A data-driven approach to monitoring data collection in an online panel

Jessica M. E. Herzing*

LINES/FORS, University of Lausanne, Switzerland, and University of Mannheim, Germany

jessica.herzing@unil.ch

Caroline Vandenplas

KU Leuven, Belgium

Julian B. Axenfeld

MZES, University of Mannheim, Germany

Abstract

Longitudinal or panel surveys suffer from panel attrition which may result in biased estimates. Online panels are no exceptions to this phenomenon, but offer great possibilities in monitoring and managing the data collection phase and response-enhancement features (e.g., reminders), due to real-time availability of paradata. This paper presents a data-driven approach to monitor the data collection phase and to inform the adjustment of response-enhancement features during data collection across online panel waves, which takes into account the characteristics of an ongoing panel wave. For this purpose, we study the evolution of the daily response proportion in each wave of a probability-based online panel. Using multilevel models, we predict the data collection evolution per wave day. In our example, the functional form of the data collection evolution is quintic. The characteristics affecting the shape of the data collection evolution are characteristics of the specific wave day and not of the panel wave itself. In addition, we simulate the monitoring of the daily response proportion of one panel wave and find that the timing of sending reminders could be adjusted after 20 consecutive panel waves to keep the data collection phase efficient. Our results demonstrate the importance of re-evaluating the characteristics of the data collection phase, such as the timing of reminders, across the lifetime of an online panel to keep the fieldwork efficient.

Keywords: reminder procedure, response rates, web survey, fieldwork, panel attrition

Introduction

Longitudinal or panel surveys offer analytic benefits for social science research with regard to gaining knowledge about causes and effects of individual changes, differentiation between age and cohort effects, and the investigation of measurement errors (see Andreß, Golsch, & Schmidt, 2013; Elder & Giele, 2009; Firebaugh, 2008; Glenn, 2005; Halaby, 2004). Since, panel surveys rely upon the same sample units at certain points in time (consecutive panel survey waves), an important source of error in panel surveys is panel attrition (for examples see Behr, Bellgardt, & Rendtel, 2005; Cheng, Zamarro, & Orriens, 2018; Das, Toepoel, & van Soest, 2011; Dennis & Li, 2003; Frick, Grabka, & Groh-Samberg, 2012; Lugtig, 2014; Lynn, 2009). The term panel attrition covers panel survey members that are no longer able to participate in consecutive panel waves (due to, changing contact details, refusal, incapacity or death, see Watson & Wooden, 2009) and it adds up to unit nonresponse of sample units in the first panel wave.

If the variables of interest are correlated with panel members' response propensities, then estimates of these variables of interest might be biased (see Bethelehem, 2002; Groves, 2006; Groves & Peytcheva, 2008). Biased estimates are a threat for drawing inference in substantive research. Hence, numerous methods to correct for these nonresponse biases have been developed (for examples see Frick et al., 2012; Rubin, 1987; Rosenbaum & Rubin, 1983; Roßmann & Gummer, 2016; Vandecasteele & Debels, 2007). However, as Allison (2001, p. 5) states the best solution to the missing data problem is prevention. Consequently, survey practitioners implement response-enhancement features during data collection to prevent panel attrition a priori and hence, to avoid potential nonresponse bias in the first place.

As aforementioned, panel attrition can result in nonresponse bias when sample units with specific characteristics systematical fail to respond to a survey request (for an explanation see Billiet, Philippens, Fitzgerald, & Stoop, 2007). Nonresponse bias partly depends on response rates as well as covariances between response propensities and survey variables (see Bethelehem, 2002; Billiet et al., 2007; Groves, 2006; Groves & Peytcheva, 2008). As response rates are easy to calculate in real-time, the evolution of response rates is often used as a data quality indicator (Stoop, 2005) to signal problems during the data collection (for examples of data collection monitoring see Laflamme, Maydan, & Miller, 2008; Malter, 2013; Schouten & Shlomo, 2017). To complement the evolution of response rates as a data quality indicator, survey organizations often investigate socio-demographic or other variables (for examples see Couper, Kapteyn, Schonlau, & Winter, 2007; Kaplowitz, Hadlock, & Levine, 2004). Hence, one can monitor the evolution of socio-demographic or other key survey variables to monitor potential nonresponse biases, in addition to monitoring the evolution of response rates.

The data collection evolution and as a consequence, panel attrition is influenced by various survey characteristics, such as the day a survey is launched, invitation letters, the announced survey topic, the announced survey length, incentives, pre-notifications, and reminders (see Fan & Zheng, 2010; Groves, Presser, & Dipko, 2004; Göritz, 2014; Liu & Wronski, 2018; Vehovar, Batagelj, Lozar Manfreda, & Zaletel, 2002; Weible & Wallace, 1998). All these survey characteristics are usually kept constant across the lifetime of a panel survey with the initial purpose to enhance responses (for examples see Blom et al., 2016). However, it needs to be studied whether the effects of reminders remain stable across the lifetime of a panel (see also Göritz & Crutzen, 2012, p. 248). Yet the implementation of the aforementioned response-enhancement features is often driven by the practical needs to reduce costs, and to increase response rates (Groves & Heeringa, 2006; Göritz & Crutzen, 2012) and fail to consider the performance and characteristics of the ongoing data collection (considering survey characteristics when introducing reminders is suggested by Couper, 2008, p. 341; Dillman, Smyth, & Christian, 2014, p. 336-337). However, response-enhancement features should be planned strategically and should consider survey characteristics (in line with the principle of responsive and adaptive survey design for examples see Groves & Heeringa, 2006; Wagner et al., 2012; Schouten, Peytchev, & Wagner, 2017). In this regard, it is important to know which response-enhancement features may have an influence on data collection efficiency in terms of panel attrition and whether the efficiency of response-enhancement features changes across panel waves.

To remedy this lack of consideration of data collection efficiency in panel surveys, we propose to model the evolution of the data collection across panel waves. Vandenplas and Loosveldt (2017) propose to increase response rates by rendering data collection more efficient. For this purpose, they model the response proportion - number of completed interviews/questionnaires divided by the sample size - per time unit (e.g., month, week, day, hour) with the aim to understand the evolution of the data collection efficiency and the factors that can influence it (Vandenplas & Loosveldt, 2017). In the following, we transfer this approach to a probability-based online panel.

Online panels are well suited for monitoring the data collection evolution, due to the real-time availability of paradata - and more specifically, data about the surveys' fieldwork progression and the data collection evolution. Furthermore, it is possible to adjust response-enhancement features on short notice in online panels. Moreover, online panels have the advantage that the survey request/invitation letters are kept the same for all respondents over time and that the target population stays unchanged over time. In consequence, there is no or little variation of unknown or difficult to control survey characteristics over waves in an online panel. This allows controlling for many known survey characteristics when modeling the data collection evolution. However, panel attrition (also referred to as panel fatigue, see Behr et al., 2005; Dennis & Li, 2003) can change the sample composition. A change in sample composition (e.g., an increase of elderly respondents) can affect the data collection efficiency, as the aim to reach the target population (e.g., the general population) might be achieved earlier than in previous waves or vice versa. Thus, a reconsideration of data collection efficiency and response-enhancement features across the lifetime of a panel might be valuable.

Methods

Data

The data used in this study comes from the German Internet Panel (GIP). The GIP is a probability-based online panel (for further information see Blom, Gathmann, & Krieger, 2015), as the sample is based on face-to-face recruitment interviews and is representative for the German population aged 16-75 (see Blom et al., 2017). Persons without internet and/or computer are provided with the necessary equipment to enable them to participate in the online panel. The panel is conducted every second month and collects panel members' attitudes and opinions about political, social and economic issues.

