



How does broadband supply affect participation in panel surveys?: Using geospatial broadband data at the district level to analyze mode choice and panel attrition

Maikel Schwerdtfeger¹ · Ruben L. Bach²

Accepted: 29 May 2024 / Published online: 12 June 2024
© The Author(s) 2024

Abstract

Good broadband supply is crucial for different aspects of participation behavior in web surveys. In this study, we combined geospatial broadband data and the survey data of 16 waves of the mixed-mode GESIS Panel to investigate mode choice and panel attrition. Since small-scale geospatial data is often unavailable or difficult to access for research purposes, we report here results of a feasibility study investigating whether publicly available broadband data at the district level are sufficient to draw meaningful conclusions about participation behavior. In the first part of the analysis, we investigated the effects of broadband supply on mode choice. The results showed a positive effect of the broadband category with the highest coverage rate on choosing the online mode. We also found effects for internet familiarity, age, and education. In the second part, we investigated the effects of broadband supply on panel attrition within the 16 survey waves. We did not find significant effects of broadband supply on panel attrition. However, we found effects for evaluation of duration, measured duration, and age. Overall, we conclude that the granularity of the geospatial data on broadband supply is not ideal. A detailed discussion on the granularity of the broadband data in the context of the general availability of fine-grained geospatial data is provided at the end of this study.

Keywords Broadband supply · Mode choice · Panel attrition · Panel data · Geospatial data

✉ Maikel Schwerdtfeger
maikel.schwerdtfeger@gesis.org

Ruben L. Bach
r.bach@uni-mannheim.de

¹ GESIS – Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, 68072 Mannheim, Germany

² Mannheim Centre for European Social Research (MZES), University of Mannheim, 68131 Mannheim, Germany

1 Introduction

Over the last decades, collecting data online via web surveys has become crucial for quantitative research in the social sciences. Reflecting this circumstance, several survey studies (e.g., the German Internet Panel, LISS Panel, Understanding America Study) even provided participants with internet access and participation devices to enable them to participate in their web surveys. With the increasing prevalence of web surveys, many studies have examined the impact and consequences of the coverage problems resulting from the fact that not all participants are able or willing to participate online (Scherpenzeel & Bethlehem 2011; Bandilla et al. 2009; Baur & Florian 2009). From the perspective of this research on coverage, having broadband access often implies that people can participate in web surveys without any technical constraints, although the quality and speed of this broadband access are not considered. In reality, broadband speed can vary massively and thereby affect the survey experience. The present study aims to expand the survey methodological research by investigating the effects of broadband supply on panel participation in a feasibility study on the use of publicly available geospatial data. This geospatial broadband data is based on the *Breitbandatlas* of the Federal Ministry for Digital and Transport (Bundesministerium für Digitales und Verkehr 2023).

We measured the effects of broadband supply on panel participation by using two essential characteristics of a mixed-mode panel survey: mode choice and panel attrition. This approach enabled us to examine the impact of broadband supply at two different stages of participation. First, *mode choice* refers to a situation in which people in a recruitment interview of a panel survey decide whether they want to participate online via web surveys or offline via paper-and-pencil surveys. This decision may depend on several factors, such as respondents' preferences, access to technology, and comfort levels with the selected participation mode. Usually, this initial decision sets the participation mode for all subsequent survey waves. For this early stage of participation, we investigated whether people in regions with poor broadband supply are more inclined to use the offline mode. Second, among those who chose the online mode, we investigated whether poor broadband supply led to increased panel attrition. Increased panel attrition is a severe problem since even moderate attrition rates can substantially reduce the number of participants over the course of a long-term panel survey. As a result, panel attrition can undermine statistical power, and selective panel attrition can lead to biased survey estimates (see, e.g., Lugtig 2014). To assess broadband supply, we used geospatial data on regionally available broadband supply at the district level and combined it with geocoded panel survey data. Therefore, our research objectives involve exploring how broadband supply affects mode choice in mixed-mode panel surveys and understanding its impact on the attrition of online participants in panel surveys.

In summary, this study aims to expand methodological research in the context of panel surveys in two different ways. First, the approach of using geospatial data on broadband supply is a novelty in survey methodology that overcomes the restrictions of existing survey data. Geospatial broadband data can replace survey questions about existing broadband speed, which can reduce survey time and the respondent burden. Furthermore, broadband data is not affected by the motivated or unintentional misreporting of respondents and can be applied retrospectively. Second, analyzing participation behavior in the context of regionally available broadband supply enables us to draw conclusions about whether participants in regions with a poor broadband supply avoid the online mode; and if not, whether they have a higher probability of unit nonresponse than participants in regions with good broadband supply. These conclusions can be used to develop targeting strategies

that actively guide the participation mode choice based on the panelists' residence, thereby reducing the likelihood of panel attrition.

2 Theoretical background and hypotheses

2.1 Online waiting

Slow broadband speeds cause online waiting situations, although the delay often is just a matter of seconds. For example, it takes one additional second to load the search results of a search engine and then two additional seconds to load the suggested website. In total, the delay is only three additional seconds, but the flow experience is interrupted each time.

The crucial factor in evaluating waiting times is not the absolute duration but the waiting-time gap. The *waiting-time gap* is the difference between the time someone is willing to wait and their perceived waiting time (Chebat et al. 2010). In an environment in which every request is expected to be processed and answered within seconds, a 20-s wait for a page download may be perceived as unreasonable, just as a 20-min wait for a comparable traditional service would be perceived as unreasonable (Chebat et al. 2010). An experiment with Google users showed that even minimal changes in waiting time matter. Slowing down the search results page by 400 ms had an average impact of -0.6% on the number of searches per user (Brutlag 2009). Thus, even half a second delay can have a measurable impact on internet users.

The Weber-Fechner Law explains such findings as the ability of the human sensory system to perceive so-called just noticeable differences (Reichl et al. 2010). A *just noticeable difference* describes the minimum difference between two stimuli that is required for a person to perceive that the stimuli are not the same. Transferred to our use case, the following basic thresholds apply in expert communities for system programming: the limit for having the user feel that a system reacts instantaneously is about 0.1 s; the limit for keeping the user's flow uninterrupted is about 1.0 s, even though the user will notice the delay (Nielsen 1994).

To summarize, the decisive aspect for evaluating online waiting situations is not the absolute duration of the delay but the noticeable difference between the acceptable and the perceived waiting time. In online environments, acceptable waiting times are much shorter than in face-to-face situations. As a result, one could argue that even minimal delays can lead to worse evaluations or increased dropouts. In the following two sections, we apply these general mechanisms of online waiting to mode choice and panel attrition.

2.2 Mode choice

According to Smyth et al. (2014), access to and familiarity with a participation mode are the strongest predictors of mode preference, whereas measures of safety concerns, physical abilities, and normative concerns remained "unexpectedly weak" predictors. Accordingly, Herzing and Blom (2019) found that their indicator of digital affinity, which includes digital access and internet usage, influences whether people participate in online panel surveys. Both access to and familiarity with the internet can be seen as minimum requirements for choosing the online mode since people cannot participate online without internet access and a certain level of internet familiarity. Another study found that the familiarity, comfort, and convenience with a communication medium reflected a preference for a particular survey mode (Olson et al. 2012). Thus, in addition to the minimum requirements of internet access and familiarity, comfort and convenience with the internet play a crucial role in

choosing a participation mode. Similarly, Bretschi and Weiß (2022) found that short-term mode switching from offline to online mode is positively affected by three dimensions of internet use: frequency, variety, and number of devices.

With respect to the above-mentioned online waiting mechanisms, we expect that the degree of comfort and convenience is determined by past experiences with the internet. Positive experiences, like intuitive navigation and fluid operation, are expected to increase preference towards choosing the online mode. Negative experiences, like the inability to do something or unexpectedly long waiting times, are expected to increase preference towards choosing the offline mode. With regard to broadband supply speeds, we expect that increased page load times due to poor broadband supply will have a negative impact on the comfort and convenience with the internet, and thus also on choosing the online mode.

Thus, our first hypothesis is: Living in a region with a better broadband supply increases the probability of choosing online participation in a mixed-mode panel.

