Concurrent, Web-First, or Web-Only? How Different Mode Sequences Perform in Recruiting Participants for a Self-Administered Mixed-mode Panel Study

Christmann, Pablo; Gummer, Tobias; Häring, Armando; Kunz, Tanja; Oehrlein, Anne-Sophie; Ruland, Michael; Schmid, Lisa

*Journal of Survey Statistics and Methodology (2024) **00**, 1–26*

CONCURRENT, WEB-FIRST, OR WEB-ONLY? HOW DIFFERENT MODE SEQUENCES PERFORM IN RECRUITING PARTICIPANTS FOR A SELF-ADMINISTERED MIXED-MODE PANEL STUDY

PABLO CHRISTMANN *
TOBIAS GUMMER
ARMANDO HÄRING
TANJA KUNZ
ANNE-SOPHIE OEHRLEIN
MICHAEL RULAND
LISA SCHMID

During the COVID-19 pandemic, many survey programs switched to self-administered modes of data collection, often offering respondents both web-based and paper-based questionnaires. However, there is little guidance as to when to offer which mode, especially when the aim is to recruit participants for a panel survey. This study examines the effectiveness of different mode-sequence designs by analyzing an experiment implemented in the recruitment wave of the German panel study "Family Research and Demographic Analysis." We randomly assigned 108,256 individuals aged 18–49 years to one of three mode-sequence-design conditions: concurrent, web-first including a paper-based questionnaire with the second reminder (web-first-2), and web-first including a paper-based questionnaire with the third reminder (web-first-3). A fourth, simulated group did not receive a paper-based

Pablo Christmann is a Senior Researcher with the GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany. Tobias Gummer is a Senior Researcher with the GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany and Professor at the School of Social Sciences, University of Mannheim, Mannheim, Germany. Armando Häring is a Senior Researcher with the infas Institute for Applied Social Science, Bonn, Germany. Tanja Kunz is a Senior Researcher with the GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany. Anne-Sophie Oehrlein is a Doctoral Researcher with the GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany. Michael Ruland is a Senior Researcher with the infas Institute for Applied Social Science, Bonn, Germany. Lisa Schmid is a Senior Researcher with the GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany.

This work was funded by the German Federal Ministry of Education and Research (BMBF) as part of FReDA (grant number 01UW2001B).
This study design and analysis were not preregistered.

*Address correspondence to Pablo Christmann, GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany; E-mail: pablo.christmann@gesis.org.

https://doi.org/10.1093/jssam/smae008

questionnaire (web-only). We analyzed how different mode-sequence designs affected outcome rates, sample composition, response distributions, data quality, share of paper-based questionnaires, postage costs, number of postal mailings in the recruitment wave, and participation in the first regular wave. Our results show no differences in response distributions and small differences in sample composition across the four mode-sequence designs. As the web-first-2 and simulated web-only designs yielded comparatively good response and recruitment rates at reasonable costs, we recommend their use when surveying adults between 18 and 49 years old.

KEY WORDS: Mixed-mode; Mode-choice design; Panel survey; Recruitment experiment; Self-administered modes.

Statement of Significance

This study examines the effectiveness of four mode-sequence designs by analyzing an experiment implemented in the recruitment wave of the German panel study "Family Research and Demographic Analysis." The first group received a paper-based questionnaire with the invitation letter. The second group received a paper-based questionnaire with the second reminder only, while the third group received a paper-based questionnaire only with an additional third reminder. A fourth simulated group did not receive a paper-based questionnaire. Our results show no differences in response distributions and little differences in sample composition across the designs. As the second and fourth groups yielded good response and recruitment rates at reasonable costs, we recommend their use when surveying adults aged 18–49 years.