The empirical analysis was performed on the data collected between November 2014 and March 2018, which results in 21 panel waves (Blom et al., 2018). We chose to work with the refreshment sample of the GIP that started in September 2014 as we wanted a homogeneous group of respondents with the same amount of panel experience. Furthermore, we excluded the September 2014 wave from the analysis because this wave had a longer data collection phase than all other panel waves, due to the recruitment of new respondents via face-to-face interviews. In November 2014, a total of 2,064 panel members were invited (Blom et al., 2018). Over the 21 panel waves the invitation e-mail, the three reminder e-mails for nonrespondents (reminder 1 = second Friday of a month; reminder 2 = third Friday of a month; reminder 3 = Tuesday after the third Friday of a month), the start of the wave (first day of a month) and the end of the wave (last day of a month), and the timing and amount of incentives were kept constant (paid in May and November).

Depending on the length of a month the data collection phase was between 30 and 31 days[i] long; and depending on the weekday when the survey was launched the first reminder was sent between day 6 and day 12, the second reminder was sent between day 13 and day 19, and the third reminder was conducted via telephone between day 17 and day 23 of the data collection.

Comparing the GIP to other population-based online panels (the Longitudinal Internet Studies for the Social sciences (LISS panel), Étude Longitudinale par Internet Pour les Sciences Sociales (ELIPSS panel), and the GESIS Panel) shows that online panels mostly agree on a common practice on sending reminders (Blom et al., 2016). All four panel surveys field their waves monthly or every two month with a field-period of one or two months. They send their first reminders one week or two weeks after launching a wave. However, only the GIP uses the third reminder. Concerning the timing of reminders, slight differences between the GIP and other panel surveys remain. For instance, the GESIS panel sends its two reminders one week (day 7) and two weeks (day 14) after the survey was fielded (Bosnjak et al., 2018, p. 108).

Analytical approach

To investigate the data collection efficiency and how response-enhancement features - in our case reminders - increase responses to an online panel (defined as filling out the whole questionnaire), we model the daily response proportion - the proportion of completed questionnaires in a given day divided by the total number of invited panelists. To model the daily response proportion, we consider a multilevel model with two levels[ii]: days of data collection (referred to as wave day; level one) are nested within panel waves (level two). The dependent variable, the response proportion per day, is modeled in terms of elapsed time since the start of the data collection, expressed in days, which we consider as a continuous variable.

First, we examine the shape of the data collection evolution (as proposed by Vandenplas & Loosveldt, 2017). This allows us to understand key features of the evolution of the data collection, such as how long does the daily response proportion decrease, increase or when does the increase or decrease level off. Then, we attempt to understand the characteristics that influence the shape of the data collection evolution by introducing panel wave and day characteristics to explain some of the between wave variance (random intercept) and within wave variance (residual variance).

Possible covariates of the first level of the multilevel model are characteristics of the day, such as weekday, holidays, the day of reminder; whereas possible level two variables are panel wave characteristics such as survey length of the previous wave, starting weekday, or satisfaction with the previous wave. Finally, we use the shape of data collection evolution to monitor the data collection evolution of one specific panel wave (see Vandenplas, Loosveldt, & Beullens, 2017).

The shape of the data collection evolution

To understand the shape of the evolution of the GIP data collection, the daily response proportion is defined as the number of completed interviews of a specific day of data collection in one specific wave, divided by the total number of panelists that have been invited to participate in that specific wave separately for waves 1 to 21. The specific days of data collection are considered as a repeated measurement (30 measurements) within each wave and hence, we have 30 daily response proportions for each wave.

To model the shape of the data collection evolution, we estimate a multilevel model with a random intercept (β_{0w}) that allows the response proportion on the first day of data collection (d = 0) to vary between panel waves (w). Further, the daily response proportion is expressed as a polynomial of the days elapsed since the first day of data collection. For this purpose, we built the model in multiple steps to find the functional form of the data collection evolution.

To find the functional form of the data collection evolution, a random intercept with a linear slope ($\beta_{1w}d$) is estimated, then a quadratic function ($\beta_{2w}d^2$), a cubic function ($\beta_{3w}d^3$), a quartic function ($\beta_{4w}d^4$), a quintic function ($\beta_{5w}d^5$), and further polynomial functions (sextic, and so forth) of the days elapsed since a panel wave was launched are added step by step to the multilevel model.

All slopes of the functional form of the data collection evolution ($\beta_{1w}, \beta_{2w}, \beta_{3w}, \beta_{4w}, \beta_{5w}, \dots$) are first specified as fixed slopes and then as random slopes reflecting that the shape of the evolution can vary from wave to wave. Thus, the *basic model* describes the evolution of the daily response proportion during the data collection phase given wave days (d = 1, ⋯, 30) and waves (w = 1, ⋯, 21):

$$RP_{wd} = \beta_{0w} + \beta_{1w}d + \beta_{2w}d^2 + \cdots + \beta_{nw}d^n + \varepsilon_{wd}, \qquad (1)$$

$$\beta_{0w} = \gamma_{00} + u_{0w},$$

$$\beta_{nw} = \gamma_{n0} + u_{nw},$$

where RP_{wd} represents the response proportion in wave (*w)* at day (*d)*; β_{0w} represents the random intercept with fixed part γ_{00} and random part u_{0w}; β_{nw} represents the random slope of the linear function (*n* = 1), the quadratic function (*n* = 2), and higher polynomial functions of wave day d^n (*n* = 3, $\cdots$, *n*) with the fixed part γ_{n0} and the random part (wave specific) u_{nw}. ε_{wd} represents the residuals at level one, the specific days of a wave (with $\varepsilon_{wd} \sim N(0, \sigma_\varepsilon^2)$, and $cov(\varepsilon_{wd}, \varepsilon_{wd'}) = 0$). The level two covariance matrix of Equation 1 can be parametrized by

$$\begin{bmatrix} \sigma_0^2 & \cdots & \sigma_{0n} \\ \vdots & \ddots & \vdots \\ \sigma_{n0} & \cdots & \sigma_n^2 \end{bmatrix}.$$

This covariance structure reflects the assumption that the covariance between the response proportion of two consecutive days are fully captured by the functional form, as the variance of the residuals is fixed, and the covariance is zero. Furthermore, no restriction is imposed on the level two covariance matrix. A full-unstructured covariance matrix, with non-zero covariance between the residual errors at the day-level, might be more accurate. However, we do not have enough data to estimate such a model (this also depends on the number of independent variables and the algorithm used for the estimation), which opens scope for new research.

Characteristics influencing the shape of the data collection evolution

To explore which survey characteristics influence the shape of the evolution of the daily response proportion, we consider two types of variables: (1) variables that describe characteristics of the day (level one), and (2) variables that describe characteristics of the wave (level two). The variables on level one comprise the weekday, whether the first, the second or the third reminder was sent on a specific weekday (e.g., Monday, Tuesday, and so on), and whether the day of data collection was a public holiday or a school holiday in Germany. The variables of level two entail wave characteristics such as the weekday on which the data collection of a specific wave started, the length of the previous wave and a satisfaction indicator of the previous wave.