In addition to the influence of the regional broadband supply, we add important predictors from previous mode choice research to strengthen the explanatory power of the model that we will use to investigate the first hypothesis. As mentioned above, several studies have found a direct effect of internet familiarity on participation mode preference (Olson et al. 2012; Smyth et al. 2014). Also, male participants were more likely to choose the web mode (Diment & Garrett-Jones 2007) and possess lower tolerance, acceptance, and satisfaction levels for slower system response times (Yu et al. 2020). Finally, participation mode preferences are affected by age, as older participants have lower preferences for web mode (Diment & Garrett-Jones 2007; Millar et al. 2009), and education level, as higher education increases the likelihood of preferring the web mode (Millar et al. 2009; Smyth et al. 2014). Therefore, we included internet familiarity, gender, year of birth, and education level as control variables. The hypotheses in this regard are:

Second hypothesis: Being more familiar with the internet increases the probability of choosing online participation in a mixed-mode panel.

Third hypothesis: Being male increases the probability of choosing online participation in a mixed-mode panel.

Fourth hypothesis: Younger participants have a higher probability of choosing online participation in a mixed-mode panel.

Fifth hypothesis: Participants with a higher level of education have a higher probability of choosing online participation in a mixed-mode panel.

Next, we review the literature on panel attrition and apply it to the mechanism of online waiting.

2.3 Panel attrition

Galesic (2006) has emphasized that interest in survey questions and the burden experienced while answering them are the most important aspects of studying survey

behavior. Specifically, interest in the survey and other initial factors of motivation (e.g., incentives) become weaker over the course of a panel survey, and the influence of negative aspects that increase the burden (e.g., boredom) becomes stronger (Galesic 2006). Consistent with these findings, Gummer and Daikeler (2018) investigated respondents' experience in a mixed-mode panel survey and found that particularly the perceived length of surveys influences further participation, regardless of the participation mode.

With respect to survey burden, many studies have focused on the length of a survey as a crucial factor (see, e.g., Vicente & Reis 2010). Shorter web surveys yield higher response rates (Deutskens et al. 2004; Marcus et al. 2007) and lower dropout rates (Ganassali 2008). Both of these studies manipulated the number of questions to investigate the effects of survey length. However, it does not require a higher number of questions to increase the survey length and, thus, the survey burden. Crawford et al. (2001) found that their web survey with server problems, compared to a second identical web survey without server problems, increased the average length of a survey for those who completed it to 21.6 vs. 17.8 min, and the nonresponse to 66.1% vs. 63.9%, and the breakoffs to 10.6% vs. 8.8%.

In view of these findings and the above-described mechanisms of online waiting, we expect that even short delays in page load time can affect the survey burden negatively and increase the probability of dropping out of a panel survey.

Therefore, the sixth hypothesis is: Living in a region with a better broadband supply decreases the risk of attrition in an online panel survey.

In line with the mode choice analysis, we added important predictors from existing research on panel attrition to the model we used to investigate this hypothesis empirically. In the first place, many studies on web and mixed-mode surveys have found that survey duration is a crucial factor of panel attrition (Crawford et al. 2001; Deutskens et al. 2004; Ganassali 2008; Vicente & Reis 2010). With respect to the sociodemographic variables, the attrition rate in face-to-face interviews is lower among women (Behr et al. 2005; Lepkowski & Couper 2002) and people with higher education (Watson & Wooden 2009). The findings relating to participants' ages were mixed. For example, Lipps (2009) measured an increased risk of attrition for the oldest and youngest participants in face-to-face and telephone interviews, and Struminskaya (2014) found a small negative effect in an online panel survey. Consequently, we included the following control variables: evaluation of survey duration, measured survey duration, gender, year of birth, and education level. The hypotheses in this regard are:

Seventh hypothesis: Evaluating the survey duration as less long decreases the risk of attrition in an online panel survey.

Eight hypothesis: Taking more time to complete the survey increases the risk of attrition in an online panel survey.

Ninth hypothesis: Being male increases the risk of attrition in an online panel survey.

10th hypothesis: Younger participants have a higher risk of attrition in an online panel survey.

11th hypothesis: Participants with a higher level of education have a lower risk of attrition in an online panel survey.

2.4 Feasibility of using publicly available geospatial data

Since small-scale geospatial data is often unavailable or difficult to access for research purposes, we decided to conduct a feasibility study to test whether publicly available broadband data at the district level is sufficient for our examination. This is an exploratory approach for which we have no comparative or empirical values to draw from.

12th Hypothesis: Using publicly available geospatial broadband data at the district level is sufficient to draw meaningful conclusions about participation behavior in a panel survey.

3 Data and methods

3.1 GESIS Panel survey data

The survey data we used in this paper is the GESIS Panel – Extended Edition version 31.0.0 (GESIS 2019). The GESIS Panel is a probability-based panel survey of the GESIS – Leibniz Institute of the Social Sciences that started in 2013. It uses two self-administered participation modes: online via web surveys or offline via paper and pencil surveys. The target population of the GESIS Panel comprises all German-speaking people between 18 and 70 years of age who permanently reside in Germany (Bosnjak et al. 2018). After the initial sampling, the GESIS Panel carried out refreshment samples in 2016 and 2018 from the German General Social Survey (ALLBUS). Schaurer and Weyandt (2018) and Schaurer et al. (2020) have pointed out that the participants of the ALLBUS were asked whether they were willing to participate in a subsequent self-administered panel survey. If they agreed, those who use the internet were nudged to use the online mode. Those participants who did not use the internet or did not want to participate online were offered the offline mode.

For this paper, we used the second cohort of the GESIS Panel since the geospatial broadband data were not available before 2016. This cohort was recruited in 2016 with a minimum recruitment rate (RECR) of 18.36% (Schaurer & Weyandt 2018). The resulting data comprises 16 regular waves from June 2016 ("wave dc") to December 2018 ("wave ff") in addition to the recruitment interviews of the second cohort ("d11" and "d12").

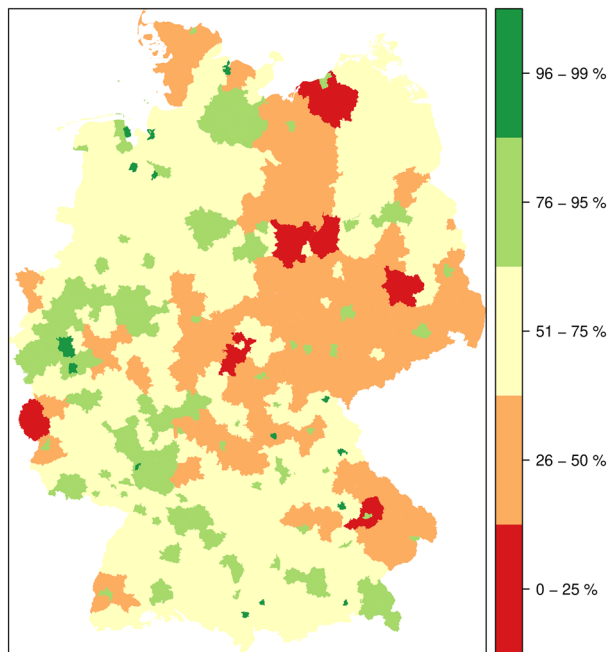
3.2 Broadband supply in Germany

The geospatial data on broadband supply we used in this paper is from Germany's Federal Agency for Cartography and Geodesy, which provides a machine-readable map on their website called *Broadband Supply with 50 Mbit/s*¹ (Bundesamt für Kartographie und Geodäsie 2016; cf. Internet Archive 2018). The data is based on the *Breitbandatlas* of the Federal Ministry for Digital and Transport (Bundesministerium für Digitales und Verkehr 2023). As shown in Fig. 1, the map provides information on the proportion of broadband supply with at least 50 Mbit/s available in each German administrative district in mid-2016. The administrative districts comprise 432 units categorized by rural districts ("*Landkreise*"), independent cities ("*Kreisfreie Städte*"), districts ("*Kreise*"), and city districts

¹ Although the map is not available anymore on the website of the Federal Agency for Cartography and Geodesy, a screenshot of the website from October 2018 can be obtained from Internet Archive (2018).