1. INTRODUCTION

Even before the outbreak of the COVID-19 pandemic, large-scale cross-sectional survey programs, such as the Generations and Gender Programme, the European Social Survey, and the European Values Study, had started experimenting with self-administered modes (e.g., Cernat and Revilla 2021; Luijkx et al. 2021; Wolf et al. 2021; Cernat et al. 2022; Piccitto et al. 2022). This was prompted primarily by declining response rates and the rising cost of interviewer-administered modes (de Leeuw and de Heer 2002; Brick and Williams 2013; de Leeuw et al. 2018; Olson, Smyth et al. 2021; Wolf et al. 2021). Disruptions to face-to-face data collection during the pandemic also forced ongoing and planned survey programs to switch to self-administered modes (e.g., the German Institute for Employment Research Establishment

Panel [Sakshaug et al., 2020] and Understanding Society—The UK Household Longitudinal Study [Burton et al. 2020]).

Some probability-based panel studies in the social sciences have been using self-administered modes to survey their panel members for about 10 to 15 years. Examples include the Longitudinal Internet Studies for the Social Sciences (LISS) panel in the Netherlands and the German Internet Panel (GIP) and the GESIS Panel in Germany. These panel studies rely on self-administered modes mainly for reinterviews in the ongoing panel. In the recruitment stages, interviewers usually conduct interviews face-to-face and directly obtain respondents' consent to join the panel (e.g., Blom et al. 2015; Bosnjak et al. 2018). During recruitment, interviewers serve two main purposes: first, to motivate respondents to give their consent to be reinterviewed (i.e., to increase panel consent rates); second, to establish a relationship between the panel study and the respondents (i.e., to increase panel retention rates). However, declining response rates and increasing survey costs associated with the face-to-face mode are also a challenge for this recruitment strategy. Consequently, probability-based panel studies such as the GIP have recently begun exploring whether recruitment surveys can also be conducted in self-administered (mixed) modes rather than face-to-face (Cornesse et al. 2022).

There is considerable leeway regarding the modes that can be offered to respondents to participate in a survey and—if multiple modes are offered—their optimal sequencing. The most common self-administered modes in general population surveys rely on web-based and paper-based questionnaires. Several different strategies for designing mode sequences can be distinguished (de Leeuw 2018). In mixed-mode designs, respondents are offered a choice of multiple modes to participate in the survey. If there are several contact attempts, a basic distinction can be made between concurrent and sequential mixed-mode designs. In a concurrent (or simultaneous) mixed-mode design, respondents are offered several modes at the same time during the initial contact. In a sequential mixed-mode design, only one mode is offered initially, and other modes are offered only in later contact attempts (Dillman 2017). If in this case a web-based questionnaire is offered first, the strategy is referred to as "web-first," "push-to-web," or "web-push." If a paper-based questionnaire is offered first, it is typically called "paper-first." In a single-mode design, by contrast, only one mode is offered across all contact attempts, and, therefore, respondents do not have a choice of modes. If only the web mode is offered, the strategy is often called "web-only." In panel studies, mode-sequence-design decisions are particularly important because they can affect not only participation behavior in the recruitment wave but also in subsequent waves.

The present study contributes to the literature by analyzing a survey experiment on the effects of four different mode-sequence designs (one concurrent, two web-first, and one web-only) for the recruitment of participants for a panel

study. We perform several analyses based on a comprehensive set of key performance indicators, including response rates, sample composition, distributions of substantive variables, data quality, share of paper-based questionnaires, and survey costs. In addition, we analyze how these mode-sequence designs impact panel consent, recruitment rates, and participation behavior in the first regular wave. Thus, we ask two research questions:

(1) How does the mode-sequence design affect participation behavior and key performance indicators in the panel recruitment wave?
(2) How does the mode-sequence design in the recruitment wave affect panel consent and participation behavior in the first regular wave?

2. PREVIOUS RESEARCH

In the total survey error (TSE) framework, Groves et al. (2009) distinguish two dimensions of possible sources of error that can affect the quality of survey statistics: measurement and representation. The representational dimension comprises all steps that lead from the target population to the survey respondents and the extent to which the latter accurately reflect the characteristics of the target population. Regarding representation, our primary interest lies in evaluating nonresponse and nonresponse error to determine whether survey statistics are impacted by varying participation within the different mode-sequence-design groups. In survey practice, nonresponse is often assessed through response rates, although Groves (2006) and Groves and Peytcheva (2008) demonstrated that there is no strong statistical relationship between nonresponse bias and response rates.

