriting with a threshold in order to determine whether respondents were predicted to stay or attrit. As the threshold selection was arbitrary, we randomly drew one individual threshold for every respondent out of the range of 90% of all predictions. The thresholds were distributed uniformly. All respondents with a higher probability of attriting than the threshold were predicted to attrit, and all respondents with a lower probability of attritting than the threshold were predicted to stay in the panel. In order to enable comparability, each respondent's individual threshold had to remain the same for all the models that were compared. Whenever the aim is to identify most cases of future attrition, the threshold should be chosen empirically by evaluating the correct predictions of attrition and the correct predictions overall.

3.5. Results

Our results are presented in Table 3.3. The size of the regression coefficients – which are presented in Column 2 (i.e., "Coefficients") – is of smaller relevance to us. For most of the operationalisations, we found a statistically significant relationship between response time and panel attrition. It is unexpected that online respondents who participated on the first day, among the first 10% and among the first 50% were more likely to attrit than later respondents. The same can be found for mail participants who participated among the first 50%. This may be reason for a neglectable difference or may show that participation timing has a nonlinear relationship to panel attrition. However, the main results of the analysis pertain to the amount of attrition that each model correctly predicted (i.e., "CPA", Column 4) and to the percentage of correct predictions (i.e., "OCP", Column 5) made by each model.

Table 3.3. Results of random-effects logistic regressions of panel attrition on different operationalisations of response time (34 waves).

Model	Coefficient	AIC	CPA	OCP	Product of CPA & OCP	Rank
Web mode: n = 3508						
Days: metric	0.05*** (0.01)	2528	0.61	0.60	0.37	3
Days binary: on first day	0.61** (0.18)	2593	0.59	0.57	0.34	11
Days binary: after first reminder	1.24*** (0.14)	2531	0.62	0.60	0.37	3
Days binary: after second reminder	1.10*** (0.16)	2564	0.60	0.58	0.35	7
Days binary: seven days after field start	1.12*** (0.15)	2545	0.64	0.59	0.38	1
Days binary: fourteen days after field start	1.21*** (0.14)	2546	0.59	0.59	0.35	7
Number of contacts	0.70*** (0.08)	2535	0.62	0.60	0.37	3
Respondents: metric	0.02*** (0.00)	2538	0.64	0.59	0.38	1
Respondents binary: First 5%	0.69 (0.42)	2603	0.60	0.56	0.34	11
Respondents binary: First 10%	0.84* (0.33)	2598	0.62	0.57	0.35	7
Respondents binary: First 50%	0.91*** (0.15)	2568	0.63	0.58	0.37	3
Respondents binary: Last 10%	1.32*** (0.16)	2550	0.60	0.59	0.35	7
Respondents binary: Last 5%	1.25*** (0.20)	2577	0.58	0.58	0.34	11

Note: Response on the first day could not be estimated for mail respondents due to a dearth of cases. "Number of contacts" and "after first/second reminder" were omitted for mail participants because these participants had not received reminders. * = p<0.05, ** = p<0.01, *** = p<0.001, standard errors in parentheses. Coefficient = coefficient of a random-effects logistic regression, AIC = Akaike information criterion, CPA = correctly predicted attrition, OCP = overall correct predictions. CPA, OCP, and the product of both all range from 0 to 100, where 0 means no correct predictions and 100 means that all cases were correctly predicted. The “Rank” column shows the ranking among the modes based on the product of the CPA and OCP (the higher, the better).

Table 3.3, continued. Results of random-effects logistic regressions of panel attrition on different operationalisations of response time (34 waves).

Model	Coefficient	AIC	CPA	OCP	Product of CPA & OCP	Rank
Mail mode: n = 1558						
Days: metric	0.04*** (0.01)	1350	0.68	0.55	0.37	1
Days binary: seven days after field start	0.58** (0.21)	1382	0.66	0.51	0.34	4
Days binary: fourteen days after field start	0.74*** (0.19)	1375	0.68	0.52	0.35	4
Respondents: metric	0.01** (0.00)	1382	0.64	0.51	0.33	8
Respondents binary: First 5% -	1.09*** (0.29)	1379	0.65	0.52	0.34	4
Respondents binary: First 10% -	0.53 (0.28)	1387	0.67	0.51	0.34	4
Respondents binary: First 50%	0.39* (0.19)	1386	0.63	0.51	0.32	9
Respondents binary: Last 10%	1.19 (0.22)***	1365	0.67	0.54	0.36	2
Respondents binary: Last 5%	1.40 (0.26) ***	1368	0.66	0.54	0.36	2

Note: Response on the first day could not be estimated for mail respondents due to a dearth of cases. "Number of contacts" and "after first/second reminder" were omitted for mail participants because these participants had not received reminders. * = p<0.05, ** = p<0.01, *** = p<0.001, standard errors in parentheses. Coefficient = coefficient of a random-effects logistic regression, AIC = Akaike information criterion, CPA = correctly predicted attrition, OCP = overall correct predictions. CPA, OCP, and the product of both all range from 0 to 100, where 0 means no correct predictions and 100 means that all cases were correctly predicted. The "Rank" column shows the ranking among the modes based on the product of the CPA and OCP (the higher, the better).

Our models reveal that a metric operationalisation of response time that used the days until survey return correctly predicted 61% of attrition among web respondents. Among mail respondents, this value was higher. Moreover, the model that included days until survey return as a metric variable correctly predicted 68% of attrition.

An analysis of response on the first day was only relevant for web respondents, and 59% of attrition was correctly predicted using this model. The analyses of response after the first and second reminders were also relevant only for web respondents. 62% of attrition among web respondents was correctly predicted by the model that used response after the first reminder, and 60% of attrition among web participants was correctly predicted by the model that used response after the second reminder. The model that used response after seven days correctly predicted 64% of attrition among web panelists and 66% of attrition among mail-mode panelists. The model that used response after 14 days correctly predicted 59% of attrition among web respondents and 68% of attrition among mail respondents. When the number of contacts was used to predict attrition among web respondents, this model correctly predicted 62% of attrition.

When response time was operationalised as the proportion of panel members who had participated before a given respondent, this model correctly predicted 64% of attrition in both modes. Models that included an operationalisation of response time for whether a respondent had been among the first 5% of respondents to return a survey correctly predicted 60% of attrition for web panelists and 65% of attrition for mail panelists. When response time was operationalised as being among the first 10% of respondents to participate or not, this model correctly predicted 62% of attrition among web respondents and 67% of attrition among mail respondents. The model with an operationalisation of response time that split the first 50% and the last 50% of participation responses correctly predicted 63% of attrition for both web panelists and mail panelists. The models that included the operationalisation of response time as being among the latest 10% to respond correctly predicted 60% of attrition for web panelists and 67% of attrition for mail panelists. When response time was defined based on whether or not a respondent had been among the latest 5% to respond, this model correctly predicted 58% of attrition for web respondents and 66% of attrition for mail respondents.

Overall, the 22 above-described models predicted attrition similarly well, which can also be seen in the close range of the AICs. In order to fairly evaluate these models, the OCP must be taken into account. The overall correct predictions are also similar to one another. There are three possible explanations for this similarity: First, the operationalisation may truly have had a similar relationship to attrition. Second, the similarity could be attributed to the random threshold, which works well when a variable is equally distributed. In our case, we had a skewed distribution, with which the randomly chosen threshold did not fit very well. In cases for which maximising correct predictions is more important than finding a fair method of comparing models, thresholds should be chosen empirically based on the threshold at which the highest OCP can be predicted. Third, the models predicted similar relationships to attrition because the covariates were identical. We believe that all these explanations apply to a certain extent, but the similarity between the correct predictions of attrition can be especially attributed to the third explanation. Two models that share all variables except for one – which was differently operationalised in the two models – should lead to similar predictions of attrition. In order to be able to more easily evaluate the models, Column 6 provides the product of the CPA and OCP. The seventh column (i.e., "Ranking") provides a ranking of the operationalisations based on the product of the CPA and OCP. As we can see, for web respondents, response time can predict attrition best when it is operationalised either (i) as response before or after seven days after the field start or (ii) as a metric share of prior respondents. For mail respondents, the metric operationalisation of response time works best. Due to the similarity of the results, we recommend that future work that investigates response time conduct a robustness check using multiple operationalisations.

3.6. Conclusion & discussion

In the present study, we operationalised response time in multiple ways and estimated the association between each operationalisation and panel attrition. Then, we compared the predictions of attrition that had been estimated by each model. We focused on 13 operationalisations of response time that are commonly used in the literature. The operationalisations were based on (i) the number of days it had taken a respondent to respond, (ii) the number of contacts, and (iii) the proportion of respondents who had participated prior to a given respondent. We found that many – albeit not all – of the operationalisations of response time were significantly related to panel attrition. The regression models predicted attrition similarly well. With regard to the mixed findings in the extant literature, the source of the variance was clearly not the operationalisation of response time, at least not when panel attrition was the variable of interest. Instead, the mixed findings may be attributed to different survey design features, such as field time or response rate.