Fig. 1 Broadband supply in Germany in 2016

Proportion of broadband supply with at least 50 Mbit/s



("Stadtkreise"). The lowest broadband category (0–25%) means that a maximum of 25% of the households in such an administrative district have a broadband connection with a data transmission rate of at least 50 Mbit/s. This 50 Mbit/s speed was considered to be a threshold value for sufficient data transmission (Bundesamt für Kartographie und Geodäsie 2016; cf. Internet Archive 2018).

3.3 Operationalization and methods

3.3.1 Preparation of the analysis data

The geospatial broadband data and the geocoded survey data needed to be combined to investigate the effects of broadband supply on the mode choice and panel attrition of the participants of the GESIS Panel. The first step was to retrieve and process the machine-readable data of the map of broadband supply with a geographic information system (GIS). We accomplished this by using the open-source software QGIS. We used the resulting shapefile in the statistical programming language R as a `SpatialPolygonsDataFrame` (R Core Team 2021). Next, we transformed the geospatial broadband data and the coordinates of the GESIS Panel participants to the same coordinate reference system (CRS). The CRS provides information on the coordinate origins and curvature of the earth. Finally, we generated a variable with broadband categories for each participant in each survey wave and appended these variables to the GESIS Panel survey data. The new variables assign each participant the broadband category of their district for each survey wave.

In the recruitment interview, participants were asked whether they used the internet for private purposes. For the mode choice analysis, participants who did not have internet access at all were excluded from the analysis due to the missing choice situation regarding the participation mode. Furthermore, 15 participants were excluded due to switching their participation mode between the recruitment interview and the first survey wave. Four of them switched from online to offline because they did not provide a valid email address. The other 11 participants switched from offline to online since they used the URL in the invitation letter of the profile survey and then provided a valid email address. The resulting dataset for the mode choice analysis contained 1,455 observations. For the panel attrition analysis, we used the online participants and converted their data into a long format to analyze the longitudinal panel data of the 16 waves. The resulting dataset for the panel attrition analysis contained 13,095 observations.

3.3.2 Mode choice analysis

The first set of analyses was concerned with mode choice as the dependent variable. We used a binomial logistic regression model since mode choice is a binary variable (0=Offline, 1=Online). Choosing the offline mode serves as a reference category.

In the recruitment interview, participants who reported that they used the internet for private purposes were asked whether it was acceptable for them to answer the panel survey online. Those who disagreed became offline participants. The main independent variable of the mode choice analysis was the broadband category at the time of the first wave. It was classified according to the proportion of broadband supply with at least a 50 Mbit/s speed in the administrative district of the participant (1=0–25% of at least 50 Mbit/s, 2=26–50% of at least 50 Mbit/s, 3=51–75% of at least 50 Mbit/s, 4=76–95% of at least 50 Mbit/s, 5=96–99% of at least 50 Mbit/s).

Additionally, we included four control variables in the mode choice analysis: internet familiarity (measured as the reported frequency of private internet usage: 1=Rarer, 2=About once a week, 3=More than once a week, 4=About once a day, 5=Several times a day), gender (0=Female, 1=Male), year of birth, and the highest level of education (corresponding to the general school-leaving qualification: 1=Low, 2=Medium, 3=High²). The lowest categories of broadband, familiarity, and education, as well as female gender, serve as reference categories. See Table 1 in the Appendix for the descriptive statistics of all the variables we used in the mode choice analysis.

3.3.3 Panel attrition analysis

The dependent variable of the second analysis was panel dropout, which is a dichotomous indicator for attrition in each wave (0=No dropout, 1=Dropout). Not dropping out serves as a reference category. The main independent variable of the panel attrition analysis was the time-dependent broadband category with a broadband value for each participant in

² Low education level: Individuals without a formal school-leaving qualification or those with a "Hauptschulabschluss," which is typically obtained after completing lower secondary education. Medium education level: Participants with a "Realschulabschluss," indicating completion of a higher level of secondary education. High education level: Individuals with a "(Fach-)Hochschulreife" or "Abitur," representing the highest level of secondary education in Germany.

Table 1 Descriptive statistics of the mode choice analysis

Variable	Statistics/Values	Frequencies (%)
Mode choice: Online	0=Offline 1=Online	394 (27.1%) 1,061 (72.9%)
Broadband category	1=0–25% of at least 50 Mbit/s 2=26–50% of at least 50 Mbit/s 3=51–75% of at least 50 Mbit/s 4=76–95% of at least 50 Mbit/s 5=96–99% of at least 50 Mbit/s	25 (1.7%) 287 (19.7%) 542 (37.3%) 545 (37.5%) 56 (3.8%)
Internet familiarity	1=Rarer 2=About once a week 3=More than once a week 4=About once a day 5=Several times a day	31 (2.1%) 39 (2.7%) 146 (10.0%) 265 (18.2%) 974 (66.9%)
Gender: Male	0=Female 1=Male	727 (50.0%) 728 (50.0%)
Year of birth	Mean (sd): 1968 (15.7) min < med < max: 1928 < 1967 < 1997 IQR (CV): 25 (0)	N = 1445
Education	1=Low 2=Medium 3=High	232 (15.9%) 502 (34.5%) 721 (49.6%)

each wave.³ Consistent with the previous analysis, the independent variable was classified according to the proportion of broadband supply with at least a 50 Mbit/s speed in the administrative district of the participant (1=0–25% of at least 50 Mbit/s, 2=26–50% of at least 50 Mbit/s, 3=51–75% of at least 50 Mbit/s, 4=76–95% of at least 50 Mbit/s, 5=96–99% of at least 50 Mbit/s).

Furthermore, in the panel attrition analysis, we included five control variables. The first was the evaluation of duration. At the end of each survey wave, the participants evaluated whether they experienced the questionnaire as long (1=Not at all, 2=Rather not, 3=Partially agree, 4=Rather, 5=Very). However, the evaluation of duration could not be included in the analysis as a time-dependent variable due to missing events (dropouts) in some of the response categories. The solution was to generate a categorized mean for each participant by calculating their mean evaluation of the duration across all waves and grouping this mean into one of the five original categories. We included the categorized mean of the evaluation of duration as a time-independent variable in the analysis. The second control variable was the actual duration in seconds. Due to the same issues as the time-dependent evaluation of duration, we calculated the mean duration for each participant and categorized it into four groups by dividing it into quartiles. We included the other three control variables—gender (0=Female, 1=Male), year of birth, and education (1=Low, 2=Medium, 3=High)—as time-independent variables. The lowest categories of broadband, evaluation of duration, and education, as well as female gender, serve as reference categories. See Table 2 in the Appendix for the descriptive statistics of all the variables we used in the panel attrition analysis.

³ The actual number of changes in the broadband category of a participant between waves was low because they only occurred when a participant changed residence.

Table 2 Descriptive statistics of the panel attrition analysis

Variable	Statistics/Values	Frequencies (%)
Panel dropout	0 = No dropout 1 = Dropout	12,882 (98.4%) 213 (1.6%)
Broadband category	1 = 0–25% of at least 50 Mbit/s 2 = 26–50% of at least 50 Mbit/s 3 = 51–75% of at least 50 Mbit/s 4 = 76–95% of at least 50 Mbit/s 5 = 96–99% of at least 50 Mbit/s	192 (1.5%) 2,550 (19.5%) 4,796 (36.6%) 4,950 (37.8%) 607 (4.6%)
Evaluation of duration	1 = Long–Not at all 2 = Long–Rather not 3 = Long–Partially agree 4 = Long–Rather 5 = Long–Very	1,210 (9.2%) 7,160 (54.7%) 4,057 (31.0%) 611 (4.7%) 57 (0.4%)
Measured duration (quartiles)	Mean (sd): 2.5 (1.1) min < med < max: 1 < 2 < 4 IQR (CV): 2 (0.4)	1: 3,296 (25.2%) 2: 3,328 (25.4%) 3: 3,264 (24.9%) 4: 3,207 (24.5%)
Gender: Male	0 = Female 1 = Male	6,360 (48.6%) 6,735 (51.4%)
Year of birth	Mean (sd): 1,968.8 (15.6) min < med < max: 1,933 < 1,968 < 1,997 IQR (CV): 24 (0)	N = 13,095
Education	1 = Low 2 = Medium 3 = High	1,586 (12.1%) 4,074 (31.1%) 7,435 (56.8%)

We modeled panel attrition with a Cox regression, which is a method that enabled us to fit multivariate survival models with time-dependent and time-independent covariates. To fit our Cox regression model with time-dependent covariates, we converted the data to a long format. Moreover, we split the episodes so that each participant had one row for each wave in which they participated (Broström, 2018, pp. 67–70). The purpose of a Cox regression model is to evaluate simultaneously the effects of multiple covariates on an event, which is dropout in the case of the panel attrition analysis. The Cox model is expressed by the hazard function, and its outcome can be interpreted as the risk of an event at each point in time. Like the logistic regression, it is common to take the exponential function of the coefficients of the Cox regression to inverse the logarithmic function, which provides hazard ratios that are easier to interpret. A hazard ratio above 1 indicates a covariate that is positively associated with an event probability and, thus, is negatively associated with length of survival.