Equation 2 represents the basic model of Equation 1 with the additional functions for covariates for wave day and wave. The wave day variables were introduced to explain the residual (within) variance (ε_{wd}). The wave variables were introduced to explain the variance of the random intercept (u_{0w}) and the random slopes (u_{nw}). Hence, the final model that we consider is as follows:

$$RP_{wd} = \beta_{0w} + \beta_{1w}d + \beta_{2w}d^2 + \cdots + \beta_{nw}d^n + \beta_{n+1}x_1 + \cdots + \beta_{n+c}x_c + \varepsilon_{wd}, \qquad (2)$$

$$\beta_{0w} = \gamma_{00} + \gamma_{01}z_1 + \cdots + \gamma_{0s}z_s + u_{0w},$$

$$\beta_{nw} = \gamma_{n0} + \gamma_{n1}z_1 + \cdots + \gamma_{ns}z_s + u_{nw},$$

where RP_{wd} represents the response proportion in wave *w* at day d. β_{0w} is the random intercept with the fixed part defined as the fixed intercept γ_{00}. γ_{0s} represents the effect of the wave variables z_s ($s = 1, \cdots, n$), and random part u_{0w}. Furthermore, β_{nw} represents the random slope of the linear (n = 1), quadratic (n = 2), and the higher polynomial functions of wave day d^n ($n = 3, \cdots, n$) with the fixed part γ_{n0}. γ_{ns} ($s = 1, \cdots, n$) is the slope for the effect of the wave variables z_s, on the random slopes β_{nw} and the random part u_{nw}. β_{n+c} represents the fixed slope for the effect of wave day characteristics x_c ($c = 1, \cdots, n$). Moreover, ε_{wd} represents the residuals at level 1 of the multilevel model. Finally, Equation 2 has the same variance-covariance structure as for the basic model (see Equation 1).

Monitoring the data collection for a specific wave

Using the parameters of the fixed part of the basic model (see Equation 1) - the potential linear, quadratic, and higher polynomial functions of wave days d^n - we can display the functional form graphically as a curve:

$$E(RP_d) = \gamma_{00} + \gamma_{10}d + \gamma_{20}d^2 \cdots + \gamma_{n0}d^n. \qquad (3)$$

The curve in Equation 3 expresses the expected daily response proportion $(E(RP_d))$ of a panel wave and hence, the curve can be used as a benchmark to monitor the daily response proportion of ongoing panel waves. We model a 95% confidence band around the predicted curve to allow some uncertainty of the model. This confidence band serves as a benchmark when we monitor other data collection waves. We then simulate the monitoring of the daily response proportion for a specific wave (in our case wave 21) by plotting the daily response proportion of wave 21 against the predicted curve of the data collection evolution of wave 1-20 and its corresponding 95% confidence band. With this analytical approach, we can monitor the data collection for each day of a panel wave and investigate the effect of response-enhancement features during the data collection.

Next to the evolution of the daily response rate, we monitor the evolution of some key survey variables and their sampling error (for further information see Vandenplas et al., 2017). The aim of monitoring the evolution of key survey variables is to detect when the data collection has reached its maximum "phase capacity", meaning that additional respondent do not contribute to the sample composition in terms of reducing potential nonresponse bias. The sampling error represents the precision of the estimates and indicates how the estimates of the variable of interest get closer to a pre-defined "true" value (in our case the first panel wave).

Results

The shape of the data collection evolution

The general shape of the evolution of the daily response proportion during the data collection is described in Figure 1 , which represents the response proportion per day for each wave. In most waves, the daily response proportion first drops (except for wave 1, 8[iii], 13, and 14), then slightly increases around the middle of the data collection phase (although less obvious in waves 5, 6, 8, 14, 17, and 21), falls again, increases slightly again before it levels-off. This shape of the data collection evolution could be caused by the reminder structure and will be further investigated in the following sections.

To model the shape of the data collection evolution, we estimated the basic model of Equation 1 (see table Table 1, model 1). In model 1 we investigate the shape of the data collection evolution by including step-wise polynomial functions of the wave day as explained in the method section. The final model contains a linear function ($\beta_{1w}d$), a quadratic function ($\beta_{2w}d^2$), a cubic function ($\beta_{3w}d^3$), quartic function ($\beta_{4w}d^4$), and a quintic function ($\beta_{5w}d^5$) for the wave day, as these functional forms of wave days were significant. However, adding a sextic function ($\beta_{6w}d^6$) for wave days resulted in non-convergence of the model. Moreover, most of the panel waves described in Figure 1 represent a quintic function for the response proportions per wave days (e.g., panel waves 1, 2, 6, 7, 9, 10, 13, 16, 17, 20, and 21). Wave 8 seems to be particular, starting with a lower response proportion on day 0 and not displaying the steep decrease in the first days. All slopes (β_{nw}, *n* = 1, …, 5) were entered both as fixed and random slopes. Only the linear slope (β_{1w}) of the wave days was retained as random (variance σ_{1w} significant different from 0), all other slopes (quadratic, cubic, quartic, and quintic) were kept fixed ($u_{2w} = u_{3w} = u_{4w} = u_{5w} = 0$) as the variances were tiny and not significantly different from 0 (results not presented).

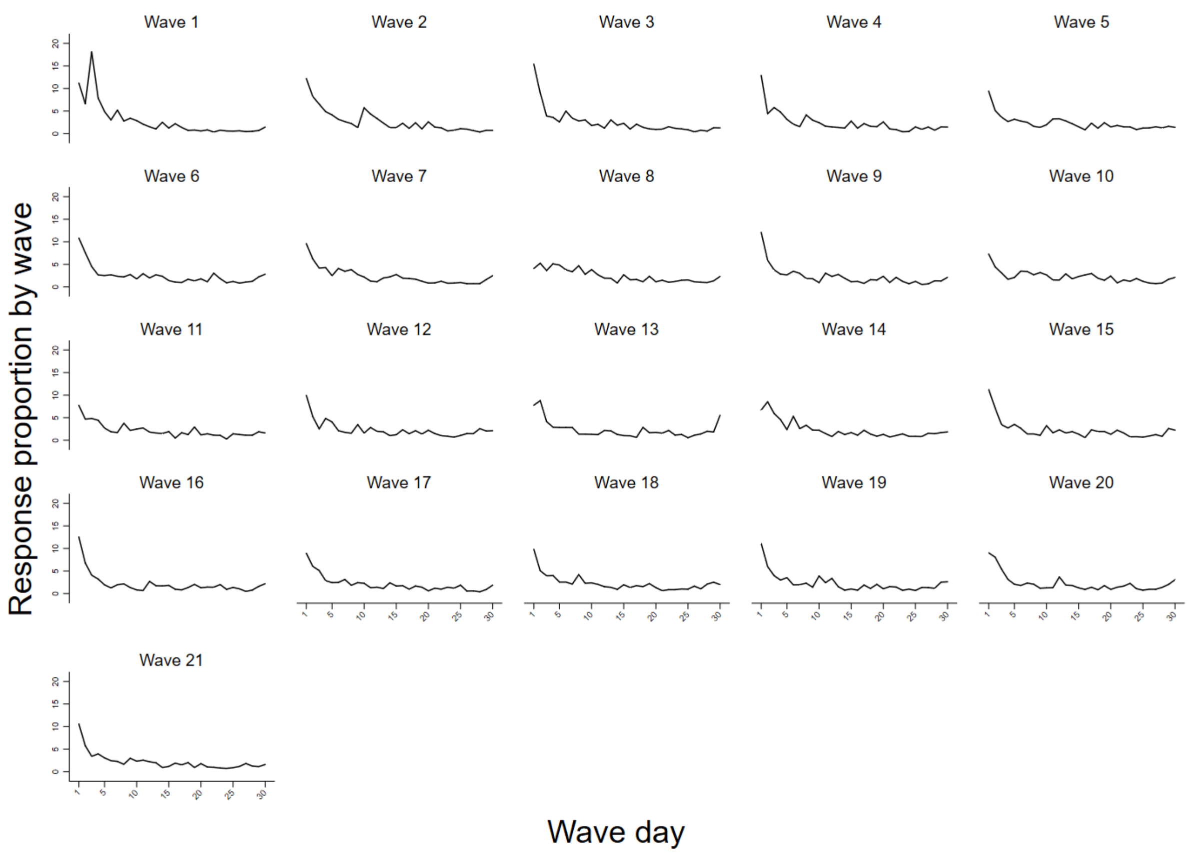

Figure 1. Daily response proportion per wave day in percent for each panel wave separately.