Studies based on cross-sectional surveys have found that "web-intensive" mode-sequence designs have negative effects on overall response rates. Wolf et al. (2021) reported a higher response rate with a concurrent design compared with a sequential web-first design. Holmberg et al. (2010) found that a concurrent design achieved higher response rates than web-first and paper-only designs. Patrick et al. (2018) found that a web-first design yielded lower overall response rates compared with paper-only and sequential paper-first designs. In the 2018 recruitment round of the GIP, Cornesse et al. (2022) tested four self-administered mode-sequence designs: concurrent, web-first, paper-first, and web-only. They found that the response rate for their recruitment survey was highest in the concurrent design followed by the sequential paper-first and web-first designs. The lowest response rate was in the web-only design.

Previous research shows that offering respondents different ways of participating in a survey increases their likelihood of doing so. Each mode caters to different respondent preferences (e.g., for privacy, which may reduce their motivation to participate via the web) and abilities (e.g., to use a smartphone

or computer to participate via the web). The various mode-sequence designs differ in the number of options that are offered simultaneously or sequentially: concurrent designs offer at least two modes from the start, whereas sequential designs usually offer only one mode initially and further modes in subsequent contact attempts. Consequently, in the present study, we expect that the concurrent design will lead to a higher response rate than the two web-first designs, and that the web-only design will lead to a lower response rate than the other three designs.

There is limited research on the recruitment of participants for self-administered mixed-mode panel studies. As part of the 2018 GIP recruitment process, participants in an initial recruitment survey were invited to take part in a subsequent online registration survey, where they were asked for their consent to become a member of the online panel (Cornesse et al. 2022). To evaluate the success of the panel recruitment process, the authors estimated panel registration rates for each initial experimental group, which they defined as the number of completed interviews in the online registration survey divided by the number of eligible cases in the gross sample of the recruitment survey. The highest panel registration rate was achieved in the web-only design, followed by the web-first and the concurrent designs; the paper-first design lagged far behind. One limitation of the study by Cornesse et al. (2022) is that panel registration was possible only via an online registration survey, even if respondents had chosen the paper-based option in the initial recruitment survey.

Other studies that also relied on panel data mixed interviewer-administered modes (i.e., face-to-face, telephone) and self-administered modes (e.g., Klausch et al. 2017; Mauz et al. 2018; Sakshaug et al. 2019; Lynn 2020; Legleye and Charrance 2021). For example, Rao et al. (2010) found that a sequential paper-first design with a telephone follow-up resulted in higher response and recruitment rates compared with a telephone-only mode. However, using telephone follow-up interviews impairs the generalizability of the results to panels recruited using only self-administered modes. Based on the limited body of research, we were unable to formulate clear expectations on how the different mode-sequence designs in our study would impact recruitment rates.

Approaches to assessing nonresponse error often compare survey data with auxiliary data from the gross sample, for example, R-indicators (Schouten et al. 2009); or they compare sociodemographic characteristics with reliable population statistics to indicate sample balance. Again, evidence is limited, although Gummer et al. (2022) and Cornesse et al. (2022) found that different self-administered mode-sequence designs led to similarly representative samples in terms of sociodemographic characteristics. Whereas both studies focused on the general population, we focus on a younger population (aged 18–49 years) who are likely to be more tech savvy, and who live in Germany, a country with high internet penetration and smartphone coverage rates

(Poushter 2016; Taylor and Silver 2019). Thus, we expect no large differences in sample composition across the four mode-sequence designs, as most respondents will be able to participate via a web-based questionnaire.

The second dimension of the TSE framework is measurement. Mixed-mode designs face the risk of "measurement mode effects," which may arise from variations in question presentation across modes that may cause systematically different answers to the same questions from the same respondents (de Leeuw 2005). Previous research has shown that mode effects are more problematic when mixing between rather than within interviewer-administered and self-administered modes (Klausch et al. 2013; Cernat et al. 2016). As we compare mode-sequence designs that rely solely on self-administered modes, we expect that substantive key variables will be similarly distributed between these different designs.