A key assumption of the Cox regression model is that the hazard curves for groups of observations are proportional and do not cross, which is why it is also called the *proportional hazards model* (STHDA 2019). Using scaled Schoenfeld residuals, we tested the proportional hazards (PH) assumption underlying the Cox regression model.

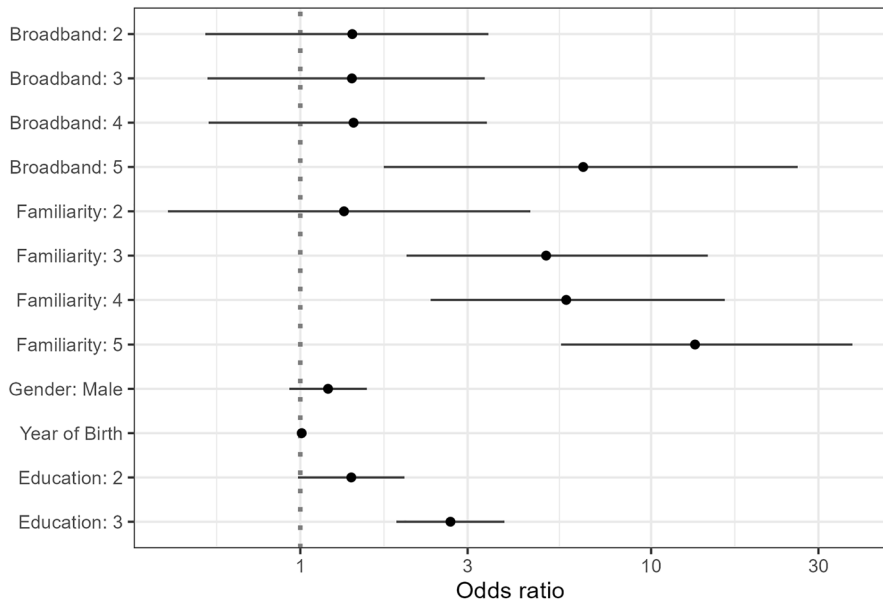


Fig. 2 Odds ratios of the covariates of the binomial logistic regression on mode choice (online) with 95% confidence intervals (Model 1). *Note:* Mode choice (offline), lowest categories of broadband, familiarity, and education, as well as female gender, serve as reference categories

4 Results

4.1 Mode choice

Figure 2 presents the results of the mode choice logistic regression model. The dots represent the odds ratios of each covariate of the binomial logistic regression for mode choice, and the horizontal lines represent the 95% confidence intervals. The odds ratio of the fifth broadband category is 6.405, which means that holding the other covariates at a fixed value, living in a region where 96–99% of the households have a data transmission rate of at least 50 Mbit/s (broadband category five) increases the odds of choosing the online mode by a factor of 6.405 ($p < 0.01$) compared to a region where 0–25% of the households have a data transmission rate of at least 50 Mbit/s (broadband category one). However, broadband categories two, three, and four are not statistically different from broadband category one. The odds ratios of internet familiarity ranges from 5.021 in the third category (private internet usage: more than once a week) to 13.342 in the fifth category (private internet usage: several times a day), which indicates that having an internet familiarity of category three, four, or five increases the odds of choosing the online mode significantly ($p < 0.002$) compared to an internet familiarity of category one. The control variables year of birth⁴ (1.001) and education: 3—which is high education—(2.680) also have a significant positive effect on choosing the online mode, whereas the coefficient of gender is not

⁴ In a separate model, we also tested categorizing year of birth into different generational cohorts but found no significant effects and no improvement in model fit.

Fig. 3 Marginal effects of broadband category on mode choice (Model 1)

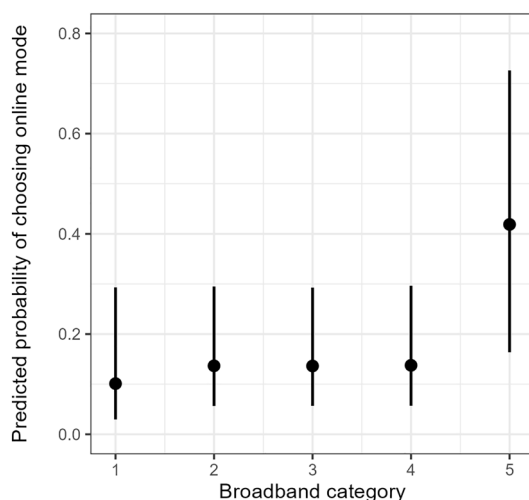
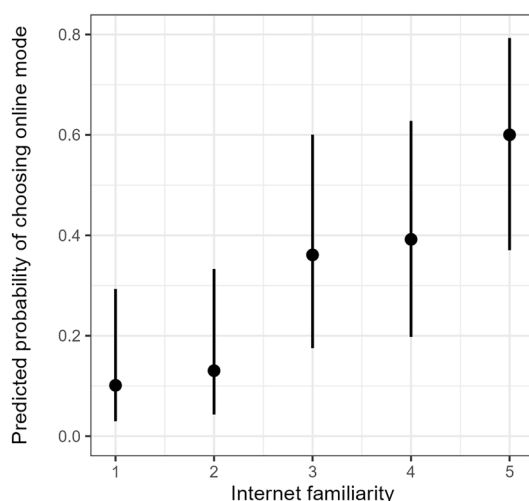


Fig. 4 Marginal effects of internet familiarity on mode choice (Model 1)



statistically significant and, therefore, has no relevant effect in this model. Consequently, hypotheses 1 (broadband supply), 2 (internet familiarity), 4 (age), and 5 (education) cannot be rejected, while hypothesis 3 (gender) must be rejected.

Figure 3 depicts the marginal effects of the independent variable broadband on mode choice. According to the predicted probabilities for broadband, categories one to four, which cover a data transmission rate of at least 50 Mbit/s for 0–95% of the households, indicate low probabilities (ranging from about 10% to 15%) of users choosing the online mode. However, in the highest broadband category, where 96–99% of the households have a data transmission rate of at least 50 Mbit/s, the probability of choosing the online mode is over 40%, and the regression model indicates that this effect is significantly different from the first category. Overall, there is a small difference in gradient between categories one to four and a large increase with wide confidence intervals in category five. Based on

Fig. 5 Marginal effects of education on mode choice (Model 1)