To visualize the general shape of the evolution of the daily response proportion of Equation 3, the parameter estimates from Table 1, Model 1 are plotted for each wave separately in Figure 2. The general shape of the data collection evolution displays a steep decrease and then a leveling-off of the decrease in the first six to seven days, followed by a slight increase around the reminders. Then, the daily response proportion drops again and increases towards the last days of the data collection phase. Finally, the last increase in the response proportion levels-off.

Looking at Figure 2, the leveling off of the daily response proportions occurs between day 6 and day 12 depending on the wave and corresponds with the sending of the first reminder. The cubic function of day of panel wave (γ_{30} = -0.01), is negative which causes the leveling-off of the initial decrease and the slight increase that follows to slow down to lead to a new decrease around day 19, which is, for most of the waves, after the second reminder is sent. Finally, the second decrease levels-off leading to a new increase at the end of the data collection phase, in line with the extremely small positive quartic function (γ_{40} = 0.00). This last increase levels-off due to the extremely small negative quintic function (γ_{50} = -0.00).

The intercept in the fixed part of Table 1, Model 1 (γ_{00} = 8.91) shows that the expected response proportion on the first day of data collection is 9.36, meaning that 9.36 percent of the invited panelists are expected to complete the questionnaire on day 1. The fixed linear term of the day of a wave (d^1 equivalent to γ_{10} = -2.71) is negative, meaning that a decrease of roughly 3 percentage points per day in response proportion can be expected in the first days of data collection. However, the quadratic function (d^2 equivalent to γ_{20} = 0.38) is positive, meaning that the decrease in daily response rate is expected to level off.

The variances of the random intercept (σ_0^2= 0.64) and the random slope of the linear function (σ_1^2= 0.00) show that the response proportion on the first day and the way in which it decreases in subsequent days may vary from wave to wave. The negative covariance term (σ_{01} = -0.04) of the intercept and the linear function means that the higher the response proportion is on the first day, the faster the decrease in response proportion the following days.

The daily response proportion decreases with each wave, as we would expect from previous research on panel attrition. One should note that all curves go through the same point on day 19 (see the inflection point Figure 2). This inflection point is the same for all panel waves and as no substantive meaning, as it is an artifact of the model estimation[iv]. The model lacks stability at this point as it only includes 21 waves, repeating the analysis with more waves would provide a more stable model (Maas & Hox, 2005).

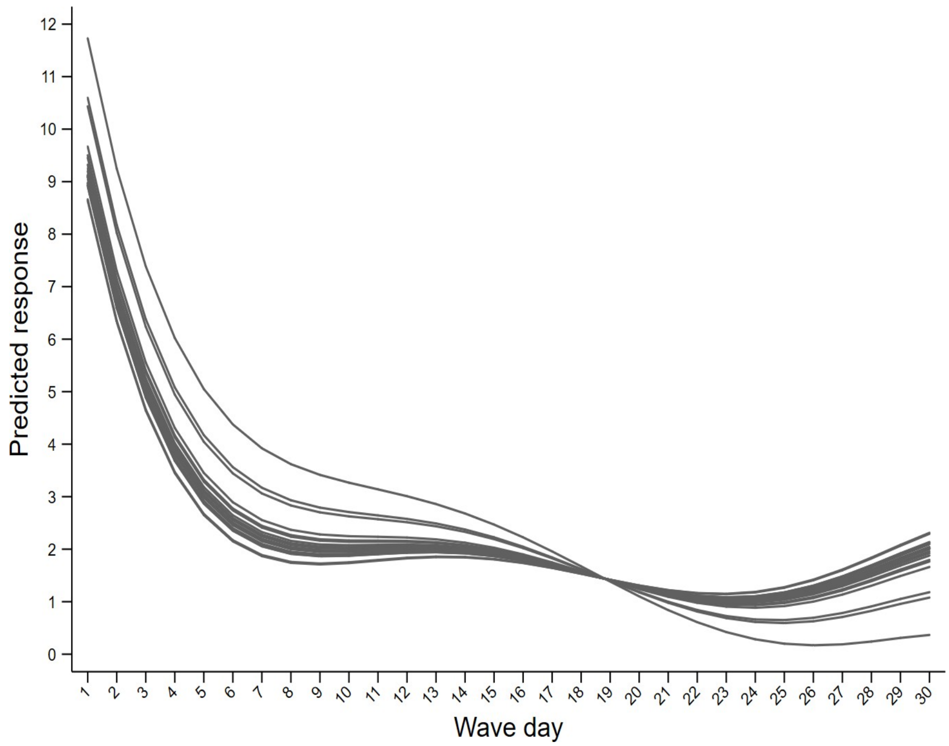

Figure 2. Shape of predicted daily response proportion for all panel waves representing the data collection evolution for each panel wave separately.

Characteristics influencing the shape of the data collection evolution

Next, we consider the quintic model with covariates (see Equation 2 with a quintic function for day of panel wave), which is presented in Table 1, Model 2. On the wave day level the weekday and the day sending reminders significantly influence the data collection evolution.

The daily response proportion is significantly lower on Tuesdays, Wednesdays, Thursdays, Fridays, and Saturdays compared to Mondays. A next step would be an experimental test to check whether data collections are better started on Mondays and whether reminders are better sent on Mondays as suggested by these results. Whether there was a public or school holiday anywhere in Germany on a specific wave day had no significant effects (results not presented).

Table 1. Parameters of the multilevel model describing the shape of data collection evolution for all panel waves.

	Model 1 basic model		Model 2 with covariates	
	$\hat{\beta}$	Std. err.	$\hat{\beta}$	Std. err.
Day of wave	-2.71***	0.13	-2.78***	0.12
Day^2 of wave	0.38***	0.03	0.39***	0.03
Day^3 of wave	-0.02***	0.00	-0.02***	0.00
Day^4 of wave	0.00***	0.00	0.00***	0.00
Day^5 of wave	-0.00***	0.00	-0.00***	0.00
Ref. Response on Monday				
Tuesday			-0.37*	0.14
Wednesday			-0.33*	0.13
Thursday			-0.71***	0.13
Friday			-0.77***	0.16
Saturday			-0.57***	0.13
Sunday			-0.02	0.13
Reminder 1			2.05***	0.24
Reminder 2			0.98***	0.24
Reminder 3			0.88***	0.24
γ_{00}	9.36***	0.25	9.84***	0.25
σ_0^2	0.64	0.24	0.62	0.23
σ_1^2	0.00	0.00	0.00	0.00
σ_{01}	-0.04	0.01	-0.03	0.01
σ_ε^2	0.96	0.06	0.79	0.05
Number of waves	21		21	
Number of days	630		630	

Note. – $\hat{\beta}$=coefficients , Std. err.=standard errors, Ref. = Reference category
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The first reminder increases the daily response proportion by 2.05 percent, the second reminder by 0.98 percent and the third by 0.88 percent. These findings show that there is an effect of reminders, which cause small peaks in the data collection evolution or stop the response proportion from further decreasing but that the effect diminishes with the number of reminders. Furthermore, we see that there is an increase of the response proportion on the last few days of the wave, which may suggest that some panelists systematically wait until the final days of data collection to participate. At the wave-level, respondent satisfaction and questionnaire length of the previous wave, as well as the weekday of launching the data collection were considered. However, none of the explanatory variables on the wave level were significant (results not presented).