Almost all of the operationalisations in our study were significantly related to panel attrition, and these operationalisations resembled one another in terms of both how well they could predict attrition and the accuracy of their overall predictions. Although the predictions of attrition were similar, we found that the operationalisations of response time as a metric share of prior respondents and as response before or after seven days after the field start worked best for online respondents and that the operationalisation of response time as a metric count of days until response worked best for mail respondents. Using these operationalisations, the balance between overall correct predictions and correctly predicted attrition was better than when using the alternative operationalisations with the data we used. One drawback of most of the operationalisations of response time that were applied in our study was that a variable that offered metric information – namely the day of the response or the share of prior respondents – was only used as a binary variable. A metric variable would have offered a richer potential for analysis, and the behaviour of the respondents could have been analysed in greater detail. Thus, when researching response time, it may also be advisable to apply a metric operationalisation.

The present study is not without limitations. First, we only focused on the association between the operationalisations of response time and panel attrition. However, other data-quality indicators also exist, such as item nonresponse, straightlining, and the consistency between two measurements, any of which can be associated with response time. Future studies should thus examine the impact of different operationalisations of response time on these data-quality indicators. Second, our study considered the entire process of participation in our operationalisation of response time. In other words, the time it took a response to reach the researchers was counted for the operationalisations of response time in the mail mode. An alternative would have been to count only the time it took a participant to respond. Indeed, after responding, mail-mode participants still had to take the questionnaire to a mailbox. We regard this process as an important part of participation; however, it is possible that this process is not relevant to other studies. Third, the operationalisations of response time could not always be applied to the mail mode and were thus not always disjointed. For instance, reminders were sent on a specific day, and we thus cannot know whether a given response was related to the day or to the reminder. This distinction could be experimentally varied in a future study.

Despite these limitations, our study offers added value for survey practitioners and survey methodologists alike. Indeed, it provides a summary of many operationalisations of response time and also indicates the advantages and disadvantages of the different operationalisations. Our method can aid survey practitioners in identifying potential future attrition, and it can help survey methodologists in determining which operationalisation should be used in future studies on response time. Survey practitioners can easily test which respondents are most likely to attrit by evaluating respondents' response time. This process is easy to implement and does not require much time or many resources; therefore, the test can be performed after each survey wave. However, survey practitioners should establish guidelines as to how to target interventions (e.g., regarding which respondents to address) that fit their survey with respect to aspects such as field length and attrition rate. In order to establish such rules, practitioners could evaluate questions such as the budget needed for the interventions, the size of the interventions, and the number of respondents who should be targeted by the interventions. Particularly late respondents could be the target of cost-effective interventions, such as additional incentives. Survey methodologists could additionally apply the method used in our study to other time-related data (e.g., website-login timestamps) and compare the way in which different operationalisations of these data are related to a given variable of interest. In terms of response time, our study revealed that the specific operationalisation is of smaller importance when predicting attrition. Hence, of all the operationalisations examined in our study, researchers are free to use the operationalisation that is easiest for them to calculate.

3.7. References

Allison, P. D. (2009). Fixed effects regression models. SAGE publications.

Armenakis, A. A., & Lett, W. L. (1982). Sponsorship and follow-up effects on response quality of mail surveys. Journal of Business Research, 10(2), 251–262. https://doi.org/10.1016/0148-2963(82)90031-5

Bates, N., & Creighton, K. (2000). The last five percent: What can we learn from difficult interviews? In Proceedings of the Annual Meetings of the American Statistical Association, 13-17.

Bethlehem, J. G. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. Groves, D. A. Dillman, J. L. Eltinge & R. J. Little (Eds.) Survey Nonresponse. Wiley

Boehmke, B., & Greenwell, B. (2019). Hands-on machine learning with R. Chapman and Hall/CRC.

Bollinger, C. R., & David, M. H. (2001). Estimation with response error and nonresponse: Food-stamp participation in the SIPP. Journal of Business & Economic Statistics, 19(2), 129–141. https://doi.org/10.1021/jf0483366

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. Social Science Computer Review, 36(1), 103–115. https://doi.org/10.1177/0894439317697949

Brehm, S. S., & Brehm, J. W. (2013). Psychological reactance: A theory of freedom and control. Academic Press.

Cannell, C. F., & Fowler, F. J. (1963). Comparison of a self-enumerative procedure and a personal interview: A validity study. Public Opinion Quarterly, 27(2), 250–264. https://doi.org/10.1086/267165

Cohen, S. B., & Carlson, B. L. (1995). Characteristics of Reluctant Respondents in the National Medical Expenditure Survey. Journal of Economic and Social Measurement, 21(4), 269–296. https://doi.org/10.3233/JEM-1995-21402

Cohen, S. B., Machlin, S. R., & Branscome, J. M. (2000). Patterns of survey attrition and reluctant response in the 1996 medical expenditure panel survey. Health Services and Outcomes Research

Methodology, 1(2), 131–148. https://doi.org/10.1023/A:1012543121850

Curtin, R., Presser, S., & Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. Public Opinion Quarterly, 64(4), 413–428. https://doi.org/10.1086/318638

de Leeuw, E. D., & Hox, J. J. (1988). Artifacts in Mail Surveys: The Influence of Dillman's Total Design Method on the Quality of the Responses. In W. E. Saris and I. N. Gallhofer (Eds.) Sociometric Research. Vol. 2: Data Analysis. (pp. 61–73). Springer. https://doi.org/10.1007/978-1-349-19054-6_3

Diaz de Rada, V. (2005). The Effect of Follow-up Mailings on The Response Rate and Response Quality in Mail Surveys. Quality & Quantity, 39(1), 1–18. https://doi.org/10.1007/s11135-004-5950-5

Donald, M. N. (1960). Implications of Nonresponse for the Interpretation of Mail Questionnaire Data. Public Opinion Quarterly, 24(1), 99-114. https://doi.org/10.1086/266934

Eckland, B. K. (1965). Effects of prodding to increase mail-back returns. Journal of Applied Psychology, 49(3), 165–169. https://doi.org/10.1037/h0021973

Friedman, E. M., Clusen, N. A., & Hartzell, M. (2003). Better Late? Characteristics of Late Respondents to a Health Care Survey. Proceedings of the Survey Research Methods Section of the American Statistical Association, 992–98.

GESIS. (2020). GESIS Panel—Extended Edition. GESIS, Cologne. ZA5664 Datafile Version 35.0.0. https://doi.org/10.4232/1.13435

Gilbert, G. H., Longmate, J., & Branch, L. G. (1992). Factors influencing the effectiveness of mailed health surveys. Public Health Reports, 107(5), 576-584.

Green, K. E. (1991). Reluctant Respondents: Differences between Early, Late, and Nonresponders to a Mail Survey. The Journal of Experimental Education, 59(3), 268–276. https://doi.org/10.1080/00220973.1991.10806 566

Groves, R. M. (2005). Survey errors and survey costs. John Wiley & Sons.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. Public opinion quarterly, 70(5), 646–675.

Gummer, T., & Struminskaya, B. (2020). Early and Late Participation during the Field Period: Response Timing in a Mixed-Mode Probability-Based Panel Survey. Sociological Methods & Research 0(0), 1-24. 0049124120914921. https://doi.org/10.1177/0049124120914921

Helasoja, V. (2002). Late response and item nonresponse in the Finbalt Health Monitor Survey. The European Journal of Public Health, 12(2), 117–123. https://doi.org/10.1093/eurpub/12.2.117

Kay, W. R., Boggess, S., Selvavel, K., McMahon, M. F., & Kay, W. R. (2001). The Use of Targeted Incentives to Reluctant Respondents on Response Rate and Data Quality. In Proceedings of the Annual Meeting of the American Statistical Association, 1-7.

Kennickell, A. B. (2017). What do the 'late 'cases tell us? Evidence from the 1998 Survey of Consumer Finances. Statistical Journal of the IAO 33(1), 81-92. https://doi.org/10.3233/SJI-160302

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H., & Koskenvuo, M. (2001). Non-response and related factors in a nation-wide health survey. European journal of epidemiology, 17(11), 991–999. https://doi.org/10.1023/A:1020016922473

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

Kreuter, F., Müller, G., & Trappmann, M. (2014). A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents. Sociological Methods & Research, 43(3), 452–464. https://doi.org/10.1177/0049124113508094

Kunz, F. (2010). Empirical Findings on the Effects on the Response Rate, the Responses and the Sample Composition. methods, data, analyses, 4(2), 127-155. https://doi.org/10.12758/mda.2010.009.

