Now, later, or never? Using response-time patterns to predict panel attrition

Minderop, Isabella; Weiß, Bernd

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - Projektnummer 491156185 / Funded by the German Research Foundation (DFG) - Project number 491156185

International Journal of Social Research Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsrm20

Now, later, or never? Using response-time patterns to predict panel attrition

Isabella Minderop & Bernd Weiß

To cite this article: Isabella Minderop & Bernd Weiß (2022): Now, later, or never? Using response-time patterns to predict panel attrition, International Journal of Social Research Methodology, DOI: 10.1080/13645579.2022.2091259

To link to this article: https://doi.org/10.1080/13645579.2022.2091259

INTERNATIONAL JOURNAL OF SOCIAL RESEARCH METHODOLOGY
https://doi.org/10.1080/13645579.2022.2091259

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Now, later, or never? Using response-time patterns to predict panel attrition

Isabella Minderop and Bernd Weiß

Leibniz Institute for the Social Sciences, GESIS, Mannheim, Germany

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 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.

ARTICLE HISTORY
Received 02 September 2021
Accepted 14 June 2022

KEYWORDS
Reluctant response; data quality; panel survey; paradata; late response

Introduction

Panel attrition is a central problem in longitudinal surveys. Indeed, it is a special case of unit nonresponse (Alwin, 2007, p. 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. Eisnecker & Kroh, 2016; Herzing & Blom, 2019; Lugtig, 2014; Roßmann & Gummer, 2016; Struminskaya, 2014).

A good predictor of panel attrition is response behavior in previous waves, such as participation or nonresponse (Lugtig, 2014; Roßmann & Gummer, 2016). 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

CONTACT Isabella Minderop isabella.minderop@gesis.org GESIS, Leibniz Institute for the Social Sciences c/o Isabella Minderop, B6 4-5, Mannheim 68159, Germany

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 (Lugtig, 2014; Nicoletti & Peracchi, 2005; Roßmann & Gummer, 2016). 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.

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 (Göritz, 2008; Haunberger, 2011; Nicoletti & Peracchi, 2005; Roßmann & Gummer, 2016).

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 and 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; Armenakis & Lett, 1982; Eckland, 1965; Gilbert et al., 1992; Kreuter et al., 2010). 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 and 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.

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 (Esser, 2011; Kroneberg, 2014) 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.

Data

The data analyses 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.[1] 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).[2] 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 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,[3] 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 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.

Table 1. Descriptive statistics.

			Overall (n = 5340)		Participant (n = 4429)		Attrited (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

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

Method

One 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. 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 or whether the effects are cumulative.

For each model and respondent, we calculated the predicted probability that this 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).

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.[4] 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 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

Table 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 R^2	0.10	0.25	0.29
AIC	4416.8	3663.9	3497.9
n	5340	5340	5340

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

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 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 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 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.

Table 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

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

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.

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.

Notes

1. 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.
2. 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).
3. 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.
4. 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 ($p < 0.001$).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Isabella Minderop is research associate at GESIS - Leibniz Institute for the Social Sciences. Her research interests lie in survey methodology, especially paradata and data quality.

Bernd Weiß is head of the GESIS Panel, a probabilistic mixed-mode access panel. He also serves as Deputy Scientific Director of the Department Survey Design and Methodology at GESIS – Leibniz Institute for the Social Sciences in Mannheim. His research interests range from survey methodology, research synthesis, open science to family sociology and juvenile delinquency.

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 & 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. https://scholar.google.com/scholar_lookup?hl=en&publication_year=2009b&author=D.+A.+Dillman&author=J.+D.+Smyth&author=L.+M.+Christin&title=Internet%2C+Mail+and+Mixed%E2%80%90Mode+Surveys%3A+The+Tailored+Design+Method

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. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210300

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

Friedman, E. M., Clusen, N. A., & Hartzell, M. (2003). Better Late? Proceedings of the Survey Research Methods Section of the American Statistical Association, 992–98. https://ideas.repec.org/p/mpr/mprres/5b036c2fea4745a4b941073c30ad8f23.html

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*, *9*(5), 576. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1403701/

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., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.

Gummer, Tobias , & Struminskaya, Bella . (2020). Early and late participation during the field period: response timing in a mixed-mode probability-based panel survey. *Sociological Methods & Research*, 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, *Analytical Sociology: Actions and Networks*, pp. 95–123. John Wiley and Sons.

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*, *30*, 127–155. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210146

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: *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

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

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 / to deny or to confess? Determinants of truthful answers in surveys. *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* [Doctoral thesis]. Utrecht University. https://dspace.library.uu.nl/handle/1874/301751

Tancreto, J. G., & Bentley, M. (2005). Determining the Effectiveness of Multiple Nonresponse Followup Contact Attempts on Response and Data Quality. Proceedings of the Survey Research Methods Section of the American Statistical Association, 3626–3632 http://www.asasrms.org/Proceedings/y2005/files/JSM2005-000340.pdf

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. *Methodology of Longitudinal Surveys*, *1*, 157–182.

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 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.

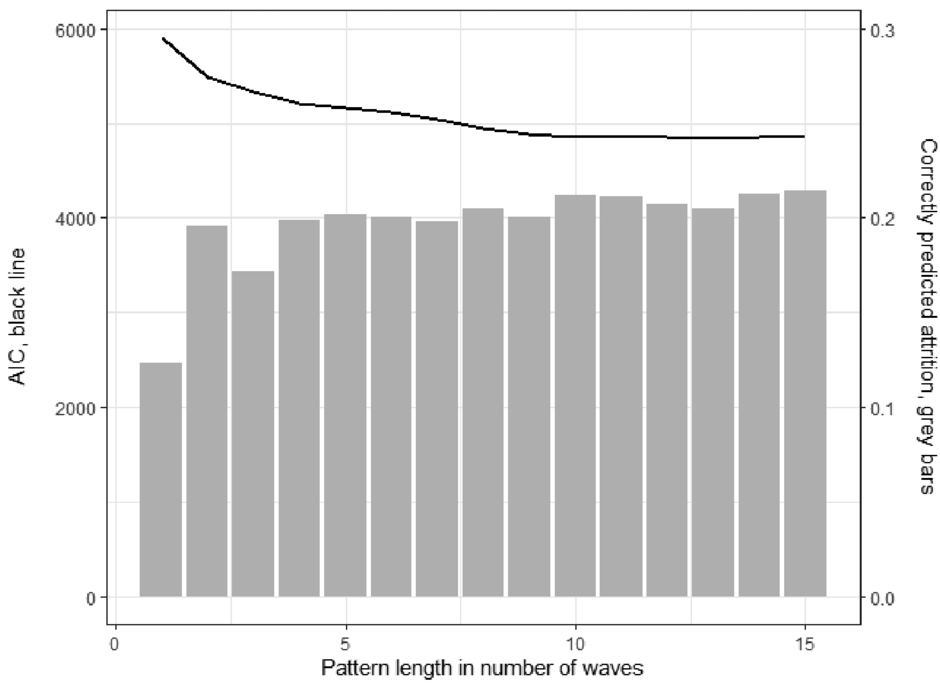

Figure 1. Estimation of the optimal pattern length.