The variance of the random slope σ_1^2 and the random intercept σ_0^2 (both significantly different from 0) and their covariance σ_{01} do not change much from the model without the covariates (Table 1, Model 1) to the model with covariates (Table 1, Model 2). However, the residual variance σ_ε^2 is reduced by introducing the covariates (0.96 vs. 0.79). This means that the explanatory variables introduced do not explain variation between the shape of the data collection evolution, but they do explain some of the residual errors (of the basic model) within the waves.

Monitoring the data collection of wave 21

Monitoring response proportions

Figure 3 displays the monitoring graph of the data collection evolution for wave 21. We used the general shape of the quintic function of the data collection evolution to create a confidence interval based on the first 20 waves, which serves as a benchmark to simulate the monitoring of wave 21. The grey benchmark curve represents the 95% level confidence band based on the first 20 waves. The crosses represent the daily response proportion for each day of wave 21. The vertical lines represent the timing of the three reminders in wave 21 (reminder 1 = day 9;

reminder 2 = day 16; reminder 3 = day 20; CATI reminder for sample units who did not respond in the three previous waves of wave 21= day 23). A day of data collection should be flagged when the daily response proportion falls outside the confidence band of the data collection evolution of previous panel waves for two or more days.

In Figure 3 we see that the predicted response proportion of day 1 in wave 21 is above the confidence band, but days 2 and 3 are below the confidence band. This could have led to sending the reminder earlier, although day 4 to 7 fall again in the confidence band. Day 8 is under the confidence band. On day 9 the reminder is sent, and the daily response proportion lies above the confidence band. The reminder effect seems to persist on day 10 and 11. The daily response proportion then falls back in the confidence band for days 12 to 13 and under the confidence band for days 14 and 15 until it increases again as a consequence of the second reminder on day 16. After day 16 the daily response proportion of wave 21 oscillates around the confidence band. These results indicate that the first and second reminder in wave 21 could have been sent earlier. In wave 21, there is an effect of reminder 3 on day 20 on predicted response proportion, not as effective as reminder 1 on day nine but almost as effective as the effect of the second reminder on day 15, which is slightly in contradiction with Couper's (2008) results that more than two reminders are often inefficient. This could be due to different behavior from long-term panelists, who might wait for a specific reminder until they participate in a panel wave. Finally, we find no significant effect on the predicted response proportion of the CATI reminder on day 23.

Monitoring a key survey variable

To monitor whether additional respondents contribute to the sample composition and to investigate the efficiency of the data collection evolution, we monitor the daily participation of respondents by predicted mean age based on the previous 20 panel waves using the 95% confidence intervals for predicted mean age as a benchmark.

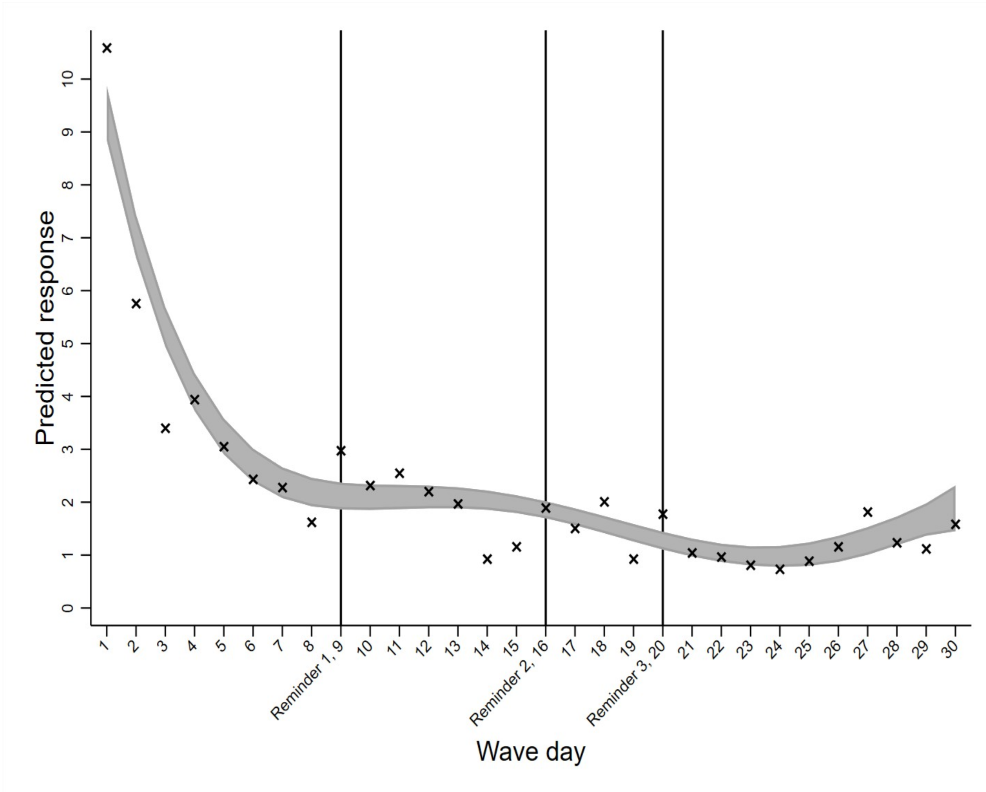

Figure 3. Monitoring predicted daily response proportion represented by the crosses in wave 21 (grey bands represent 95% confidence intervals for cumulative estimates of previous panel waves; vertical lines represent the sending of reminders).

In Figure 4 we see that until wave day five the mean age during data collection in wave 21 falls within 95% confidence interval band of previous panel waves. Between day six and eight the mean age is above the confidence band meaning that more older respondents (or less young respondents) participated in the panel wave than in previous panel waves. Furthermore, in previous panel waves, the stabilization of the mean age starts at about day eight. This stabilization of mean age seems to be later (day 27) in the case of wave 21.

On wave day 9, the day of the first reminder in wave 21, the mean age falls back in the confidence interval suggesting that the reminder activates younger respondents to participate. The mean age stays in the confidence interval (slightly above on day 10 and 15) until the day after the second reminder (day 16).