Lin, I.-F., & Schaeffer, N. C. (1995). Using survey participants to estimate the impact of nonparticipation. Public Opinion Quarterly, 59(2), 236–258. https://doi.org/10.1086/269471

Lynn, P., & Lugtig, P. J. (2017). Total Survey Error for Longitudinal Surveys. In P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), Total Survey Error in Practice (pp. 279–298). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119041702.ch13

Olson, K. (2013). Do non-response follow-ups improve or reduce data quality? A review of the existing literature. Journal of the Royal Statistical Society: Series A (Statistics in Society), 176(1), 129–145. https://doi.org/10.1111/j.1467-985X.2012.01042.x

Preisendörfer, P., & Wolter, F. (2014). Who Is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys. Public Opinion Quarterly, 78(1), 126–146. https://doi.org/10.1093/poq/nft079

Roßmann, J., & Gummer, T. (2016). Using Paradata to Predict and Correct for Panel Attrition. Social Science Computer Review, 34(3), 312–332. https://doi.org/10.1177/0894439315587258

Schoenman, J. A., Berk, M. L., Feldman, J. J., & Singer, A. (2003). Impact Of Differential Response Rates On The Quality Of Data Collected In The CTS Physician Survey. Evaluation & the Health Professions, 26(1), 23–42. https://doi.org/10.1177/0163278702250077

Skarbek-Kozietulska, A., Preisendörfer, P., & Wolter, F. (2012). Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. Zeitschrift für Soziologie, 41(1), 5-23. https://doi.org/10.1515/zfsoz-2012- 0103

Struminskaya, B., & Gummer, T. (2022). Risk of Nonresponse Bias and the Length of the Field Period in a Mixed-Mode General Population Panel. Journal of Survey Statistics and Methodology 10(1), 161-182. https://doi.org/10.1093/jssam/smab011

Treat, J. B., & Stackhouse, H. F. (2002). Demographic comparison between self-response and personal visit interview in Census 2000. Population Research and Policy Review, 21(1), 39–51. https://doi.org/10.1023/A:1016541328818

Voigt, L. F., Koepsell, T. D., & Daling, J. R. (2003). Characteristics of telephone survey respondents according to willingness to participate. American journal of epidemiology, 157(1), 66–73. https://doi.org/10.1093/aje/kwf185

Yan, T., Tourangeau, R., & Arens, Z. (2004). When less is more: Are reluctant respondents poor reporters? In Proceedings of the Annual Meetings of the American Statistical Association. 4632-4651.

4. Now, later, or never? Using response-time[5] patterns to predict panel attrition[6]

4.1. Abstract

Preventing panel members from attriting is a fundamental challenge for panel surveys. Research has shown that response behavior in earlier waves (response or nonresponse) is a good predictor of panelists' response behavior in upcoming waves. However, response behavior can be described in greater detail by considering the time until the response is returned. In the present study, we investigated whether respondents who habitually return their survey late and respondents who switch between early and late response in multiple waves are more likely to attrit from a panel. Using data from the GESIS Panel, we found that later response is related to a higher likelihood of attrition (AME = 0.087) and that response-time[7] stability is related to a lower likelihood of attrition (AME = −0.013). Our models predicted most cases of attrition; thus, survey practitioners could potentially predict future attriters by applying these models to their own data.

4.2. Introduction

Panel attrition is a central problem in longitudinal surveys. Indeed, it is a special case of unit nonresponse (Alwin, 2007:135; Smith, 2011) that occurs when respondents who once regularly participated in a longitudinal survey drop out of a sample and are no longer invited to future survey waves. As opposed to unit nonresponse in a cross-sectional survey, nonresponse due to relocation is much more likely in panel surveys. Apart from that, the decision of whether to participate may be influenced by prior survey experience (Groves et al., 2011; Lynn, 2009; Watson & Wooden 2009). Recruiting new panelists is costly, and the value of panelists' data increases with each panel wave both because more information offers more potential for analysis and because repeated measurements over a longer time period are necessary to analyze societal changes. The topic of panel attrition has thus been examined by many researchers (e.g.,

[5] This chapter has been published in the journal "International Journal of Social Research Methodology". In the published article, the term "response time" has been used to describe what has previously been called "participation time" in this dissertation. In this chapter, I maintain this choice of words for reasons of similarity to the published article.
[6] This chapter is joint work with Bernd Weiß. It has been published in the journal "International Journal of Social Research Methodology".

Herzing & Blom, 2019; Eisnecker & Kroh, 2016; Roßmann & Gummer, 2016; Lugtig, 2014; Struminskaya, 2014).

A good predictor of panel attrition is response behavior in previous waves, such as participation or nonresponse (Roßmann & Gummer, 2016; Lugtig, 2014). However, the recent developments of a growing number of online panels and the shorter times between any two given panel waves have enabled attrition to be investigated in greater detail using response time in previous panel waves. We argue that the specific time that a respondent needs to return their response may also be connected to panel attrition. Some evidence exists to indicate that response time in one panel wave is related to participation in the next wave (Cohen et al., 2000). This response time is the time that a respondent requires for returning a self-administered questionnaire to the survey agency. For web interviews, this is the time until a respondent submits the last question they answer, regardless of whether this question is also the last question in the questionnaire. For mail questionnaires, response time is the time until the return envelope is delivered to the survey agency. Theoretically, response times can also be calculated for interviewer-administered surveys. However, in self-administered surveys, respondents have the possibility to participate at any time, whereas in interviewer-administered surveys, respondents' participation depends on the presence of an interviewer.

In comparison with other variables that can help researchers to predict attrition, response times provide three advantages: First, response times can be assessed from survey paradata. Accordingly, considering response times neither requires additional survey time nor places any additional burden on respondents. Moreover, paradata – such as response times – are usually available for every online survey. Hence, response time is an easily available tool that researchers can use for secondary analysis when they have not collected the data themselves. Second, respondents do not directly report their response times, and they are therefore less likely to falsify them. Third, one advantage of considering response times instead of other possible variables is that response times are comparable across surveys, whereas content variables are often assessed with different wording or different scales across surveys. Therefore, responses to content variables from different surveys are not necessarily appropriate for comparison. Overall, response times could enable cost-effective methods – such as targeted interventions – to be applied in order to convince respondents who are about to attrit to stay in the panel.

Even though respondents do not have to report their response time because it is automatically collected as paradata in most online surveys (Kreuter & Casas-Cordero, 2010), research on response time is limited. While we have information about the influence of response time on item nonresponse from cross-sectional surveys (Olson, 2013), longitudinal analyses of response time using panel data are lacking. We regard this lack as a considerable research gap because response times may offer many opportunities to analyze panel data. First, the development of response times over multiple waves may provide valuable information about the identification of attriters. By considering response times, we can investigate the widely accepted hypothesis that participation in one panel wave influences participation in the next wave (Roßmann & Gummer, 2016; Lugtig, 2014; Nicoletti & Peracchi, 2005). Respondents who respond habitually should be more likely to maintain this habit than to attrit. A central indicator for attrition could be habitual response time, which suggests that a respondent who maintains the same response time as in the previous wave will stay in the panel, regardless of whether they respond early or late. Second, response time might serve as an indicator of latent factors that influence panel attrition, such as the motivation to participate in the panel or the time available to panelists. We assume that respondents do not continuously delay participation without a reason and that they instead either participate later due to time constraints or have low motivation to participate. Reasons for participating later may also be reasons to attrit in future waves.

The present study focuses on the question of whether response time can be used as a tool to identify respondents who are likely to attrit from a panel. To answer this question, we generated patterns of longitudinal response time and analyzed the relationship between response time and attrition on the one hand and between response-time habits over multiple waves and attrition on the other hand. We focused on three research questions: First, we were interested in the frequency of previous late responses and asked whether late respondents are more likely to attrit from a panel. Second, we investigated whether response time can substitute for available time and survey motivation – either of which may not be accessible for every survey – and asked whether response time can be an indicator of available time and survey motivation with respect to explaining attrition. Third, we focused on patterns of previous behavior – and particularly of response behavior – and asked whether respondents who respond habitually are less likely to attrit from a panel.

4.3. Previous Research

Many variables have been used to explain attrition. Therefore, we provide below an overview of the results of the few longitudinal studies that have addressed response time and attrition.

Subsequently, we turn to the relationship between response time and data quality in cross-sectional surveys. Finally, we conclude with an overview of how response time has been operationalized thus far.

Since response times have rarely been researched using panel surveys, empirical evidence on response-time habits is lacking. Cohen et al. (2000) found evidence that late respondents are more likely to become nonrespondents in the second wave of a panel survey, but unfortunately, their study was limited to only two panel waves. Evidence also exists showing that respondents are more likely to participate in a panel wave if they have already participated in previous waves (Roßmann & Gummer, 2016; Haunberger, 2011; Göritz, 2008; Nicoletti & Peracchi, 2005).

Due to the lack of research on the influence of response times on panel attrition, we wish to stress that panel attrition is a dimension of data quality. Although some researchers have not found response time to affect data quality (Preisendörfer & Wolter, 2014 (face-to-face survey); Diaz de Rada, 2005; Helasoja, 2002), many other researchers have found that late response is related to reduced data quality. For example, Kunz (2010), Tancreto & Bentley (2005), Friedman et al. (2003), and Donald (1960) found a higher item-nonresponse rate among late respondents. Other studies have also shown that the data provided by late respondents are less likely to be consistent with other data sources, such as administrative data (Preisendörfer & Wolter, 2014 (mail survey); Kreuter et al., 2010; Gilbert et al., 1992; Armenakis & Lett, 1982; Eckland, 1965). In addition, Green (1991) found that participants who respond later are less likely to answer open questions than are early respondents.

It is interesting to note that response time has been operationalized in several ways in previous studies. For example, Gummer and Struminskaya (2020), Preisendörfer & Wolter (2014), and Skarbek-Kozietulska et al. (2012) all operationalized response time as a metric variable reflecting days of response (e.g., 13 days vs. ~2 days). A survey-design-oriented approach – such as that used in the studies by Kreuter et al. (2014) and Helasoja (2002) – counts the number of contacts or reminders (e.g., 1 vs. ~3). In another approach, a distinction can be made between first respondents and residual respondents on the one hand or between last respondents and residual respondents on the other hand using a fixed number (the 100 first/last respondents) or percentage (the first/last 10% of respondents), as is made in Gummer and Struminskaya's (2020) study. Voigt et al. (2003) and Friedman et al. (2003) chose time intervals (everyone who responds before/after 14 days). An empirical approach could calculate the mean number of days of response and classify as late all respondents who reply later than the mean or more than one standard deviation later than the mean.