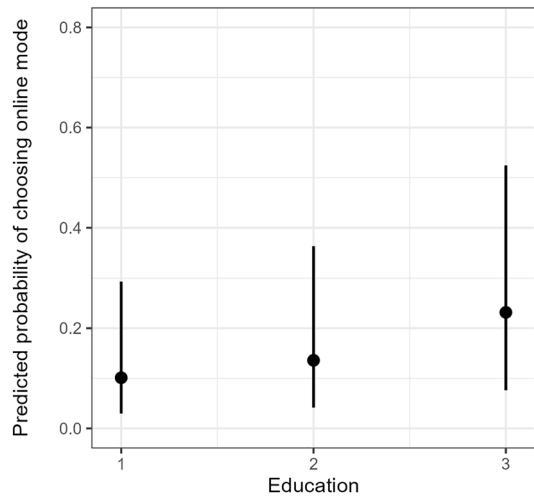
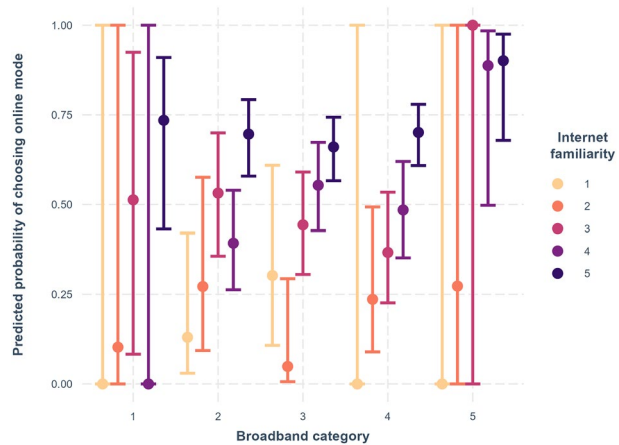


Fig. 6 Interaction of broadband and internet familiarity on mode choice (Model 2)



these findings, we find partial support for the hypothesis that living in a region with a better broadband supply increases the probability of choosing to participate online in a mixed-mode panel (hypothesis 1).

In Fig. 4, we see that there is a strong linear relationship between the lowest internet familiarity of about 10%, which includes participants who use the internet less than once a week, and the highest internet familiarity of over 60% of online mode choices, which includes participants who use the internet several times a day. However, we also see large standard deviations in each category of internet familiarity. The regression model shows that categories three, four, and five are significantly different from the first category. Overall, we find that higher internet familiarity increases the probability of choosing the online mode in a mixed-mode panel (hypothesis 2).

Figure 5 indicates a linear relationship of education on choosing the online mode, with the lowest level of education having a predicted probability of about 10% and the highest level of education having a predicted probability of about 22%. Again, we see large standard deviations in all three education categories, but the regression model shows that the

highest education category has a significantly higher effect on mode choice than the lowest education category. Overall, we find that a high level of education increases the probability of choosing the online mode in a mixed-mode panel (hypothesis 5).

In a second model, we include the interaction of broadband and internet familiarity to test whether the effect of broadband is dependent on internet familiarity. The interaction model shows no true interaction effects between broadband and internet familiarity (see Fig. 6). However, the model shows the tendency that the lower the level of internet familiarity, the higher the uncertainty in mode choice. This finding can be interpreted in conformity with the previously mentioned considerations that internet familiarity functions as a precondition for other predictors of mode choice.

Compared to the model without the interaction effect, the residual deviance decreased by 25.6 to 1,458.6, but the Akaike information criterion (AIC) increased by 0.4 to a value of 1,510.6, which can be interpreted as a deterioration of the model fit. Therefore, the first model fits the data best and serves as the basis for the hypothesis tests. See Table 3 in the Appendix for the complete table of results of both models.

4.2 Panel attrition

In Fig. 7, the dots represent the hazard ratios of each covariate of the Cox regression on panel dropout, and the horizontal lines represent the 95% confidence intervals. The hazard ratios of broadband categories two to five are not significantly distinct from broadband category one. This suggests that the varying proportions of households with a data transmission rate of at least 50 Mbit/s do not affect panel attrition. The hazard ratios of the evaluation of duration are at 6.107 ($p < 0.001$) for the fourth category and at 20.851 ($p < 0.001$) for the fifth category, indicating that evaluating the survey as long significantly increases the risk of dropout compared to not evaluating the survey as long (category one). The control variables measured duration (1.215) and year of birth (1.029) have a significant positive effect on panel attrition, whereas the coefficients of gender and education are not statistically significant and, therefore, have no relevant effects in this model. Consequently, hypotheses 7 (evaluation of duration), 8 (measured duration), and 10 (age) cannot be rejected, while hypotheses 6 (broadband supply), 9 (gender), and 11 (education) must be rejected.

Figure 8 shows the adjusted survival curves of the five broadband categories on panel dropout. The survival rates of each broadband category are almost identical, with a difference of about five percentage points between the first and fifth broadband categories after 16 survey waves. Given these results, we cannot confirm that living in a region with a better broadband supply decreases the risk of attrition in an online panel survey (hypothesis 6).

In contrast, the survival rates of the evaluation of duration are widely spread (see Fig. 9). After 16 waves, the survival rate for the participants who rated the survey as not at all long (first category) is about 88%, whereas the survival rate for participants who rated the survey as very long (fifth category) is about 10%. Thus, participants who rate the duration on average as long have an exceedingly higher risk of panel dropout than the participants who did not rate the duration as long (hypothesis 7).

While testing the proportional hazards assumption of model 1, the individual Schoenfeld test results revealed that each covariate had a p -value above 0.05, except for the year of birth and measured duration. As the curves appear to be sufficiently flat and the global Schoenfeld test yielded an insignificant result of $p = 0.205$, we do not see a

Table 3 Results table of the logistic regression models on mode choice

	DV: Mode choice—Online	
	(Model 1)	(Model 2)
Broadband: 2	0.341 (0.469)	14.778 (1,028.170)
Broadband: 3	0.339 (0.459)	15.842 (1,028.170)
Broadband: 4	0.349 (0.460)	0.995 (1,162.783)
Broadband: 5	1.857** (0.683)	1.189 (0.973)
Internet familiarity: 2	0.287 (0.599)	14.509 (543.076)
Internet familiarity: 3	1.614** (0.497)	16.729 (1,028.170)
Internet familiarity: 4	1.746*** (0.485)	0.584 (1,254.084)
Internet familiarity: 5	2.591*** (0.480)	17.696 (1,028.169)
Gender: Male	0.182 (0.130)	0.180 (0.131)
Year of birth	0.009* (0.005)	0.009 (0.005)
Education: 2	0.335 (0.178)	0.305 (0.181)
Education: 3	0.986*** (0.181)	0.963*** (0.183)
Broadband: 2×Familiarity: 2		− 13.594 (543.077)
Broadband: 3×Familiarity: 2		− 16.641 (543.078)
Broadband: 4×Familiarity: 2		NA
Broadband: 5×Familiarity: 2		NA
Broadband: 2×Familiarity: 3		− 14.702 (1,028.170)
Broadband: 3×Familiarity: 3		− 16.121 (1,028.170)
Broadband: 4×Familiarity: 3		− 1.592 (1,162.783)
Broadband: 5×Familiarity: 3		14.379 (1,455.398)
Broadband: 2×Familiarity: 4		0.879 (1,254.084)
Broadband: 3×Familiarity: 4		0.466 (1,254.084)
Broadband: 4×Familiarity: 4		15.038 (1,366.623)

Table 3 (continued)

	DV: Mode choice—Online	
	(Model 1)	(Model 2)
Broadband: 5×Familiarity: 4		16.969 (718.051)
Broadband: 2×Familiarity: 5		− 14.967 (1,028.169)
Broadband: 3×Familiarity: 5		− 16.196 (1,028.169)
Broadband: 4×Familiarity: 5		− 1.162 (1,162.783)
Broadband: 5×Familiarity: 5		
Constant	− 20.172* (8.920)	− 34.268 (1,028.209)
Observations	1,455	1,455
Null deviance	1,699.6	1,699.6
Residual deviance	1,484.2	1,458.6
AIC	1,510.2	1,510.6