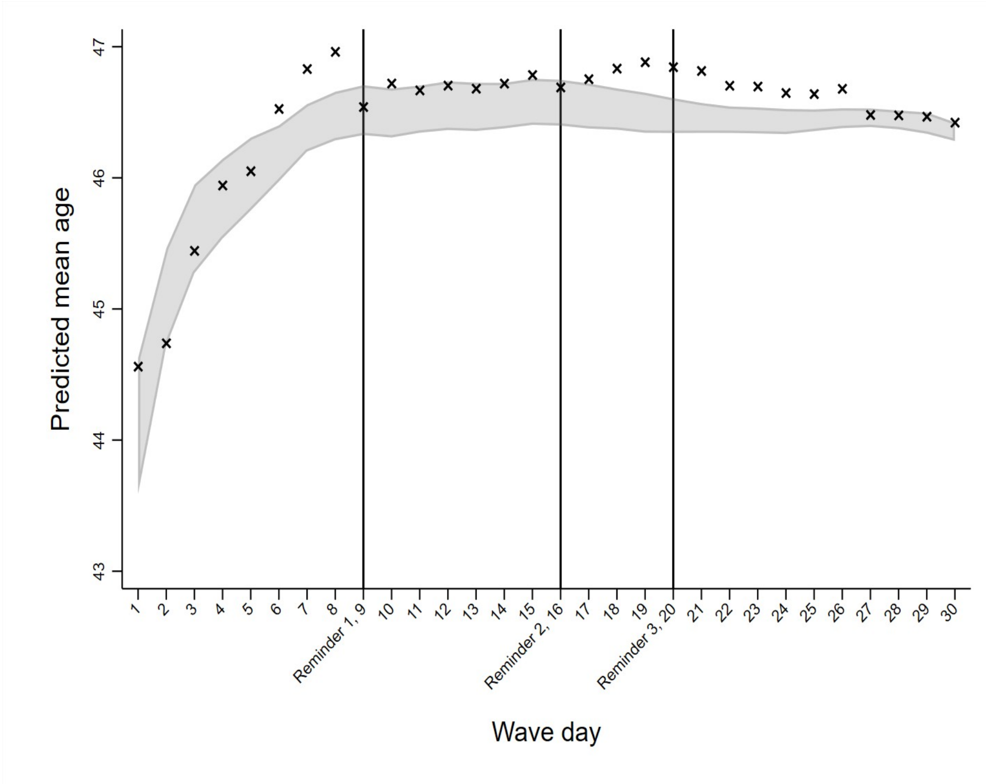

Figure 4. Monitoring daily participation in wave 21 for mean age represented by crosses (grey bands represent 95% confidence intervals for cumulative estimates of previous panel waves; vertical lines represent sending of reminders).

On wave day 17, the day after the second reminder in wave 21, the mean age falls above the confidence interval until day 26. After the third reminder, the mean age of wave 21 decreases slightly meaning this reminder may activate younger panelists. The mean age drops back in the confidence interval on the last day showing that a large proportion of younger respondents participated in the last days. The final obtained mean age is in the confidence band, showing that the same age distribution is obtained in wave 21 than in the previous wave, although younger respondents seem to wait longer until they participate. Under the assumption that later participation is an indicator for dropout in consecutive waves, the late participation of younger panelists could be a sign of an increased risk to dropout. Hence, future research may test whether groups of panelists that are at risk to dropout should be the target of specific interventions.

Monitoring the sampling error

In Figure 5 (for more details see Table A1) we plot the daily sampling error of predicted mean age of wave 21 (crosses) against the sampling error of predicted mean age of the first panel wave (dots). In general, the sampling error of both panel waves is always below 1.0. The sampling error for mean age decreases until day 12. From day 13 on the sampling errors are relatively stable for wave 1 with some jumps (day 15 to 19 and day 26 to 30) and the sampling errors do not become much smaller than in previous wave days.

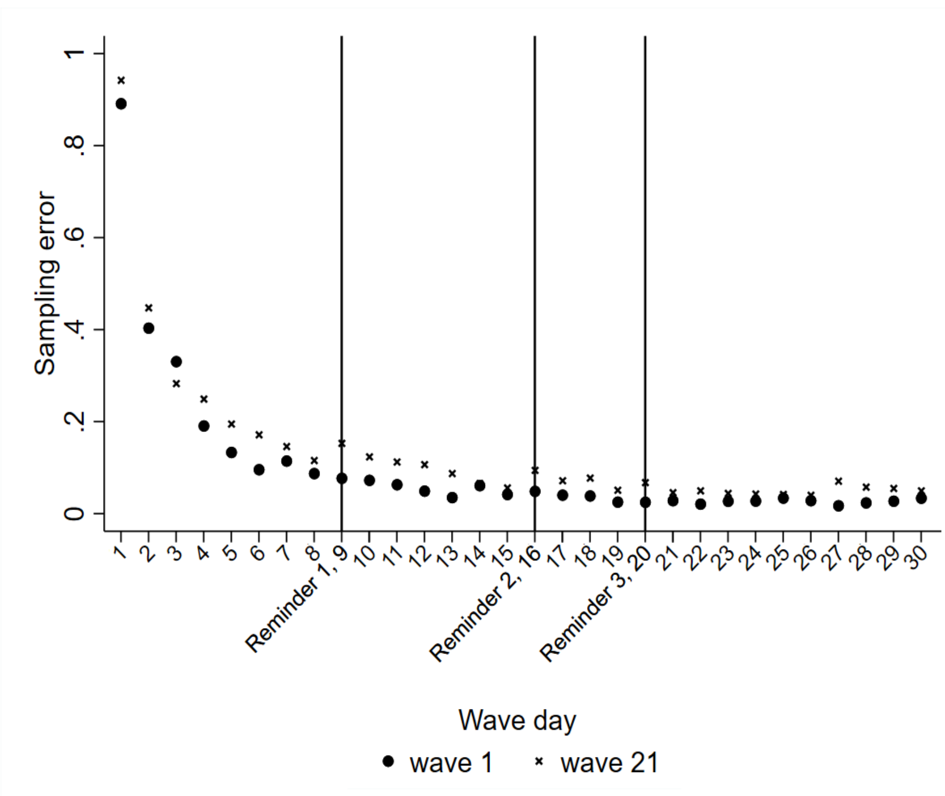

Figure 5. Daily sampling error for mean age for wave 1 (dots) and wave 21 (crosses). Vertical lines represent the sending of reminders**.**

Comparing wave 1 and wave 21 in Figure 5, we see that from day four until day 30 the sampling error of wave 21 is higher compared to the sampling error for wave 1 except for day 13. From day

13 forward the graph does not show a large difference between the sampling errors in both waves. For wave 1 the precision of the estimate (sampling error) has reached its minimum already at day seven, while for wave 21 the precision of the estimates has reached its minimum at day 13. This means that we need a longer data collection phase and possibly more reminders to obtain comparable precision for mean age in wave 21 than in wave 1.

Practically, we see that the evolution of the mean age of the participating panelists during the data collection phase is different and that the precision of the estimate takes longer to be reached. This shows that not only the sample size is reduced by panel attrition, but that either the sample composition changes (maybe less young people participate) or that the behavior of the panelists by age category changes (younger panelist answering later) across panel waves.

Conclusions

This paper aimed to present a method to re-evaluate and optimize the data collection phases to increase response rates, decrease panel attrition, and save costs for each wave in an online panel. In particular, we seek to understand the evolution of the daily response proportion (the number of completed questionnaires in one day divided by the number of invited panelists in the considered wave) and the factors that influence the shape of the data collection evolution across online panel waves. More precisely, the goal was to re-evaluate the number and timing of reminders in a data-driven manner to adapt to the panelists changing response behavior and the evolution of the sample composition across panel waves.

First, we model the shape of the evolution of the daily response proportion. The results of a multilevel model with days of data collection clustered in 21 panel waves show that the data collection evolution is quintic for the day of data collection within a panel wave: starting with a decrease in response proportion over the first days of the fieldwork, followed by a leveling-off of the decrease, followed by an increase around the sending of the reminders, which drops again before the response proportion increases, and to finally level off at the last days of the data collection phase.

Second, characteristics that can influence this shape were introduced in the model on the wave-level and wave-day level. We find that none of the wave-level characteristics affected the shape of the data collection evolution (e.g., weekday on which the wave started, mean respondent satisfaction with the previous wave, mean questionnaire length). At the wave day level, both the day of the week and the day a reminder was sent had a significant effect on the daily response proportion. The results also showed that the first reminder is the most efficient (largest effect on response proportion), whereas the second and third reminders have smaller effects on participation and hence, panel attrition.

Third, we modeled the shape of the data collection evolution of multiple panel waves to estimate a benchmark (here the 95% confidence interval) to monitor the data collection of one specific panel wave. We find several daily response proportions that fall below the benchmark, indicating that the timing and the amount of sending reminders could be adjusted in the online panel to potentially achieve a higher data collection efficiency for future waves. However, this adjustment needs experimental support. Finding the optimal number and timing of reminders, given specific survey characteristics, is one possible approach to increase response rates and avoid panel attrition during the data collection. For example, Lugtig and Blom (2018) showed that respondents are more likely to attrite the longer they wait until they respond to an online panel wave. Hence, experimental testing might be valuable to investigate the impact of optimizing the timing and amount of reminders on time to participate; and whether fast participation avoids potential panel attrition in the first place.