Another indicator for measurement error is item nonresponse. A recent meta-analysis on item nonresponse in survey modes found it to be higher in paper-based questionnaires than in web-based questionnaires (Čehovin et al. 2023). Patrick et al. (2018) found that item nonresponse was lower in web-first designs than in paper-first and paper-only designs, and that it was generally higher in paper-based than in web-based questionnaires. We reason that completing a questionnaire via the web is less difficult for respondents, as the questionnaire is displayed in sections (on separate screens), and respondents are not burdened with branching instructions. Due to the lower respondent burden in web surveys, we expect that fewer web respondents will use item nonresponse as a strategy to make the response process easier (e.g., Olson et al. 2018; Silber et al. 2021). In mode-sequence designs that elicit a higher share of paper-based questionnaires (see below), more item nonresponse should occur. Consequently, we expect item nonresponse to be highest in the concurrent design, followed by the two web-first designs and the web-only design.

Mixing modes means respondents have different ways of answering questions (i.e., using web-based or paper-based questionnaires). Mode sequence has been shown to impact the share of respondents in each mode. Patrick et al. (2018) reported that a web-first design encouraged participation via the web, which Holmberg et al. (2010) and Wolf et al. (2021) also found to be substantially higher in a web-first than in a concurrent design. We reason that this is simply the result of offering respondents the chance to participate in these modes: If parts of the sample have already been pushed to web in earlier contacts in a sequential design, fewer invitees are later offered the possibility to use a paper-based questionnaire. Based on this reasoning and on existing evidence, we expect the concurrent design to result in a higher share of paper-based questionnaires than the web-first designs.

Finally, the lack of research on recruiting participants for self-administered mixed-mode panel studies is compounded by a general lack of research on survey costs. As argued by Olson et al. (2021), considering survey costs is important, as they constrain survey design and operation decisions (e.g., incentive

amount, number of postal mailings, use of paper-based questionnaires). Although cost calculations are seldom comparable, the share of paper-based questionnaires is one of the main cost drivers, and web-intensive mode-sequence designs have thus been shown to be more cost-efficient (Holmberg et al. 2010; Patrick et al. 2018; Gummer et al. 2022). In the present study, depending on how often paper-based questionnaires are sent to invitees, we expect the concurrent mode-sequence design to be more expensive than the two web-first designs, and the two web-first designs to be more expensive than the web-only design.

3. METHODS AND EXPERIMENTAL DESIGN

3.1 Data

We implemented an experiment on mode-sequence designs in the 2021 recruitment wave of "Family Research and Demographic Analysis" (FreDA)—The German Family Demography Panel Study, a new data infrastructure project in Germany (Schneider et al. 2021). Data for the present study come mainly from the first recruitment wave (W1R) of FReDA, in which 37,783 respondents participated in a short survey of approximately 10 minutes (Bujard et al. 2022). Replication materials are available at: https://doi.org/10.7802/2648. A Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA) checklist (Seidenberg et al. 2023) is provided in table A1 in the supplementary data online.

FReDA is designed as a probability-based self-administered mixed-mode panel study with biannual surveys using both web-based and paper-based questionnaires. The two-stage sampling procedure for FReDA started in 2020 by drawing a stratified random sample comprising 320 sampling points in 268 municipalities. This selection process followed a probability-proportional-to-size (PPS) sampling approach considering the municipality's population aged 18–49 years. In the second stage, addresses were randomly selected from each sampling point. In the case of household duplicates—that is, more than one person from the same household—one person was randomly selected and the other was excluded. To account for individuals' unequal inclusion probabilities, design weights calculated as the inverse of the inclusion probability are provided. The sampling procedure resulted in a gross sample of $N = 108{,}256$. To avoid potential regional bias, the experimental groups within each sampling point were randomized.

In April 2021, invitation letters were sent by postal mail to the entire gross sample. In addition to the invitation letter and accompanying information for respondents (e.g., study flyer and privacy information), the mailings also included an unconditional prepaid incentive of €5. The mailings were identical in all experimental groups, except that, depending on the mode-sequence

design, a paper-based questionnaire was enclosed and mentioned in the invitation letter. During the field period, up to three reminder letters were sent by mail at 2-week intervals to all invitees who had not participated by that time. The number of reminder letters depended on the different mode-sequence-design strategies in each experimental group (see below). W1R was fielded from April 7, 2021, to June 25, 2021.