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

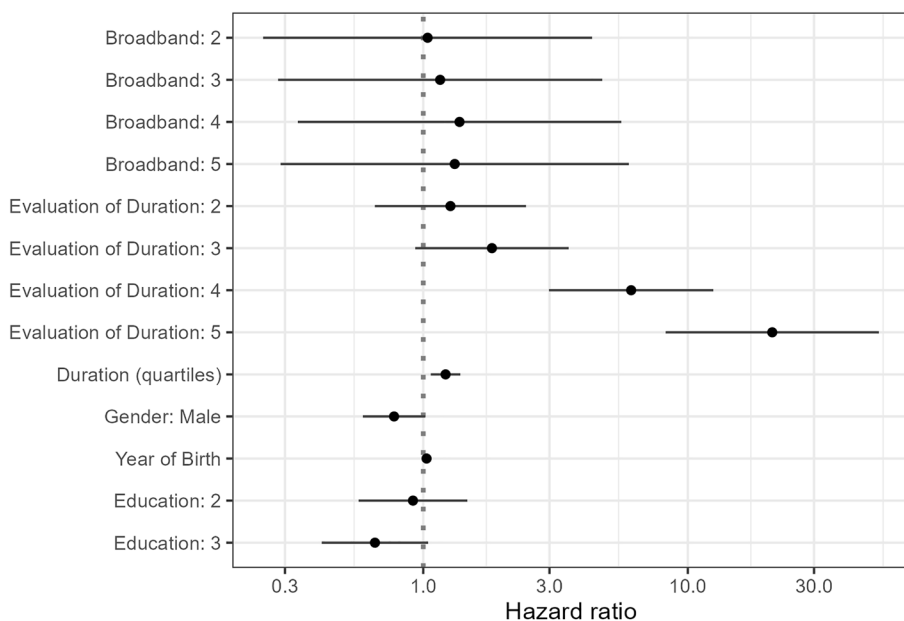


Fig. 7 Hazard ratios of the covariates of the Cox regression on panel dropout with 95% confidence intervals (Model 1). *Note:* No dropout, lowest categories of broadband, evaluation of duration, and education, as well as female gender, serve as reference categories

problem with the proportionality of the hazards. However, we tested categorizing the year of birth into different generational cohorts in a separate model. We did not find an

Fig. 8 Survival rate of broadband categories on panel dropout (Model 1)

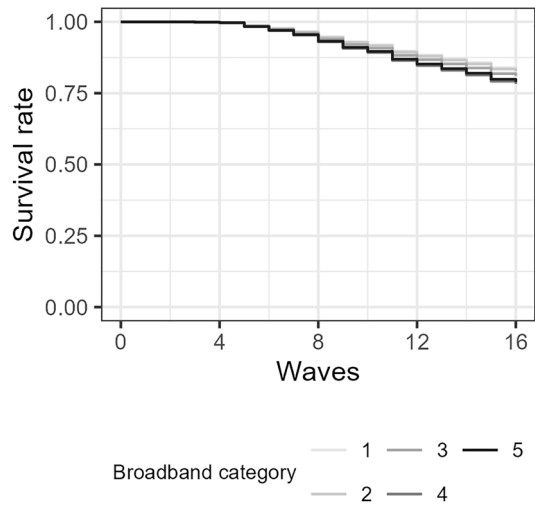
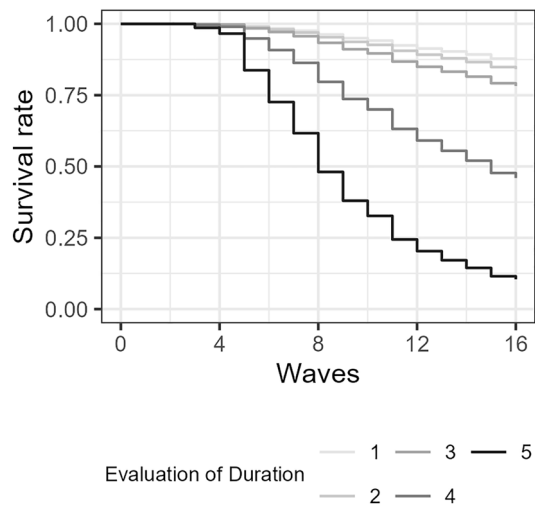


Fig. 9 Survival rate of evaluation of duration on panel dropout (Model 1)



improvement in the proportionality of the hazards. See Fig. 10 in the Appendix for the results of the proportional hazards tests and respective graphs.

In a further model, we tested for interaction effects between the broadband categories and the evaluation of duration categories. The tests for the global statistical significance of the model remained significant, but the model comparison with the LR test revealed that the interaction effects did not improve the model fit compared to the first model. See Table 4 in the Appendix for the complete table of results of both models of the panel attrition analysis.

As a robustness check, we ran the first model with two different subsets of the used dataset. The first subset excluded the participants who used a smartphone for two waves in a row, and the second subset excluded the participants who used a smartphone for three waves in a row. Smartphone participants can use the mobile network to answer the survey,

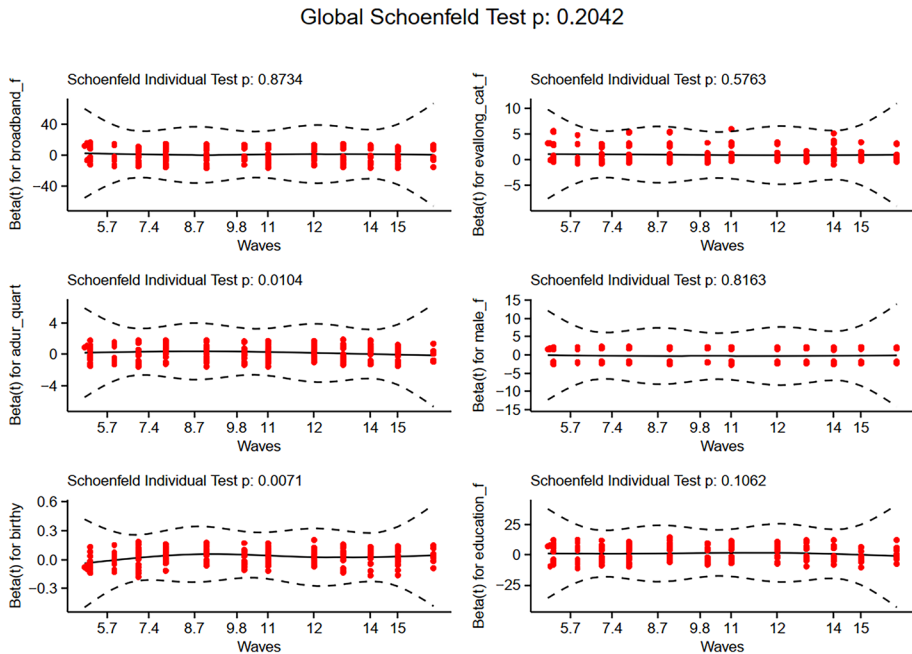


Fig. 10 Test of the proportional hazards assumption (Cox regression model 1)

bypassing the broadband network and associated limitations due to poor broadband supply. With these robustness checks with two subsets, we evaluated whether the use of smartphones, which enabled the use of both a broadband connection via a wireless network at home and a mobile network, impacted the effects of broadband supply. The results were similar to those of the complete dataset, which indicated that the findings were robust for survey participation via smartphones. See Table 5 in the Appendix for the results of the robustness check.

5 Conclusion and discussion

In the present study, we conducted a feasibility study to determine whether publicly available broadband data at the district level were sufficient to draw meaningful conclusions about participation behavior. The specific research focus was to explain how broadband supply affects the choice of participation mode in a mixed-mode panel survey and how it determines panel attrition. We will first conclude the results of the substantive research questions and then discuss the granularity of the broadband data in the context of the general availability of fine-grained geospatial data.