Fourth, monitoring the evolution of the mean age and sampling error for mean age in wave 21 indicated that the mean age estimates stabilized later than in previous panel waves. The mean age estimate increases up to the first reminder and reaches above the benchmark bands showing that older panelists participated in a larger proportion up to that point. After each reminder a decrease in the mean age can be observed, meaning that the reminders impacted the participation of younger panelists. This shows that the sample composition changed due to panel attrition and/or that panelist at different ages start to behave differently during data collection phases over the waves. The sampling error of mean age becomes small and stable from around day 23 onward, which is later than in the first panel wave. In line with the increase of response proportions during the final days of a data collection phase, additional respondents still influence the sample composition with regard to the key survey variable age, decreasing mean age and its sampling error.

In summary, in our example, the length of the data collection phase could not have been shortened in the 21st online panel wave because both the key survey variable and the sampling error show variation up to the last day. However, this might be due to younger panelists always answering to the panel wave requests until shortly before the deadline of the data collection phase exceeds (day 25). Hence, it needs to be tested whether the data collection phase can be shortened, as younger panelists may answer earlier if the deadline of the data collection phase expires earlier. Furthermore, the results indicate that the third reminder is needed, as this reminder both increases the response proportion substantially in wave 21 and influences the mean age estimate. It is important to note that a temporary stabilization of the mean age estimate and its sampling error can be observed around day 22.

In addition, panel attrition might be avoided if the second reminder would be sent two to three days earlier. However, this adjustment of the reminder procedure needs to be tested experimentally, as it is unknown whether the effects of reminders are stable (for a discussion see Göritz & Crutzen, 2012). Furthermore, the results indicate that response proportions are the highest on Mondays, suggesting that testing of whether response-enhancement strategies are more efficient on Mondays is worthwhile. These findings exemplify that survey practitioners might reconsider the fieldwork length and response-enhancement strategies, such as sending (extra) reminders at a specific time, switching the mode of reminders, or introducing an incentive across the lifetime of an online panel to avoid panel attrition.

This paper, however, has its limitations. First, we assume that the error-terms between days within a panel wave are uncorrelated. This is as strong assumptions, but we do not have enough data (waves) to fit a more complex model. Second, the results are limited to the specificity of the data collection phase of the GIP. Future research should investigate whether the presented approach could be used when several similar online surveys are aggregated instead of aggregating panel survey waves. In addition, the influence of the proposed adjustments in the data collection and potential interventions in the panel waves could enhance the research on fieldwork monitoring in online panels.

To conclude, modeling the daily response proportion across waves of an online panel can inform survey conductors about the efficiency of their data collection. By monitoring daily response proportions and the evolution of survey variable estimates, survey practitioners get informed on how to adapt the data collection phase to the wave or survey characteristics and hence, response-enhancement features might be more efficient.

Acknowledgments

This work was supported by the German Research Foundation (DFG, [SFB 884]). The authors gratefully acknowledge support from the Collaborative Research Center (SFB) 884 "Political economy of Reforms" at the University of Mannheim.

The data used in the analyses of this article are freely available as part of the Scientific Use Files (SUFs) of the German Internet Panel (GIP) survey data. They can be requested from the GESIS Data Archive for the Social Sciences (GESIS-DAS) at https://dbk.gesis.org/dbksearch/. The data sets used are cataloged under the GESIS-DAS reference numbers ZA5925 through ZA6954 (wave 14-34), which correspond to the following DOI numbers: 10.4232/1.12620, 10.4232/1.12621, 10.4232/1.12622, 10.4232/1.12623, 10.4232/1.12624, 10.4232/1.12838, 10.4232/1.12840, 10.4232/1.12841, 10.4232/1.12842, 10.4232/1.12843, 10.4232/1.12844, 10.4232/1.12755, 10.4232/1.12756, 10.4232/1.12755, 10.4232/1.12784, 10.4232/1.12889, 10.4232/1.12890, 10.4232/1.12976, 10.4232/1.12977, 10.4232/1.13011, 10.4232/1.13043,10.4232/1.13082, and 10.4232/1.13156.

References

Allison, P.D. (2001) *Missing data* (vol. 136), Thousand Oaks, CA: Sage.

Andreß, H.-J., Golsch, K. and Schmidt, A.W. (2013) *Applied panel data analysis for economic and social surveys*, Berlin: Springer.

Behr, A., Bellgardt, E. and Rendtel, U. (2005) Extent and determinants of panel attrition in the European Community Household Panel, *European Sociological Review*, 21(5): 489–512. doi: 10.1093/esr/jci037

Bethlehem, J.G. (2002) Weighting nonresponse adjustment based on auxiliary information, In R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little (eds), *Survey nonresponse*, Hoboken, NJ: John Wiley & Sons, pp. 41–54.

Billiet, J., Philippens, M., Fitzgerald, R. and Stoop, I.A.L. (2007) Estimation of response bias in the European Social Survey: using information from reluctant respondents in round one, *Journal of Official Statistics*, 23(2): 135–62.

Blom, A.G., Bosnjak, M., Cornilleau, A., Cousteaux, A.-S., Das, M., Douhou, S. and Krieger, U. (2016) A comparison of four probability-based online and mixed-mode panels in Europe, *Social Science Computer Review*, 34(1): 8–25. doi: 10.1177/0894439315574825

Blom, A.G., Bruch, C., Bossert, D., Felderer, B.I., Fickel, M., Funke, F., Gebhard, F., Herzing, J.M.E., Höhne, J.K., Holthausen, A., Krieger, U. and Rettig, T. (2018) German internet panel, wave 14 to 34. GESIS data archive. ZA5925 through ZA6954', SFB 884 – Political Economy of Reforms, University of Mannheim.

Blom, A.G., Gathmann, C. and Krieger, U. (2015) Setting up an online panel representative of the general population: the German internet panel, *Field Methods*, 27(4): 391–408. doi: 10.1177/1525822X15574494

Blom, A.G., Herzing, J.M.E., Cornesse, C., Sakshaug, J.W., Krieger, U. and Bossert, D. (2017) Does the recruitment of offline households increase the sample representativeness of probability-based online panels? Evidence from the German internet panel, *Social Science Computer Review*, 35(4): 498–520. doi: 10.1177/0894439316651584

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A. and 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): 103–15. doi: 10.1177/0894439317697949

Cheng, A., Zamarro, G. and Orriens, B. (2018) Personality as a predictor of unit nonresponse in an internet panel, *Sociological Methods & Research*, online first, doi: 10.1177/0049124117747305

Couper, M.P. (2008) *Designing effective web surveys*, New York: Cambridge University Press.

Couper, M.P., Kapteyn, A., Schonlau, M. and Winter, J. (2007) Noncoverage and nonresponse in an internet survey, *Social Science Research*, 36(1): 131–48. doi: 10.1016/j.ssresearch.2005.10.002

Das, M., Toepoel, V. and van Soest, A. (2011) Nonparametric tests of panel conditioning and attrition bias in panel surveys, *Sociological Methods & Research*, 40(1): 32–56. doi: 10.1177/0049124110390765

Dennis, J.M. and Li, L. (2003) Effects of panel attrition on survey estimates, In *Annual Meeting of the American Association for Public Opinion Research*, May 17, 2003, Nashville, TN, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.200.7528&rep= rep1&type=pdf

Dillman, D.A., Smyth, J.D. and Christian, L.M. (2014) *Internet, phone, mail, and mixed-mode surveys: The tailored design method*, Hoboken, NJ: John Wiley & Sons.