The lack of research on the relationship between response time and panel attrition and the mixed results on the relationship between response time and data quality render it difficult to draw hypotheses from the existing literature. Therefore, we next examine the theoretical background of panel attrition and response time.

4.4. Theory

Panel attrition has been dealt with mostly by rational choice theories, which explain attrition similarly to survey participation (e.g., Dillman et al., 2016; Groves et al., 2000). Rational choice theories argue that when faced with the choice between multiple alternatives, actors engage in the action that promises the best cost–benefit calculation. Hence, when evaluating whether to participate in a panel wave, panelists consider whether their benefit from participation is greater than their benefit from nonparticipation – that is, they take into account the costs of both alternatives. Respondents may draw intrinsic benefits from survey participation, for example, due to compliance with a norm of politeness, loyalty to a public institution, support for science, or having a new experience (Esser, 1986). Respondents who draw a large benefit from participation may find it easier to make the decision to participate and consequently participate earlier, whereas respondents who receive a lesser benefit may postpone their ultimate decision to participate. The costs of participation could include the time required by survey participation. With respect to respondents with time constraints, participation may be more costly, and the respondents may be more likely to participate late. These respondents may also be more likely to attrit from the panel at some point.

Hypothesis 1 (H1): Late respondents are more likely to attrit from a panel.

The argument presented above is based on the assumption that available time and survey motivation influence both response time and attrition, which means that response time only has a relationship with attrition because it acts as an indicator. Response time reflects the relationship that latent factors – such as available time and survey motivation – have with panel attrition. If such latent factors do not influence both response time and attrition, there should be no relationship between response time and attrition.

Hypothesis 2 (H2): The effect of available time and survey motivation on attrition is partially mediated by response time and response-time habit.

A vast amount of research on panel attrition has thematized a "habit of participating." If this habit is lacking, respondents are more likely to attrit (Lugtig, 2014). We can draw the same conclusion when applying the model of frame selection (Kroneberg, 2014; Esser, 2011) to the response-decision process. The model of frame selection argues that actors make decisions in an automatic or a reflective mode. This decision can be divided into smaller decisions about (1) the kind of situation (frame), (2) how actors are expected to behave in the given situation (script), and (3) how they ultimately act (action). The evaluation of whether to participate in a panel-survey wave can be broken down into the following decisions: (1) whether actors see themselves in a situation in which they are invited to participate in a survey, (2) whether they are expected to participate in a survey, and (3) whether they want to participate in a survey (Schnell, 2013). As participants face this situation on a regular basis, we expect most panel respondents to automatically decide on the kind of situation and the expected reaction and to evaluate these decisions correctly. However, the ultimate reactions to a survey invitation can be diverse. Since panelists are in a recurring situation, they may respond with the same behavior as in previous situations, which can be early or late participation. Rather than evaluating possible behavior alternatives, respondents repeat their mentally anchored behavior and only reflect on their decision to participate if specific factors – such as their costs or benefits of participating – have changed. In this case, respondents may deviate from their usual behavior. Panelists who reflect on their participation may decide that participation does not benefit them and may therefore be more likely to quit the panel.

Hypothesis 3 (H3): Respondents with a stable response-time habit are less likely to attrit from a panel.

We are aware that H1 and H3 partially contradict each other with respect to respondents who have a long history of late responses in their response-time pattern. Since research on response time is lacking, it is not yet clear whether these respondents are more or less likely than other respondents to attrit. We expect the effect of a stable response-time habit (H3) to be greater than the effect of response time (H1), which means that habitually late respondents can be expected to stay in the panel rather than to attrit.

4.5. Data

The data analyzes for the present study used data from the GESIS Panel, which is a German probability-based mixed-mode access panel (Bosnjak et al., 2018) that contains about 5,000 panelists, who must be at least 18 years old. The sample was initially recruited in 2013 and consisted of a random sample drawn from municipal population registers. It was refreshed in 2016. Respondents from both cohorts participated in a recruitment interview and were subsequently invited to complete a self-administered welcome survey. After completing the welcome survey, the participants were regarded as being regular panelists. American Association for Public Opinion Research (AAPOR) Response Rate 1 in the face-to-face recruitment interview was 35.5% for the initial recruitment cohort and 33.2% for the refreshment cohort. For the initial recruitment cohort, 79.89% of the recruited respondents participated in the welcome survey, and for the refreshment cohort, 80.51% participated in the welcome survey.

Every two months, all panelists are invited to complete a GESIS Panel survey. Almost all 5,000 panelists participate: Around 65% of respondents participate online, and the other participants receive a paper-and-pencil questionnaire together with a postal invitation. All respondents receive a 5-EUR prepaid incentive. Online respondents additionally receive an email with a web link to the survey and email reminders both one and two weeks after the field start. The GESIS Panel is open to researchers from all fields to submit questionnaire proposals, which leads to highly diverse survey topics. The surveys usually take about 20 minutes to complete and contain at least 4 different topics. For the present study, we used the data published in October 2018, GESIS Panel version 26 (GESIS, 2018). These data contain 28 waves (2013–2018). The last wave is the GESIS Panel wave "fa," which was conducted between February and March 2018. The GESIS Panel is especially well suited for testing the influence of response time on panel attrition because this panel collects data every two months.

Attrition: Our dependent variable is panel attrition. We defined panelists as attriters if they had participated in at least one regular wave in the GESIS Panel and had either actively deregistered as a panelist or not replied in three subsequent waves despite having received an invitation to respond. We used a binary indicator of whether a respondent had attritted from the panel by the February/March wave in 2018 as the dependent variable for our analyses. We defined staying in the panel as 0 and attrition as 1.

Response time and response-time habit: The central predictor variables for our analysis were response time and response-time habit. To operationalize these two variables, we generated individual response-time patterns for each panelist with respect to their response times for the previous 10 waves.[8] For the panelists who had not participated in 10 waves, we took into account all the waves in which they had participated. Drawing on the day of survey return, we calculated whether a respondent had returned the survey up to and including 14 days after the field start (early), more than 14 days after the field start (late), or not at all (nonresponse).[9] We operationalized response time as the relative frequency of late responses in a response-time pattern and therefore counted the number of late responses and the total number of times that each respondent had participated in each pattern.

We measured response time as the proportion of late responses among the last 10 times that every respondent had participated. Thus, the indicator of response time ranges from 0 to 1, with 0 indicating only early responses and 1 indicating only late responses.

We operationalized response-time habit as the longest sequence of identical consecutive response times in a pattern. Respondents could have a maximum of 10 subsequent waves with the same response behavior, and the minimum was one. A sequence length of one wave indicates that respondents had switched their response times from wave to wave, and a sequence length of 10 waves indicates that respondents had never switched their response time. If respondents had provided multiple different sequences following one another, such as early responses in two subsequent waves followed by late responses in three waves, we chose the larger number.

Available time: We measured the available time for a respondent to complete the survey using three binary variables that indicated whether the respondent had had a partner, had had children younger than 16 in the household, and had worked full time.

Survey motivation: At the end of each GESIS Panel wave, respondents were asked to evaluate the current survey as "important," "diverse," "interesting," "long," "difficult," or "too personal" on a five-point-scale. We operationalized survey motivation using the responses to these 6

[8] Our choice of 10 waves was not arbitrary. We estimated our model – which is described later in the Method section – multiple times by varying the number of waves (pattern lengths). To determine the optimal pattern length, we calculated the Akaike Information Criterium (AIC) and compared the predicted attrition with the actual attrition of each model. Based on these indicators, we found that a pattern length of 10 is optimal. Shorter response-time patterns do not perform as well in the models as patterns that were built from 10 waves. Longer response-time patterns do not improve the models.
[9] We chose the cut-off point of 14 days because the online respondents had received their second reminder to participate at this point. Choosing a cut-off point was necessary to operationalize the response-time habit. In addition, this operationalization of response time is often applied in the current literature on response times (Olson, 2006).

questions. To account for the multiple waves, we calculated the mean over time. Hence, for every respondent, we calculated the mean among all waves in which the respondent had participated for each of the 6 indicators of survey motivation.

Control variables: In addition, we controlled for education,[10] the number of waves in which respondents had participated, and the survey mode. Survey mode, education, and the number of waves in which a respondent had previously participated can be argued to be related to both survey response time and panel attrition. Therefore, we decided to include these items to account for confounding effects. To operationalize education, we calculated the mean of the available waves. We also summed the number of times that each respondent had participated in the 28 waves. Panelists chose the mode when entering the panel and could not switch.

Overview: Table 4.1 presents the means and standard deviations of all the relevant variables for the full sample and for active and attrited respondents. We found that 17% of all respondents had dropped out of the panel and that active respondents had returned an average of 23% of their last surveys 14 or more days after the field start. Respondents who had attrited from the panel returned an average of 40% of their surveys at this time. Active panelists usually had the same response time for 6.9 of up to 10 waves, and attrited panelists repeated the same response time for 4.6 of up to 10 waves. Hence, the respondents who had attrited from the GESIS Panel participated later and in a less habitual manner than did the active respondents.