The literature review in the context of mode choice revealed that internet familiarity was an important precondition for choosing the online mode, followed by online experiences, which were closely linked to a fast and stable broadband connection. In light of this

Table 4 Results table of the Cox regression models on panel dropout

	DV: Event = Panel dropout	
	(Model 1)	(Model 2)
Broadband: 2	0.038 (0.731)	13.471 (1,723.302)
Broadband: 3	0.147 (0.720)	13.128 (1,723.302)
Broadband: 4	0.316 (0.718)	12.971 (1,723.302)
Broadband: 5	0.274 (0.773)	14.554 (1,723.303)
Evaluation of duration: 2	0.237 (0.336)	13.035 (1,723.303)
Evaluation of duration: 3	0.598 (0.341)	14.070 (1,723.303)
Evaluation of duration: 4	1.809*** (0.365)	0.857 (1.167)
Evaluation of duration: 5	3.037*** (0.473)	1.761 (1.435)
Duration (Quartiles)	0.194** (0.066)	0.192** (0.067)
Gender: Male	-0.254 (0.139)	-0.266 (0.142)
Year of birth	0.029*** (0.005)	0.029*** (0.005)
Education: 2	-0.089 (0.242)	-0.042 (0.244)
Education: 3	-0.420 (0.236)	-0.405 (0.238)
Broadband: 2 × Eval. of Dur.: 2		-13.461 (1,723.303)
Broadband: 3 × Eval. of Dur.: 2		-12.883 (1,723.303)
Broadband: 4 × Eval. of Dur.: 2		-12.292 (1,723.303)
Broadband: 5 × Eval. of Dur.: 2		-14.102 (1,723.303)
Broadband: 2 × Eval. of Dur.: 3		-13.824
Broadband: 3 × Eval. of Dur.: 3		-13.231 (1,723.303)
Broadband: 4 × Eval. of Dur.: 3		-13.296 (1,723.303)
Broadband: 5 × Eval. of Dur.: 3		-15.291 (1,723.303)
Broadband: 2 × Eval. of Dur.: 4		0.781 (1.406)

Table 4 (continued)

	DV: Event = Panel dropout	
	(Model 1)	(Model 2)
Broadband: 3 × Eval. of Dur.: 4		0.659 (1.305)
Broadband: 4 × Eval. of Dur.: 4		1.352 (1.334)
Broadband: 5 × Eval. of Dur.: 4		NA (0.000)
Broadband: 2 × Eval. of Dur.: 5		0.309 (1.750)
Broadband: 3 × Eval. of Dur.: 5		1.490 (1.599)
Broadband: 4 × Eval. of Dur.: 5		1.963 (1.695)
Broadband: 5 × Eval. of Dur.: 5		NA (0.000)
Number of Events	213	213
Observations	13,095	13,095
Model comparison: Pr(> Chi)		0.745
Wald Test	166.160*** (df = 13)	175.360*** (df = 27)
LR Test	133.618*** (df = 13)	143.848*** (df = 27)
Score (Logrank) Test	234.628*** (df = 13)	271.382*** (df = 27)

The model comparison tests the improvement of model fit compared to the previous model by an LR test

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

review, we expected that living in a region with a better broadband supply would increase the probability of choosing the online mode in a mixed-mode panel survey (hypothesis 1). Our results show that only the highest broadband category, where the district has 96–99% broadband coverage of at least 50 Mbit/s, significantly increases the likelihood of choosing the online mode compared to districts with 0–25% broadband coverage of at least 50 Mbit/s. With these results, we can partially confirm the first hypothesis since it is true, at least for the outermost categories, that living in a region with a better broadband supply increases the probability of choosing the online mode in a mixed-mode panel survey. Furthermore, we can confirm previous research findings, as our model shows that higher internet familiarity (hypothesis 2), younger age (hypothesis 4), and higher education (hypothesis 5) increase the likelihood of choosing the online mode in a mixed-mode panel survey, with the effect of higher internet familiarity appearing to be particularly strong.

The literature review in the context of panel attrition revealed that survey burden and panel experience play a central role in determining panel dropouts. The crucial factor in this context is the perception of burden, which is influenced by flow experience connected to waiting times and the overall perceived survey duration. Therefore, we expected that living in a region with a better broadband supply would decrease the risk of attrition in online panel surveys (hypothesis 6). The sixth hypothesis was rejected by the results, as there was no significant effect of broadband supply on panel attrition.

Table 5 Robustness check of the Cox regression on panel dropout

	DV: Event = Panel dropout	
	Subset 1	Subset 2
	(1)	(2)
Broadband: 2	− 0.176 (0.737)	− 0.124 (0.735)
Broadband: 3	0.011 (0.722)	0.042 (0.721)
Broadband: 4	0.167 (0.722)	0.187 (0.721)
Broadband: 5	− 0.051 (0.789)	0.185 (0.775)
Evaluation of duration: 2	0.128 (0.339)	0.141 (0.337)
Evaluation of duration: 3	0.567 (0.347)	0.521 (0.345)
Evaluation of duration: 4	1.790 ^{***} (0.373)	1.705 ^{***} (0.372)
Evaluation of duration: 5	2.702 ^{***} (0.509)	2.864 ^{***} (0.491)
Duration (Quartiles)	0.205 ^{**} (0.072)	0.211 ^{**} (0.070)
Gender: Male	− 0.403 ^{**} (0.152)	− 0.361 [*] (0.146)
Year of birth	0.039 ^{***} (0.006)	0.036 ^{***} (0.006)
Education: 2	− 0.126 (0.259)	− 0.051 (0.251)
Education: 3	− 0.430 (0.254)	− 0.437 (0.249)
Number of events	182	195
Observations	9,874	10,709
Wald test (df = 13)	166.470 ^{***}	173.010 ^{***}
LR test (df = 13)	143.193 ^{***}	144.015 ^{***}
Score (Logrank) Test (df = 13)	223.616 ^{***}	242.010 ^{***}

Subset 1 Excluded participants that used the smartphone for at least two waves in a row

Subset 2 Excluded participants that used the smartphone for at least three waves in a row

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Besides that, the analysis shows that the probability of dropping out increases the more the survey is evaluated as long (hypothesis 7) and the longer the actual duration of the survey is (hypothesis 8). Also, younger respondents have a higher risk of dropping out (hypothesis 10).

Besides examining the effects of broadband supply on mode choice and panel attrition, the present study was designed as a feasibility study to test whether publicly available geospatial broadband data at the district level were sufficient for this type of analysis.

There is an ongoing discussion about the availability of geospatial data for research purposes and the associated privacy concerns (Solymosi et al. 2023). Fine-grained geospatial data offers the potential for precise analyses but also carries a high risk of identifying personal identities. In addition, fine-grained geospatial data is often not available for research purposes or subject to severe restrictions, while larger-scale geospatial data is more accessible. In the case of the broadband supply data used, the district level was accessible online, whereas we could not gain access to the more detailed grid level when designing the study.

In view of the results, the classification of about 400 administrative districts in Germany into five broadband categories was not ideal for the present analyses. Although the analysis of mode choice shows a significant effect for the highest broadband category, we expect that better broadband data would provide a more accurate picture of the impact of broadband supply. Such better broadband data would include both finer spatial units and more information about the different levels of broadband speed available per unit. Therefore, we cannot reject hypothesis 12 since the district-level data has produced meaningful results. However, we do not recommend focusing solely on district-level data for such research questions, although it was important to test this in a feasibility study. The approach of using publicly available geodata to optimize survey participation behavior has a high potential for improving survey quality with relatively simple means and low costs. When fine-grained geospatial data is limited and restricted, larger-scale geospatial data with fewer barriers can be used to find opportunities for survey optimization. This study shows that it might be reasonable to replicate these analyses—especially the mode choice part—with more fine-grained broadband data to test whether the insignificance of the broadband coefficients is attributable to the large granularity of the broadband data.

To conclude, this field of research is becoming more and more important due to the increasing number of web surveys and the ambition of full coverage of the population with web surveys. Based on the mode choice analysis results, survey researchers can extract practical implications. For instance, we can assume that in mixed-mode scenarios, respondents in areas with very good broadband supply are more likely to choose the online mode. Not only does this help in predicting online participation, but it also allows for a targeted approach to increasing online participation. Unfortunately, it is not possible to deduce any practical implications regarding broadband supply from the results of the panel attrition analysis. Therefore, the approach of using geospatial broadband data to investigate participation behavior should be re-examined with more precise data on broadband supply. In the course of this, new types of data, such as digital behavioral data, can also be examined in order to obtain more in-depth insights into, for example, familiarity, internet usage, and online habits. Findings from this research can be especially useful in survey recruitment when it is possible to offer potential participants the most suitable survey mode based on their place of residence, thereby improving participation rates and reducing panel attrition.

Appendix

See Fig. 10.

See Tables 1, 2, 3, 4, 5.