The questionnaire in W1R was specifically designed for the purpose of recruiting respondents for the FReDA panel. Thus, it was kept short (duration approximately 10 minutes) and included key sociodemographic questions (e.g., age, gender, family status, nationality), as well as 27 questions deemed interesting and motivating for participants. The W1R questionnaire concluded with a question asking respondents for their consent to be contacted again as part of the FReDA panel.

All 26,725 respondents who gave their panel consent in W1R were invited to participate in the first regular wave of the FReDA panel (W1A). In W1A, respondents were offered a web-based and paper-based questionnaire in a "tailored" mode-sequence design. In brief, "tailored" means that respondents in W1A were assigned to the respective mode-sequence designs depending on which mode they had chosen in W1R (see section 3.1 for a detailed explanation). W1A followed soon after W1R and was fielded between July 7, 2021, and September 14, 2021.

3.2 Experimental Design

Our recruitment experiment comprised three randomized experimental groups and one simulated group, each representing a distinct mode-sequence design. Figure 1 provides an overview of our experimental design. Regardless of the group, all invitation letters and all reminder letters included a web link and QR code to access the web-based questionnaire.

The first experimental group ("concurrent" design) received an invitation letter with an enclosed paper-based questionnaire and a stamped return envelope. Two weeks later, invitees (who had not yet participated in the survey) received a first reminder letter, without a paper-based questionnaire. Four weeks after the initial invitation letter, those invitees who had still not responded received a second reminder letter. This time, the paper-based questionnaire was enclosed again.

In the second and third experimental groups, we implemented two variants of a sequential web-first design. The second experimental group ("web-first-2" design) received an invitation letter and two reminder letters. The paper-based questionnaire was enclosed only with the second reminder letter and was mentioned neither in the invitation letter nor in the first reminder letter.

The third experimental group ("web-first-3" design) received an invitation letter and three reminder letters. Respondents received the third reminder letter