4.6. Method

Our aim in the present study was to test whether response time and response habit over time could be used as proxies for the variables that influence available time and survey motivation so that researchers would not have to attempt to assess these variables in a survey. Therefore, we estimated three logistic-regression models: The first model includes response time and response-time habit and shows the effect of response time and response-time habit on attrition without controlling for other variables. The second model shows the effect of available time and survey motivation on attrition without controlling for response time or response-time habit

---

[10] The German high-school system includes lower-level secondary school (Hauptschule), medium-level secondary school (Realschule), and upper-level secondary school (Gymnasium). Low, medium, and high levels of education are related to obtaining the highest educational degree from the aforementioned schools, respectively.

Table 4.1. Descriptive Statistics.

			Overall (n=5340		Participant (n=4429)		Attritted (n=911)	
	Min.	Max.	Mean	SD	Mean	SD	Mean	SD
Attrition until wave 28	0	1	0.17	0.38	0.00	0.00	1.00	0.00
Response time habit	1	10	6.51	3.08	6.91	3.02	4.56	2.58
Response time	0	1	0.26	0.30	0.23	0.29	0.40	0.31
Available time								
Full-time work	0	1	0.47	0.46	0.46	0.46	0.48	0.46
Children	0	1	0.64	0.32	0.67	0.30	0.48	0.38
Partner	0	1	0.78	0.38	0.79	0.37	0.75	0.40
Survey motivation: Evaluation								
Important	1	5	3.43	0.64	3.46	0.63	3.29	0.68
Diverse	1	5	3.73	0.57	3.76	0.56	3.59	0.56
Interesting	1	5	3.67	0.59	3.71	0.58	3.47	0.60
Long	1	5	2.34	0.66	2.30	0.64	2.57	0.70
Difficult	1	5	1.95	0.57	1.92	0.56	2.10	0.58
Too Personal	1	5	2.25	0.73	2.21	0.72	2.40	0.74
Control								
Low Education	0	1	0.19	0.39	0.19	0.39	0.21	0.41
Medium education	0	1	0.35	0.48	0.34	0.48	0.38	0.49
High Education	0	1	0.46	0.50	0.47	0.50	0.41	0.49
Participations	2	28	19.51	7.88	20.91	7.55	12.7	5.53
Mode = offline	0	1	0.34	0.47	0.33	0.47	0.38	0.49

Note: n: Number of respondents, Min.: Minimum, Max.: Maximum, SD: Standard Deviation

The third model includes all variables. By examining the explained variance of the models, we were able to disentangle whether response time explains the same variance in attrition as available time and survey motivation and whether the effects are cumulative.

For each model, we calculated the predicted probability that every respondent would attrit from the panel. We built confusion matrices (Boehmke & Greenwell, 2019) to compare this predicted probability of attriting with the information about whether the respondent had actually attritted.

Confusion matrices allowed us to (1) calculate how many cases can be correctly predicted by our model and (2) differentiate between whether the goodness-of-fit refers to a small group of interest or to a large group that behaves as expected. We had three aims with this strategy: First, we wanted to compare the predictions of the models. Second, we wanted to assess whether the goodness-of-fit referred to the true-negative (TN) values – that is, to the panelists who were correctly predicted to stay in the panel – or to the true-positive (TP) values – that is, to the panelists who were correctly predicted to attrit. Third, using our models, we wanted to determine how many of the positive (P) values – that is, the attriters – were TP.

When using confusion matrices, researchers must decide on a threshold beyond which the prediction is positive. This threshold ideally maximizes true-positive- and true-negative values and minimizes false-positive values (FP – i.e., panelists who were falsely predicted to attrit) and false-negative values (FN – i.e., panelists who were falsely predicted to stay). We calculated confusion matrices with thresholds from 0.05–0.25. Using our data and with the aim of balancing the true values (TP+TF) on the one hand and the TP on the other hand, we considered a threshold of 0.15 to be appropriate for our data. Respondents with a higher likelihood of attriting than 0.15 were predicted to attrit and became TP or FP depending on their actual outcome. Respondents whose likelihood of attriting was lower than this threshold were predicted to stay in the panel and became TN or FN depending on their actual outcome. When comparing the predictions of the three models, the threshold itself is rather unimportant; rather, it is more important that the threshold remain the same over the models. However, the threshold is important for evaluating the goodness-of-fit of the models and for identifying attriters in advance. After coding every respondent as “predicted to attrit” or “not predicted to attrit,” we compared predicted and actual outcomes. Using this method, we could distinguish between correctly predicted attriters (TP), falsely predicted attriters (FP), correctly predicted stayers (TN), and falsely predicted stayers (FN).

4.7. Results

We first present the logistic-regression results and then display the confusion matrices for the estimated models. The first model describes the effect of both response time and response-time habit on attrition.[11] The second model includes variables that we expected to influence the latent factors of available time and survey motivation. The first and second models are nested in the third model.

Our first hypothesis (H1) predicted that late respondents should be more likely to attrit from the panel in one of the subsequent waves. In Table 4.2, we see the results of the three logistic regressions that we estimated. We found that the higher the percentage of late responses among all responses was, the more likely the respondent was to attrit from the panel (AME: 0.064). We also found that when controlling factors that influence available time and survey motivation, response time leads to an even-greater likelihood of attriting (AME: 0.087). Our empirical findings support hypothesis H1.

In our second hypothesis (H2), we assumed that the effect of available time and survey motivation on attrition should be partially mediated by response time and response-time habit. That means that once we had controlled for available time and survey motivation, response time and response-time habit should not have been significantly related to panel attrition. While the effect size of response-time habit on attrition decreased, the effect was still statistically significant. The effect of response time on attrition increased in size. The effects of the variables that we used to operationalize available time and survey motivation remained stable when controlling for response time and response-time habit. Thus, when comparing Model 1 with Model 3 and Model 2 with Model 3, we must reject H2. Even though Model 3 controls the factors of available time and survey motivation, response time is still found to be significantly related to panel attrition.

Our third hypothesis (H3) stated that respondents with a stable response-time habit should be less likely to attrit from a panel. When the longest sequence of identical response time increases, the likelihood of attrition decreases in all models (AME Model 1: –0.031; AME Model 3: –0.013), which is in line with H3.

Comparing the goodness-of-fit of the models reveals that the second and third models perform better than the first model. We calculated likelihood-ratio tests to compare Model 1 with Model 3 and Model 2 with Model 3. Models 1 and 3 had a deviance of 944.84, while Models 2 and 3

[11] Response time and response-time habit are negatively correlated ($r = -0.65$). In a bivariate model, response time has an average marginal effect of 0.234 and a standard error of 0.01 on attrition. This effect is statistically significant ($p < 0.001$). Response-time habit has an average marginal effect of –0.034 and a standard error of 0.00 on attrition. This effect is also statistically significant ($p < 0.001$).

Table 4.2. Logistic regression on attrition until the 28th panel wave.

	Model 1 (AME)	Model 2 (AME)	Model 3 (AME)
Response time	0.064** (0.02)		0.087*** (0.02)
Response time habit	−0.031*** (0.00)		−0.013*** (0.00)
Available time			
Full time work		0.009 (0.01)	0.001 (0.01)
Children		−0.167*** (0.01)	−0.145*** (0.01)
Partner		−0.029* (0.01)	−0.023* (0.01)
Survey motivation			
Evaluation: important		−0.017 (0.01)	−0.016 (0.01)
Evaluation: diverse		0.020 (0.02)	0.016 (0.01)
Evaluation: interesting		−0.043** (0.02)	−0.035* (0.02)
Evaluation: long		0.001 (0.01)	−0.006 (0.01)
Evaluation: difficult		0.069*** (0.01)	0.060*** (0.01)
Evaluation: private		−0.004 (0.01)	−0.002 (0.01)
Control			
Participations		−0.016*** (0.00)	−0.015*** (0.00)
Medium education		−0.012 (0.01)	−0.018 (0.01)
High education		−0.030 (0.01)	−0.041** (0.01)
Mode = Offline		0.004 (0.01)	−0.023* (0.01)
Pseudo R2	0.10	0.25	0.29
AIC	4416.8	3663.9	3497.9
n	5340	5340	5340

Note: AME: Average Marginal Effects, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, Standard errors in parantheses, AIC: Akaike Information Criterium, n: Number of respondents had a deviance of 169.99. The results of both likelihood-ratio tests were highly significant ($p < 0.01$). Therefore, response time cannot fully replace available time or survey motivation as predictors of panel attrition. This result contradicts our second hypothesis.

In conclusion, when explaining panel attrition, response time and response-time habit can be helpful. Although Model 2 estimates attrition better than Model 1, response time and response-time habit may be good predictors of attrition when relevant content variables – such as available time and survey motivation – are lacking. In addition, including response time and response-time habit in a model that considers available time and survey motivation still improves the model. Although the empirical results do not support our second hypothesis, they clearly support our first and third hypotheses. We next aim to determine whether the predicted outcomes are consistent with the actual outcomes. If the approach of the present study is practically applied in future studies with the goal of reducing attrition, it might be wise to examine the identification of respondents who are most likely to attrit next. This knowledge would enable survey conductors to target interventions at these panelists in order to motivate them to stay in the panel. For example, panelists who are predicted to attrit might receive an additional greater incentive. Below, we discuss how predicted outcomes relate to actual outcomes.