Elder, G.H. and Giele, J.Z. (2009) *The craft of life course research*, New York: Guilford Press.

Fan, W. and Zheng, Y. (2010) Factors affecting response rates of the web survey: a systematic review, *Computers in Human Behavior*, 26(2): 132–9. doi: 10.1016/j. chb.2009.10.015

Firebaugh, G. (2008) *Seven rules for social research*, Princeton, NJ: Princeton University Press.

Frick, J.R., Grabka, M.M. and Groh-Samberg, O. (2012) Dealing with incomplete household panel data in inequality research, *Sociological Methods & Research*, 41(1): 89–123. doi: 10.1177/0049124112440796

Glenn, N.D. (2005) *Cohort analysis* (vol. 5), Thousand Oaks, CA: Sage.

Göritz, A.S. (2014) Determinants of the starting rate and the completion rate in online panel studies, In M. Callegaro, R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick and P.J. Lavrakas (eds), *Online panel research: A data quality perspective*, Chichester: Wiley, pp. 154–70.

Göritz, A.S. and Crutzen, R. (2012) Reminders in web-based data collection: Increasing response rates at the price of retention?, *American Journal of Evaluation*, 33(2): 240–50.

Groves, R.M. (2006) Nonresponse rates and nonresponse bias in household surveys, *Public Opinion Quarterly*, 70(5): 646–75. doi: 10.1093/poq/nfl033

Groves, R.M. and Heeringa, S.G. (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3): 439–57. doi: 10.1111/j.1467- 985X.2006.00423.x

Groves, R.M. and Peytcheva, E. (2008) The impact of nonresponse rates on nonresponse bias a meta-analysis, *Public Opinion Quarterly*, 72(2): 167–89. doi: 10.1093/poq/nfn011

Groves, R.M., Presser, S. and Dipko, S. (2004) The role of topic interest in survey participation decisions, *Public Opinion Quarterly*, 68(1): 2–31. doi: 10.1093/poq/ nfh002

Halaby, C.N. (2004) Panel models in sociological research: theory into practice, *Annual Review of Sociology*, 30: 507–44. doi: 10.1146/annurev.soc.30.012703.110629

Kaplowitz, M.D., Hadlock, T.D. and Levine, R. (2004) A comparison of web and mail survey response rates, *Public Opinion Quarterly*, 68(1): 94–101. doi: 10.1093/ poq/nfh006

Laflamme, F., Maydan, M. and Miller, A. (2008) Using paradata to actively manage data collection survey process, In *Section on survey research methods: American Statistical Association*, Ottawa: Statistics Canada.

Liu, M. and Wronski, L. (2018) Examining completion rates in web surveys via over 25,000 real-world surveys, *Social Science Computer Review*, 36(1): 116–24. doi: 10.1177/0894439317695581

Lugtig, P. (2014) Panel attrition: separating stayers, fast attriters, gradual attriters, and lurkers, *Sociological Methods & Research*, 43(4): 699–723. doi: 10.1177/0049124113520305

Lugtig, P. and Blom, A.G. (2018) It's the process stupid! Using machine learning to understand the relation between paradata and panel dropout, presented as a contributed paper at Methodology of Longitudinal Surveys II (MOLS), 25–27 July, Colchester.

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

Maas, C.J.M. and Hox, J.J. (2005) Sufficient sample sizes for multilevel modeling, *European Journal of Research Methods for the Behavioral and Social Sciences – Methodology*, 1(3): 86–92. doi: 10.1027/1614-2241.1.3.86

Malter, F. (2013) Fieldwork monitoring in the survey of health, ageing and retirement in Europe (SHARE), *Survey Methods: Insights from the field*, https://surveyinsights. org/?p=1974

Roßmann, J. and Gummer, T. (2016) Using paradata to predict and correct for panel attrition, *Social Science Computer Review*, 34(3): 312–32. doi: 10.1177/0894439315587258

Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70(1): 41–55. doi: 10.1093/ biomet/70.1.41

Rubin, D.B. (1987) *Multiple imputation for nonresponse in surveys*, Hoboken, NJ: John Wiley & Sons.

Schouten, B. and Shlomo, N. (2017) Selecting adaptive survey design strata with partial R-indicators, *International Statistical Review*, 85(1): 143–63. doi: 10.1111/insr.12159

Schouten, B., Peytchev, A. and Wagner, J. (2017) *Adaptive survey design*, Boca Raton, FL: CRC Press.

Stoop, I.A.L. (2005) *The hunt for the last respondent: Nonresponse in sample surveys*, (vol 8), The Hague: Sociaal en Cultureel Plan Bureau.

Vandecasteele, L. and Debels, A. (2007) Attrition in panel data: the effectiveness of weighting, *European Sociological Review*, 23(1): 81–97. doi: 10.1093/esr/jcl021

Vandenplas, C. and Loosveldt, G. (2017) Modeling the weekly data collection efficiency of face-to-face surveys: Six rounds of the European social survey, *Journal of Survey Statistics and Methodology*, 5(2): 212–32. doi: 10.1093/jssam/smw034

Vandenplas, C., Loosveldt, G. and Beullens, K. (2017) Fieldwork monitoring for the European social survey: an illustration with Belgium and the Czech Republic in round 7, *Journal of Official Statistics*, 33(3): 659–86. doi: 10.1515/jos-2017-0031

Vehovar, V., Batagelj, Z., Lozar Manfreda, K. and Zaletel, M. (2002) Nonresponse in web surveys, In R.M. Groves, D.A. Dillman, J.L. Etlinge and R.J.A. Little (eds), *Survey nonresponse*, Hoboken, NJ: John Wiley & Sons, pp. 229–42.

Wagner, J., West, B.T., Kirgis, N., Lepkowski, J.M., Axinn, W.G. and Kruger Ndiaye, S. (2012) Use of paradata in a responsive design framework to manage a field data collection, *Journal of Official Statistics*, 28(4): 477–99.

Watson, N. and Wooden, M. (2009) Identifying factors affecting longitudinal survey response, In P. Lynn (ed), *Methodology of longitudinal surveys*, Hoboken, NJ: John Wiley & Sons, pp. 159–67.

Weible, R. and Wallace, J. (1998) Cyber research: The impact of the internet on data collection, *Marketing Research*, 10(3): 19–24.

Appendix

Table A1. Sampling error for mean age in wave 21 per wave day.

Wave day	sampling error age
1	0.94
2	0.45
3	0.28
4	0.25
5	0.20
6	0.17
7	0.15
8	0.12
9	0.15
10	0.12
11	0.11
12	0.11
13	0.09
14	0.07
15	0.06
16	0.09
17	0.07
18	0.08
19	0.05
20	0.07
21	0.05
22	0.05
23	0.04
24	0.04
25	0.04
26	0.04
27	0.07
28	0.06
29	0.06
30	0.05

Endnotes

[i] We excluded day 31 from the analysis to keep days per wave constant. In case of some panel waves (12, 14, 20, and 21) the data collection lasted 32 days, due to inaccuracy of the fieldwork agency. However, less than two respondents participated during these unannounced additional days.
[ii] All analyses are conducted in Stata SE, version 15.1.
[iii] The curve of wave 8 is very flat compared to other waves. In this case, we can only speculate what happened as there is no incident in the fieldwork procedure reported. One possible explanation could be that the fielding started on a public holiday as well as that the public holidays allowed a long vacation by taking only a few days off in this specific year.
[iv] The day on which this convergence of all the curves happens can be calculated based on the covariance σ_{01} and the variances σ_0^2 and σ_1^2.