Author contributions All authors contributed to the study conception and design. Data preparation and analysis were performed by Maikel Schwerdtfeger. The first draft of the manuscript was written by Maikel Schwerdtfeger and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. A previous version of this manuscript was presented at the virtual General Online Research Conference (GOR) in 2021.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bandilla, W., Kaczmarek, L., Blohm, M., Neubarth, W.: Coverage- und Nonresponse-Effekte bei Online-Bevölkerungsumfragen. In: Jakob, N., Schoen, H., Zerback, T. (eds.) *Sozialforschung im internet: methodologie und Praxis der Online-Befragung*, pp. 129–143. VS Verlag für Sozialwissenschaften, Wiesbaden (2009). https://doi.org/10.1007/978-3-531-91791-7_8
- Baur, N., Florian, M.J.: Stichprobenprobleme bei Online-Umfragen. In: Jakob, N., Schoen, H., Zerback, T. (eds.) *Sozialforschung im Internet: Methodologie und praxis der online-befragung*, pp. 109–128. VS Verlag für Sozialwissenschaften, Wiesbaden (2009). https://doi.org/10.1007/978-3-531-91791-7_7
- Behr, A., Bellgardt, E., Rendtel, U.: Extent and determinants of panel attrition in the European community household panel. *Eur. Sociol. Rev.* **21**(5), 489–512 (2005)
- Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., Weyandt, K.W.: Establishing an open probability-based mixed-mode panel of the general population in Germany: the GESIS Panel. *Soc. Sci. Comput. Rev.* **36**(1), 103–115 (2018)
- Bretsch, D., Weiß, B.: How do internet-related characteristics affect whether members of a german mixed-mode panel switch from the mail to the web mode? *Soc. Sci. Comput. Rev.* (2022). <https://doi.org/10.1177/08944393221117267>
- Broström, G.: *Event history analysis with R*. CRC Press, Boca Raton (2018)
- Brutlag, J. (2009). Speed Matters. *Google AI Blog*. <http://ai.googleblog.com/2009/06/speed-matters.html>
- Bundesamt für Kartographie und Geodäsie. (2016). *Breitbandversorgung mit 50 Mbit/s*. https://www.geoportal.de/SharedDocs/Karten/DE/Themenkarte_Breitbandversorgung.html
- Bundesministerium für Digitales und Verkehr. (2023). *Breitbandatlas*. <https://gigabitgrundbuch.bund.de/GIGA/DE/Breitbandatlas/Vollbild/start.html>
- Chebat, J.-C., Salem, N.H., Poirier, J.-F., Gélinas-Chebat, C.: Reactions to waiting online by men and women. *Psychol. Rep.* **106**(3), 851–869 (2010). <https://doi.org/10.2466/pr0.106.3.851-869>
- Crawford, S.D., Couper, M.P., Lamias, M.J.: Web surveys: perceptions of Burden. *Soc. Sci. Comput. Rev.* **19**(2), 146–162 (2001). <https://doi.org/10.1177/089443930101900202>
- Deutskens, E., de Ruyter, K., Wetzels, M., Oosterveld, P.: Response rate and response quality of internet-based surveys: an experimental study. *Mark. Lett.* **15**(1), 21–36 (2004). <https://doi.org/10.1023/B:MARK.0000021968.86465.00>
- Diment, K., Garrett-Jones, S.: How demographic characteristics affect mode preference in a postal/web mixed-mode survey of Australian researchers. *Soc. Sci. Comput. Rev.* **25**(3), 410–417 (2007)
- Galesic, M.: Dropouts on the web: effects of interest and burden experienced during an online survey. *J. Off. Stat.* **22**(2), 313 (2006)
- Ganassali, S.: The influence of the design of web survey questionnaires on the quality of responses. *Surv. Res. Methods* **2**, 21–32 (2008)

- GESIS. (2019). GESIS Panel—Extended Edition. *GESIS Data Archive, Cologne, ZA5664 Data File (Version 31.0.0)*. <https://doi.org/10.4232/1.13319>
- Gummer, T., Daikeler, J.: A note on how prior survey experience with self-administered panel surveys affects attrition in different modes. *Soc. Sci. Computer Rev.* (2018). <https://doi.org/10.1177/0894439318816986>
- Herzing, J.M.E., Blom, A.G.: The influence of a person's digital affinity on unit nonresponse and attrition in an online panel. *Soc. Sci. Computer Rev.* **37**(3), 404–424 (2019). <https://doi.org/10.1177/0894439318774758>
- Internet Archive. (2018, October 9). *Geoportal.de—Geodaten aus Deutschland—Themenkarten—Breitbandversorgung mit 50 Mbit/s*. https://web.archive.org/web/20181009231108/https://www.geoportal.de/SharedDocs/Karten/DE/Themenkarte_Breitbandversorgung.html
- Lepkowski, J.M., Couper, M.P.: Nonresponse in the second wave of longitudinal household surveys. In: *Survey nonresponse*, pp. 259–272. John Wiley, Hoboken (2002)
- Lipps, O.: Attrition of households and individuals in panel surveys. *Soeppapers* **164**, 1–15 (2009)
- Lutig, P.: Panel attrition: separating stayers, fast attriters, gradual attriters, and lurkers. *Sociol. Methods Res.* **43**(4), 699–723 (2014)
- Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., Schütz, A.: Compensating for low topic interest and long surveys: a field experiment on nonresponse in web surveys. *Soc. Sci. Comput. Rev.* **25**(3), 372–383 (2007)
- Millar, M.M., O'Neill, A.C., Dillman, D.A.: Are mode preferences real. *Pullman: Washington State Univ.* **9**, 2–41 (2009)
- Nielsen, J.: *Usability engineering*. Morgan Kaufmann, Cambridge (1994)
- Olson, K., Smyth, J.D., Wood, H.M.: Does giving people their preferred survey mode actually increase survey participation rates? An experimental examination. *Public Opin. Quarterly* **76**(4), 611–635 (2012). <https://doi.org/10.1093/poq/nfs024>
- R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reichl, P., Egger, S., Schatz, R., D'Alconzo, A.: The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment. *IEEE Int. Conf. Commun.* **2010**, 1–5 (2010)
- Schaurer, I., Minderop, I., Bretsch, D., & Weyandt, K. (2020). *GESIS Panel Technical Report: Recruitment 2018 (Wave f11 and f12)* (p. 24). GESIS - Leibniz Institute for the Social Sciences. <https://dbk.gesis.org/dbksearch/download.asp?id=68673>
- Schaurer, I., & Weyandt, K. (2018). *GESIS Panel Technical Report: Recruitment 2016 (Wave d11 and d12)* (p. 28). GESIS - Leibniz Institute for the Social Sciences. <https://dbk.gesis.org/dbksearch/download.asp?id=63525>
- Scherpenzeel, A. C., & Bethlehem, J. G. (2011). How representative are online panels? Problems of coverage and selection and possible solutions. *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, pp. 105–132.
- Smyth, J.D., Olson, K., Millar, M.M.: Identifying predictors of survey mode preference. *Soc. Sci. Res.* **48**, 135–144 (2014). <https://doi.org/10.1016/j.ssresearch.2014.06.002>
- Solymosi, R., Buil-Gil, D., Ceccato, V., Kim, E., Jansson, U.: Privacy challenges in geodata and open data. *Area* **55**(4), 456–464 (2023)
- STHDA. (2019). *Cox Proportional-Hazards Model*. STHDA - Statistical Tools for Data Analysis and Visualization. <http://www.sthda.com/english/wiki/cox-proportional-hazards-model>
- Struminskaya, B. (2014). *Data quality in probability-based online panels: Nonresponse, attrition, and panel conditioning* [Utrecht University]. <https://dspace.library.uu.nl/handle/1874/301751>
- Vicente, P., Reis, E.: Using questionnaire design to fight nonresponse bias in web surveys. *Soc. Sci. Comput. Rev.* **28**(2), 251–267 (2010). <https://doi.org/10.1177/0894439309340751>
- Watson, N., Wooden, M.: Identifying factors affecting longitudinal survey response. *Methodol. Longitud. Surv.* **1**, 157–182 (2009)
- Yu, M., Zhou, R., Cai, Z., Tan, C.-W., Wang, H.: Unravelling the relationship between response time and user experience in mobile applications. *Internet Res.* **30**(5), 1353–1382 (2020)