Table 4.3 provides confusion matrices for the three models with varying thresholds ranging from 0.05–0.25. Each row represents one confusion matrix. The columns list the results of the comparison of actual and predicted attrition. The distribution of prediction and the actual outcome for different thresholds are given in percentages. The column "True Positive" lists the panelists who had been correctly predicted to attrit, the column "False Positive" lists the panelists who had been falsely predicted to attrit, the column "True Negative" lists the panelists who had been correctly predicted stay, and the column "False Negative" lists the panelists who had been falsely predicted to stay. For each threshold, the numbers add up to 100. The second and fourth columns show accurate predictions, whereas the third and fifth columns show false predictions. Of particular interest is the second column, which shows the percentage of TP. As the balance between accurate values (TP+TN) and TP is important for choosing a threshold, we added the product of TP and TP+TN. This column shows an indicator of how good or bad the models perform when making correct predictions. From a user's perspective, the second (True Positive) and fifth (False Negative) columns refer to respondents who would receive a possible

Table 4.3. Confusion matrixes: percentages of truly and falsely predicted attrition and stay.

Threshold	True Positive	False Positive	True Negative	False Negative	Total	Product of TP and TP+TN
Model 1						
0.05	17	0	0	83	100	289
0.10	14	3	40	43	100	756
0.15	13	4	47	36	100	780
0.20	11	6	56	27	100	737
0.25	9	8	63	20	100	648
Model 2						
0.05	17	0	34	49	100	867
0.10	15	2	53	30	100	1020
0.15	14	3	60	23	100	1036
0.20	12	5	66	17	100	936
0.25	10	7	70	12	100	800
Model 3						
0.05	16	1	38	45	100	864
0.10	15	2	53	30	100	1020
0.15	14	3	62	21	100	1064
0.20	13	4	67	15	100	1040
0.25	12	5	71	12	100	996

Note: Comparison of true and false predictions of attrition (positive) and stay (negative). We estimated each model with five thresholds and show the percentages within each estimation. The product of TP and TP+TF is an indicator of the performance of each threshold which sets the two aims of maximizing the correct predictions of attrition and the correct predictions in general into relationship. TP= True Positive, TN= True Negative

intervention targeted at future attriters because the model predicts that they will attrit. Thus, the respondents in the second (True Positive) and third (False Positive) columns would need these interventions because they actually attrit. Accordingly, it is desirable to keep the number of respondents in the third (False Positive) column low. Moreover, the respondents of the fifth (False Negative) column would needlessly increase costs since these respondents are not attriters, and investing in keeping them in the panel would be unjustified.

The thresholds themselves are an arbitrary but necessary choice for planning an intervention. When choosing a threshold, survey conductors need to balance between reaching respondents who require an intervention on the one hand and avoiding unnecessary interventions on the other hand. When balancing between true outcomes and TP, we found that a threshold of 0.15 works quite well. This threshold has the highest product of TP and TP+TF. However, for other data, other thresholds may be more appropriate. With respect to the models' predictions, we found that the first model performed slightly worse but almost as well as the other two models. The TP was very similar among the models; however, Models 2 and 3 predicted more TN cases, thereby minimizing the share of respondents who are FN and who would have increased the costs of an intervention targeted at potential attritters. The first model predicted staying in a panel at a worse rate than did the models that take into account the factors of available time and survey motivation. We also found that most correct predictions can be attributed to respondents who stay in the panel. This finding indicates that the models' goodness-of-fit mostly relies on identifying who stays in a panel rather than who attrits from a panel. The predicted attritions also include a share of FP. On the other hand, we found that simple means – such as a longitudinal analysis of response times – can have a considerable effect and that our models have high positive predicted values (PPV) – that is, we can correctly predict most panel attrition (Model 1: 77%; Model 2: 82%; Model 3: 82%). The PPV represents the share of TP among all P. Knowledge about available time and survey motivation does not considerably improve the PPV.

To summarize the confusion-matrix results, the first model predicted attrition almost as well as the third model. This finding is a major advantage because we do not need to assess any variables for the first model. We further found that most of the goodness-of-fit for all models can be attributed to the TN. The share of the FP is low in all three models. This group comprises the respondents who – in case of an intervention targeted at potential attriters – would not receive an intervention even though an intervention could possibly motivate them to stay in a panel.

4.8. Conclusion and Discussion

In the present study, we investigated the longitudinal effect of response time on attrition. We focused on the relative frequency of late responses and on alternating between early, late, and nonresponse in an individual response-time pattern. We analyzed the effects of the aforementioned variables on attrition and found that a higher frequency of late responses in an individual response-time pattern is related to a higher likelihood of panel attrition (AME = 0.087). Moreover, stronger response-time habit over multiple waves was found to be related to a lower likelihood of attrition (AME = −0.013). We disentangled whether our models explain attrition or staying in a panel and found that a considerable amount of attrition can be predicted correctly by using only response time and response-time habit.

Our findings suggest that available time and survey motivation are not the primary variables that influence response time, which also appears to suggest that response time has an association with panel attrition. We found that all the indicators of available time and survey motivation that had an effect on panel attrition before controlling for response time also had an effect on attrition when controlling for response time. Therefore, it seems reasonable to assume that other variables influence both response time and attrition. Such (often-unmeasured) variables could include elements of a respondent's personality, such as their preference for finishing an undesirable task immediately or for delaying it. Another explanation could be that the variables we used to measure available time and survey motivation do not measure these items as expected.

The present study is not without limitations. Although the existing literature primarily uses the same operationalization of late response that we used, this is not the only way to operationalize response time. Future studies could compare the different methods of operationalizing response time and the effects of different operationalizations on data quality. Furthermore, we focused on variables that are associated with available time and survey motivation in order to control response time and thus did not examine other potential factors, such as survey-design features or respondents' personalities. Apart from the existence of other influences on panel attrition, further latent factors – such as time spent on hobbies – likely influence available time. In addition, our focus was on predicting attrition and developing an easily accessible method for applying response times in future studies. Therefore, we had to accept limitations in our explanation of attrition. In future studies, the effects of response time on attrition could potentially be selected and explained more accurately.

Despite these limitations, the present study provides high added value for survey methodologists and panel infrastructures. Survey methodologists can use our approach of operationalizing response-time habit in future studies on attrition and data quality in general. This approach shifts the focus from mere response times to using derived measures, such as response-time habit. This response-time habit has not been considered in previous research and can – as has been demonstrated – be a valuable predictor of attrition in future studies. Panel infrastructures can use response times and the measures derived from these times to identify panelists with a high probability of attriting and to initiate targeted interventions for this particular group of respondents within an adaptive survey design (Schouten et al., 2018). Targeting allows for a cost-efficient method of allocating resources where they are most needed. The present study reveals that the use of response time can identify more than 80% of respondents who attrit in advance.

4.9. References

Alwin, D. F. (2007). Margins of error: A study of reliability in survey measurement (Bd. 547). John Wiley & Sons.

Armenakis, A. A., & Lett, W. L. (1982). Sponsorship and follow-up effects on response quality of mail surveys. *Journal of Business Research*, *10*(2), 251–262. https://doi.org/10.1016/0148-29638290031-5

Boehmke, B., & Greenwell, B. (2019). *Hands-on machine learning with R*. Chapman and Hall/CRC.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review*, *36*(1), 103–115. https://doi.org/10.1177/0894439317697949

Cohen, S. B., Machlin, S. R., & Branscome, J. M. (2000). Patterns of survey attrition and reluctant response in the 1996 medical expenditure panel survey. *Health Services and Outcomes Research Methodology*, *1*(2), 131–148. https://doi.org/10.1023/A:1012543121850

Diaz de Rada, V. (2005). The Effect of Follow-up Mailings on The Response Rate and Response Quality in Mail Surveys. *Quality & Quantity*, *39*(1), 1–18. https://doi.org/10.1007/s11135-004-5950-5

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2016). *Internet, phone, mail and mixed-mode surveys: The tailored design method*. John Wiley & Sons.

Donald, M. N. (1960). Implications of nonresponse for the interpretation of mail questionnaire data. *Public Opinion Quarterly, 24*(1), 99. https://doi.org/10.1086/266934

Eckland, B. K. (1965). Effects of prodding to increase mail-back returns. *Journal of Applied Psychology, 49*(3), 165– 169. https://doi.org/10.1037/h0021973.

Eisnecker, P. S., & Kroh, M. (2016). The informed consent to record linkage in panel studies: Optimal starting wave, consent refusals, and subsequent panel attrition. *Public Opinion Quarterly, 81*(1), 131–143. https://doi.org/10. 1093/poq/nfw052

Esser, H. (1986). Über die Teilnahme an Befragungen. *Zuma Nachrichten, 10*(18), 38–47.

Esser, H. (2011). Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften? In G. Albert (Ed.) *Kölner Zeitschrift für Soziologie und Sozialpsychologie/Sonderhefte*, *50* (pp. 45–62). VS Verlag für Sozialwissenschaften.