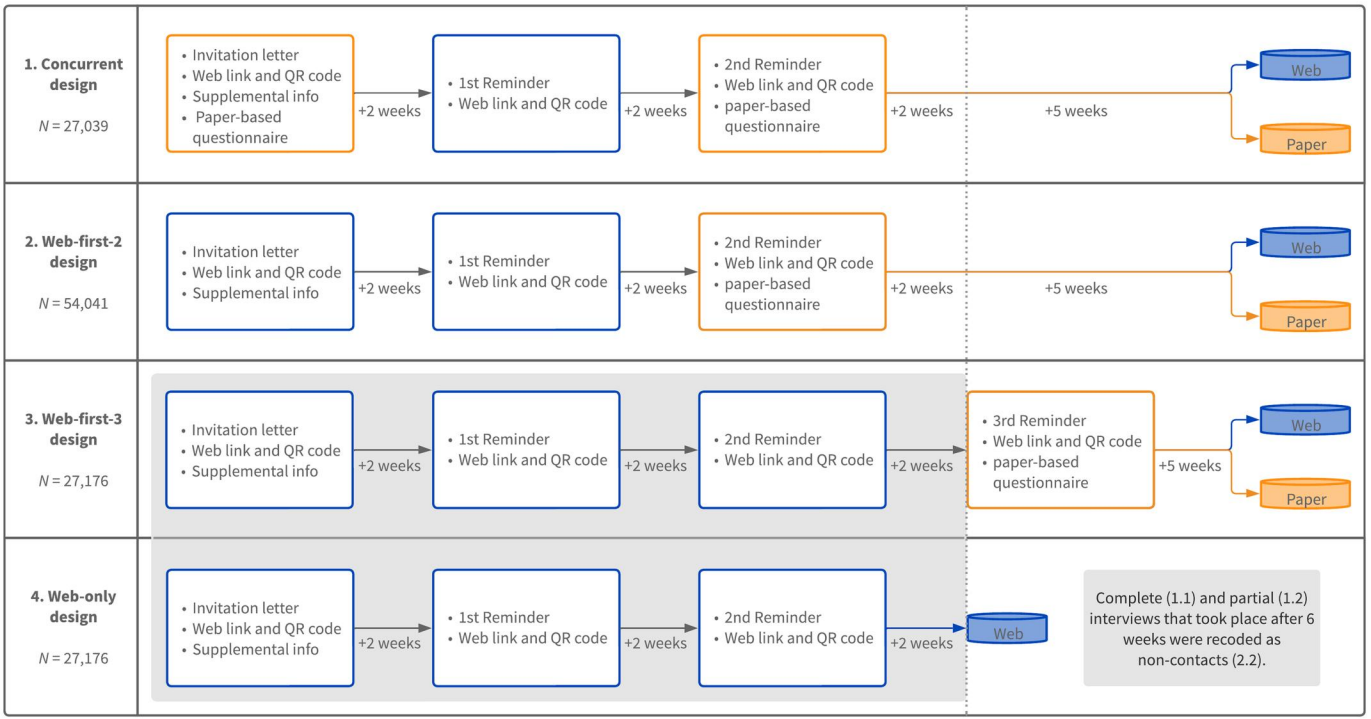

Figure 1. Experiment on Mode-Sequence Designs in the Recruitment Wave (W1R) with Three Experimental Groups (1–3) and a Simulated Group (4).

6 weeks after the initial invitation letter if they had not participated by then. The paper-based questionnaire was enclosed only with the third reminder letter and was not mentioned in the previous letters. Thus, in the web-first-3 design, even more effort was made to push respondents to the web than in the web-first-2 design. However, the web-first-3 group also received a larger number of mailings (i.e., treatments) than the other groups, which limits comparability somewhat.

We simulated a fourth group ("web-only" design), which comprised those respondents from the third experimental group who had received the initial invitation letter and up to two reminder letters. Respondents who received a third reminder letter with an enclosed paper-based questionnaire were excluded from the simulated group and treated as noncontacts. The simulated group therefore allows us to make statements about what the results would have looked like with a purely web-based recruitment approach and a shortened but still realistic field period of 6 weeks. Due to the simulation approach, the respondents in the simulated group are a subset of our third experimental group. This must be considered in the analyses.

Based on experiences reported in a previous experimental study in Germany (Wolf et al. 2021; Gummer et al. 2022), we expected the web-first-2 design to perform best in terms of the trade-off between the share of paper-based questionnaires, response rates, survey costs, and survey errors. Therefore, members of the gross sample were randomly allocated to one of the three experimental groups (concurrent, web-first-2, web-first-3) in a ratio of 1:2:1.

From W1A (i.e., the first regular wave) onward, FReDA implemented a "tailored" mode-sequence design (see figure A1 in the supplementary data online). Respondents who returned a paper-based questionnaire in W1R were assigned to the concurrent group in W1A. Web respondents who were in the concurrent group in W1R were transferred to the web-first-2 group in W1A. All other respondents remained in their initial groups. Because of this tailoring, we compare only the cumulative response rates in W1A and do not compare any other outcome measures.

3.3 Measures

We used a variety of measures to investigate differences among the three experimental groups and the simulated group with respect to key performance indicators, such as survey outcome rates in W1R and W1A and the share of paper-based questionnaires, data quality, sample composition, survey costs, and fieldwork effort in W1R. To determine statistical significance, we performed independent sample *t*-tests for all comparisons, except for comparisons between the web-first-3 group and the simulated web-only group. For these, paired *t*-tests were performed, as mean values of the same individuals were

compared at two points of time (i.e., after 6 weeks and after 11 weeks). When we use the term "all differences significant," we mean that all possible comparisons between the groups were significant at the $p < .01$ level.

We used a set of indicators to measure survey outcomes. First, we calculated the *response rate* in the experimental groups and the simulated group for the recruitment wave W1R (American Association for Public Opinion Research [AAPOR] Response Rate 2; see American Association for Public Opinion Research [AAPOR] 2023). Completed cases include respondents who completed the questionnaire fully or partially (at least 50 percent). Our adaptation of the AAPOR final disposition codes for Germany can be found in table A2 in the supplementary data online. The *panel consent rate* was computed as the number of participants who completed the W1R questionnaire and gave their consent to be surveyed again (panel consent) divided by all completed cases in W1R. The panel consent question was identical in the web and paper modes (for the wording see table A2 in the supplementary data online).