Friedman, E. M., Clusen, N. A., & Hartzell, M. (2003). Better Late? Characteristics of Late Respondents to a Health Care Survey. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 992–98.

GESIS . (2018). *GESIS Panel - Extended Edition.* GESIS, Cologne. ZA5664 Datenfile Version 26.0.0. https://doi.org/10. 4232/1.13157

Gilbert, G. H., Longmate, J., & Branch, L. G. (o. J.). (1992). Factors influencing the effectiveness of mailed health surveys. *Public Health Reports*, *107*(5), 576-584.

Göritz, A. S. (2008). The long-term effect of material incentives on participation in online panels. *Field Methods, 20*(3), 211–225. https://doi.org/10.1177/1525822X08317069

Green, K. E. (1991). Reluctant Respondents: Differences between Early, Late, and Nonresponders to a Mail Survey. *The Journal of Experimental Education*, *59*(3), 268–276. https://doi.org/10.1080/00220973.1991.10806566

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *The Public Opinion Quarterly, 64*(3), 299–308. https://doi.org/10.1086/317990

Groves, R. M., Floyd, J. F., Couper, M. P., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed). John Wiley & Sons.

Gummer, T., & Struminskaya, B. (2020). Early and Late Participation during the Field Period: Response Timing in a Mixed-Mode Probability-Based Panel Survey. *Sociological Methods & Research, 0*(0), 1-24. https://doi.org/10.1177/0049124120914921

Haunberger, S. (2011). Explaining unit nonresponse in online panel surveys: An application of the extended theory of planned behavior 1. *Journal of Applied Social Psychology, 41*(12), 2999–3025. https://doi.org/10.1111/j.1559-1816. 2011.00856.x

Helasoja, V. (2002). Late response and item nonresponse in the Finbalt Health Monitor Survey. *The European Journal of Public Health*, *12*(2), 117–123. https://doi.org/10.1093/eurpub/12.2.117

Herzing, J. M., & Blom, A. G. (2019). The influence of a person's digital affinity on unit nonresponse and attrition in an online panel. *Social Science Computer Review*, *37*(3), 404–424. https://doi.org/10.1177/0894439318774758

Kreuter, F., & Casas-Cordero, C. (2010). Paradata. *RatSWD Working Paper Series*. Working paper Berlin, Germany: German Council for Social and Economic Data (RatSWD).

Kreuter, F., Muller, G., & Trappmann, M. (2010). Nonresponse and measurement error in employment research: making use of administrative data. *Public Opinion Quarterly*, *74*(5), 880–906. https://doi.org/10.1093/poq/ nfq060

Kreuter, F., Müller, G., & Trappmann, M. (2014). A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents. *Sociological Methods & Research*, *43*(3), 452–464. https://doi.org/10.1177/0049124113508094

Kroneberg, C. (2014). Frames, scripts, and variable rationality: An integrative theory of action. In G. Manzo (Ed.) *Analytical Sociology: Actions and networks* (pp. 97-123). John Wiley & Sons. https://doi.org/10.1002/9781118762707.ch04

Kunz, F. (2010). Follow-Ups in Mail Surveys. Empirical Findings on the Effects on the Response Rate, the Responses and the Sample Composition. *methods, data, analyses,* 4(2), 127-155. https://doi.org/10.12758/mda.2010.009.

Lugtig, P. (2014). Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers. *Sociological Methods & Research*, *43*(4), 699–723. https://doi.org/10.1177/0049124113520305

Lynn, P. (2009). *Methodology of longitudinal surveys*. John Wiley & Sons.

Nicoletti, C., & Peracchi, F. (2005). Survey response and survey characteristics: microlevel evidence from the European community household panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(4), 763–781. https://doi.org/10.1111/j.1467-985X.2005.00369.x

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, *70*(5), 737–758. https://doi.org/10.1093/poq/nfl038

Olson, K. (2013). Do non-response follow-ups improve or reduce data quality?: A review of the existing literature. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 129–145. https://doi.org/10.1111/j.1467-985X.2012.01042.x

Preisendörfer, P., & Wolter, F. (2014). Who Is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys. *Public Opinion Quarterly*, *78*(1), 126–146. https://doi.org/10.1093/poq/nft079

Roßmann, J., & Gummer, T. (2016). Using Paradata to Predict and Correct for Panel Attrition. *Social Science Computer Review*, *34*(3), 312–332. https://doi.org/10.1177/0894439315587258

Schnell, R. (2013). *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Springer.

Schouten, B., Peytchev, A., & Wagner, J. (2018). *Adaptive survey design*. CRC Press.

Skarbek-Kozietulska, A., Preisendörfer, P., & Wolter, F. (2012). Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. *Zeitschrift für Soziologie*, *41*(1), 5-23. https://doi.org/10.1515/zfsoz-2012-0103

Smith, T. W. (2011). Refining the total survey error perspective. *International Journal of Public Opinion Research*, *23*(4), 464–484. https://doi.org/10.1093/ijpor/edq052

Struminskaya, B. (2014). *Data quality in probability-based online panels: Nonresponse, attrition, and panel conditioning*. Utrecht University.

Tancreto, J. G., & Bentley, M. (2005). Determining the effectiveness of multiple nonresponse followup contact attempts on response and data quality. In *Proceedings of the Survey Research Methods Section of the American Statistical Association,* 3626-3632.

Voigt, L. F., Koepsell, T. D., & Daling, J. R. (2003). Characteristics of telephone survey respondents according to willingness to participate. *American Journal of Epidemiology*, *157*(1), 66–73. https://doi.org/10.1093/aje/kwf185

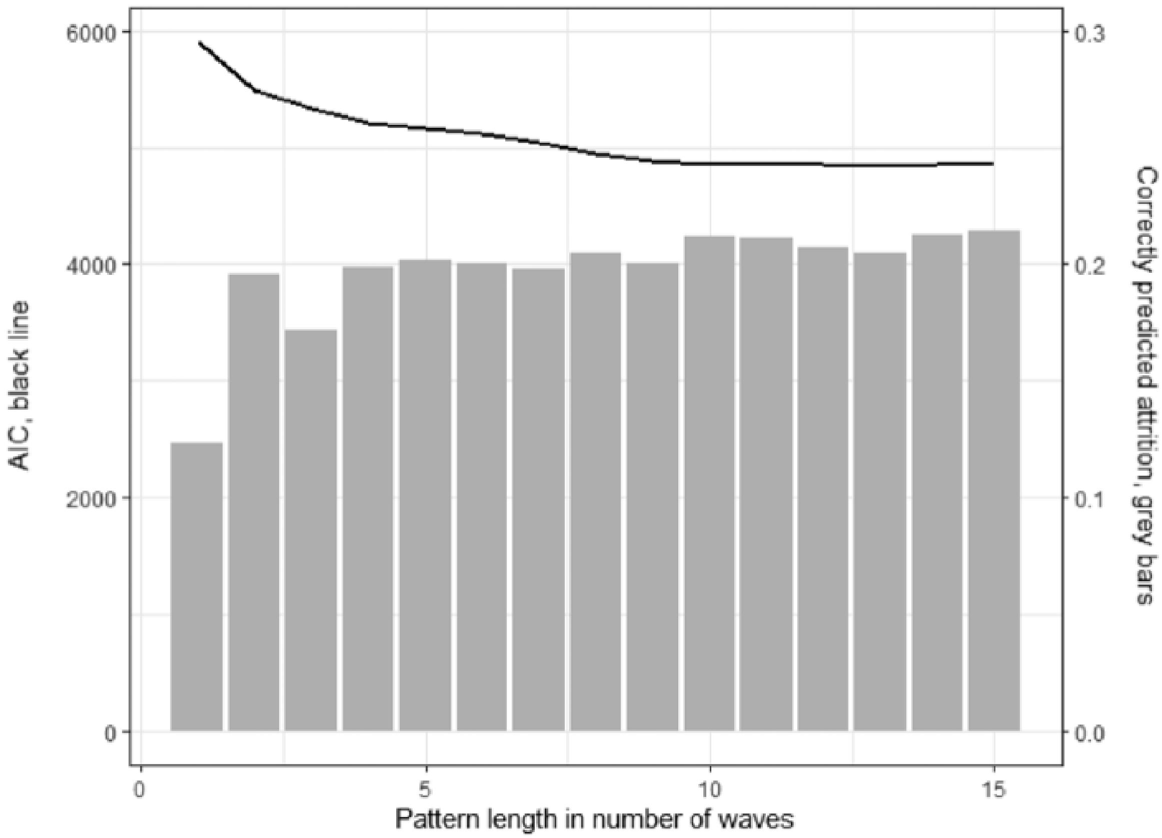
Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.) *Methodology of Longitudinal Surveys 1* (pp. 157–182). John Wiley & Sons.

4.10. Appendix

To determine the optimal pattern length, we calculated the Akaike Information Criterium (AIC) as a first indicator, and compared the predicted attrition to the actual attrition of each model as a second indicator. The first indicator, the AIC, describes the goodness of fit of a model. It enables the comparison of different models. A model with a lower AIC is superior to a model with a higher AIC. Our second indicator is the actual number of attritions that each model predicts correctly. The models predict the likelihood of each panelist to attrite, which enables us to compare the estimated prediction of attrition or stay with the actual attrition or stay in our data. Hence, this comparison showed us which of the models with different numbers of waves is able to predict attrition close to the actual behavior. Thus, the available data enabled us to verify whether the predictions of our models were reliable. We used the correctly predicted attritions (CPA) as an indicator because this value is particularly precious. The minimum of this value is 0, whereas the maximum is equivalent to the percentage of respondents who attrited. In this case, 26\% of the respondents attrited. Hence, the maximum is 0.26. Respondents who remain in the panel are not included in the CPA.

Figure A.4.1 shows the results of our estimations for the response time pattern length needed to investigate attrition. The x-axis shows the maximum number of waves included in the response time pattern that was used to calculate the model. The left y-axis refers to the line plot that is the AIC of each of the models. From the line plot, we see that the AIC decreases with a growing number of waves. When 10 waves are included in the response time pattern, the AIC stabilizes and does not decrease any more. Hence, this indicator shows that from 10 waves on, increasing the number of waves in the pattern does not increase the quality of the model. The right y-axis refers to the bar plot that visualizes the correctly predicted attrition of each model. The correctly predicted attrition strongly increases up to five waves in the response time pattern. When a pattern includes five to nine waves, the predicted attrition is almost as good as in the case of using 16 waves. However, the correctly predicted attrition slightly increases when at least 10 waves are used. Based on these indicators, we found that a pattern length of 10 is optimal. Longer response time patterns do not improve the models. Shorter response time patterns do not perform as well in the models as patterns that were built from 10 waves.

Figure A.4.1. Estimation of the optimal pattern length.



5. Conclusion

Panel surveys remain an important tool for assessing opinions and behaviors in society. However, one downside of these surveys is that they are only as powerful as the quality of their data, which is influenced by many factors. Examples of these factors are the sample, survey mode, and aspects of the questionnaire, such as question wording or scale use. Survey infrastructures can control most of these factors, nevertheless, in some respects, data quality depends on the cooperation of the panelists, their participation and the accuracy of their responses. This dissertation aimed to investigate the role of participation time — as an easily available indicator — in the identification and prediction of panelists' cooperation. In the three separate studies provided in Chapters 2, 3, and 4, I investigated the relationship between participation time and panel attrition as well as data quality indicators. Following is a summary of the results found in these chapters and what these results imply for survey infrastructures and researchers in the field of survey methodology. I conclude with the limitations of this dissertation.

In Chapter 2, I focused on the topic of data quality. Since participation time is considered a promising predictor of data quality, I investigated whether this easily available variable is related to data quality by analyzing 16 data quality indicators. I estimated 16 regression models that differed in their explanatory variable and were, depending on the distribution of the explanatory variable, logistic regressions, poisson regressions, negative binomial regressions, zero-inflated poisson regressions, or zero-inflated negative binomial regressions. I found relationships between participation time and data quality for some indicators, but not all. The relationships were diverse in their direction: whether the relationship was positive or negative, and their functional form: whether the relationship was linear, and where peaks were in the relationship.

In Chapters 3 and 4, I focused on panel attrition as a particularly severe problem for panel surveys. In Chapter 3, I addressed the following research questions: How are the different operationalizations of participation time related to panel attrition? Do the models with different operationalizations predict attrition similarly well or are there especially fitting operationalizations? I found that the different measurements of participation time mostly predicted panel attrition similarly well, even though some operationalizations of participation time were not related to panel attrition, such as the operationalization of participation time as whether or not a respondent was among the first 5% of the online respondents and the operationalization of participation time as whether or not a respondent was among the first 10% of the mail respondents. Also striking was the finding that online respondents who participated on the first day, among the first 10% and the first 50%, were more likely to attrite than later respondents. A similar finding applied to mail respondents who participated among the first 50%. The operationalizations of participation time as a metric share of prior respondents and as a response before or after seven days after the field start worked best for online respondents. The operationalization of participation time as a metric count of days until response worked best for mail respondents. However, the differences were minimal for both modes.

In Chapter 4, I found that a higher share of late responses and lower habitual behavior was related to a higher likelihood of panel attrition. Participation time and habitual participation time do not explain the same variance in attrition as available time and survey motivation. The regression model that included participation time and participation time habit predicted 77% of the attrition, which makes participation time an easily available tool for targeting potential attrition.

Based on these results, six crucial findings are summarized, and their implications described, in the following. After this summary, I conclude with a discussion of the limitations of this dissertation.

1. Data quality indicators do not equal one another.

In Chapter 2, I showed that participation time can have various relationships with different data quality indicators. Thus, data quality indicators cannot be used interchangeably. When survey infrastructures aim to target respondents to achieve the best data quality, they should be concerned with the data quality indicators that are of the most interest, rather than the indicators that are easily measurable. Multiple indicators should be used if there is no particular data quality indicator of interest. The diverse relationships between participation time and data quality indicators also may be found when using other predictors to explain data quality. Survey methodologists should be aware of this and use multiple data quality indicators when making statements about data quality. Future analyses could address the commonalities and differences in data quality indicators and attempt to build factors from multiple indicators.

2. Late respondents do not always provide poor data.

Chapter 2 showed that late respondents do not necessarily provide data of lesser quality and may even provide higher data quality than earlier respondents. Based on this information, survey infrastructures could increase or reduce their field time to include the respondents who optimize on the data quality indicators of most interest and reduce the share of respondents who satisfice on these indicators. Survey methodologists could dig deeper into the reasons for late responses and determine whether the different causes of early or late participation time are related to data quality.

3. We can predict panel attrition using participation time.

In Chapter 4, I showed that most of the attrition can be predicted using participation time. This finding offers survey infrastructures the opportunity to cost-effectively target panelists who are most likely to attrite. Most notably, the share of attritors that remain unidentified is small, which suggests that survey infrastructures do not need to rely on predictors that are harder to measure when aiming to identify future attritors. Survey methodologists could not only use participation time, but also other time-related data, such as the time spent on the internet or screen-use time of smartphones for survey methodological analyses.

4. Participation time can be measured in several ways that all predict panel attrition equally well.

In Chapter 3, I compared multiple operationalizations of participation time and found that different operationalizations predicted panel attrition equally well. This result suggests that survey methodologists do not have to worry about the operationalization of participation time or its comparability. Instead, they can use the operationalization that suits them best and, when researching panel attrition, can expect that a different operationalization would have led to a similar result. This result also shows that findings from studies that have used different operationalizations are comparable with respect to the operationalization of participation time.

5. Participation time habit built from participation time is a powerful predictor of attrition.

In participation time habit, I found a variable that can be calculated from participation time that predicts panel attrition quite well (Chapter 4). Survey infrastructures could use this variable for the identification of future attrition. This finding is particularly interesting for survey methodologists because it represents a first step for using participation time to create an innovative measure of participation behavior. Survey methodologists could think of further applications of participation time, such as measuring not only the stability of participation time but also the differences in participation time in multiple surveys. For example, researchers could investigate whether panelists who respond faster than in more recent waves have a lower likelihood of attrition than those who become slower.

6. The relationship between participation time and panel attrition cannot be explained by available time and survey motivation.

In Chapter 4, I also showed that the relationships between available time and panel attrition and survey motivation and panel attrition stayed mostly the same under control for participation time. This finding suggests that participation time is unlikely to mediate these relationships. Survey methodologists can benefit from this finding because it implies that other sources might be responsible for the association between participation time and panel attrition. Future work in survey methodology could continue working on this topic by addressing theoretical reasons why participation time could be related to attrition, for example, psychological attitudes to tasks.

This dissertation is not without its limitations. First, the panelists in the GESIS Panel are a special sample with high survey experience, since they have been committed to the panel for a long time. Therefore, the results drawn from this dissertation cannot necessarily be transferred to the general population, although evidence is based on a probability-based sample. Overcoming this limitation is difficult, especially when panel attrition is of interest, because working with a panel always implies that panelists have a certain amount of experience. Since reasons for panel attrition may change throughout a panel study, it is essential to use some panelists with longer and some with shorter participation histories for the analysis. Second, panel attrition and differences in data quality can have a vast number of causes. Because this work focused on developing a method to identify both problems, it fell short of identifying the reasons for panel attrition and differences in data quality. Similarly, an explanation is still lacking as to why participation time performs as a good predictor of panel attrition and why the relationships between participation time and the different data quality indicators are so varied. While I aimed at the practical application of participation time and could only provide potential explanations for these discrepancies, future research should empirically test the predictors of participation time with a more theoretical perspective.

Overall, this dissertation investigated the measurements and consequences of participation time in panel surveys. Concerning the measurements, this dissertation shows that it is mainly irrelevant which measurement of participation time is chosen when investigating panel attrition.

To predict attrition, survey agencies can draw on the measurement of participation time that they prefer. Concerning the consequences, I investigated lower data quality and panel attrition as two negative outcomes of the response process. This work showed that participation time can be an important tool for survey methodologists and survey infrastructures, since participation time can identify most future attrition and, therefore, can help when panelists at risk of attrition should be treated with targeted incentives. Nevertheless, this dissertation also showed the limits of participation time. With respect to predicting data quality, the relationship with participation time only applies to some indicators. Thus, research on the relationship of participation time and data quality still has to be deepened.