To compute the *recruitment rate*, the number of participants who completed the W1R questionnaire and gave their consent to be surveyed again was divided by the number of all persons who received an invitation letter to participate in W1R. Hence, the recruitment rate represents the total number of cases that gave their panel consent, regardless of whether they answered the questionnaire completely or partially, or broke off the questionnaire (i.e., answered less than 50 percent of essential questions). The formulas for all response rates can be found in table A3 in the supplementary data online. The *cumulative response rate* was computed as the number of respondents who completed the W1A questionnaire divided by the number of respondents who received an invitation letter to participate in W1R. We estimated the cumulative response rate based on respondents' initial experimental group assignment in W1R. Cases coded as "non-contacts" in the web-only group in W1R were also coded as "non-contacts" in W1A.

Second, we computed the *share of paper-based questionnaires* as the proportion of respondents among all completed cases who participated via the paper-based questionnaire in W1R. Third, we calculated the *share of no answers* as the proportion of eligible questions ($N = 66$) left unanswered in W1R ("don't know" responses were possible for 7 questions and were considered valid answers; administrative data and questions relating to the survey itself were excluded from the analysis). There was no forced response in the web-based questionnaire.

Fourth, following prior studies (e.g., Gummer 2019; Gummer and Roßmann 2019) that used dissimilarity indices to assess differences between the composition of the survey sample and reference distributions, we calculated Duncan's *dissimilarity index* (Duncan and Duncan 1955) for a set of key demographic variables: age, gender, education, employment status, household income, marital status, household size, nationality, and number of inhabitants

in the municipality. We used the German Microcensus 2020 to derive reference distributions for the variables of interest. Duncan's dissimilarity index can be interpreted as the percentage of cases that would have to be redistributed to other categories of a variable in order to match the reference distribution. It is calculated as follows:

$$D = \frac{1}{2}\sum_{i=1}^{n} |s_i - r_i|,$$

where n is the number of categories, s_i is the share of category i in the sample of interest (FReDA), and r_i is the share of category i in the reference sample (i.e., German Microcensus 2020). It should be noted that the dissimilarity index is not a directional measure and does not indicate where differences exist between categories. For this analysis, we excluded missing categories for each variable to compare our net data with the reference sample and thus reveal potential bias. Our conclusions are based on a comparison of the magnitude of differences in the dissimilarity index and, wherever necessary, on detailed comparisons presented in tables A4–A12 in the supplementary data online, where we display differences between the respective samples in the experimental groups and the German Microcensus 2020 for all variables of interest. A weighted analysis of sample composition is presented in table A13 in the supplementary data online.

Fifth, to assess differences in distributions of substantive responses, we calculated *mean values* and *variances* (with their respective confidence intervals) across the experimental groups and the simulated group for all substantive variables in the questionnaire that were measured with rating scales ($n = 27$ items; see table A14 in the supplementary data online for the respective question texts). These items cover various satisfaction variables, gender values, intergenerational values, COVID-19 variables, and subjective household income. For comparability, we rescaled all questions to range from 0 to 1. A weighted analysis of the item means is presented in figures A7 and A10 in the supplementary data online.

Sixth, to provide a comparable indication of survey costs, we calculated *fieldwork effort* and *postage costs.* As an indicator of fieldwork effort, we first estimated the total number of letters sent (initial invitation letter and all reminders) for the experimental groups and the simulated group. To compare the fieldwork effort in the different mode-sequence designs, the different gross sample sizes and outcome rates of the groups had to be considered. For this, we calculated the total number of letters sent per completed case and standardized the indicator using the concurrent group as the reference category.

To compute postage costs, the postage rates charged by the German postal service, Deutsche Post, in 2021 were used. These postage rates were multiplied by the number of letters sent (distinguishing between standard and large letters and including the number of returned paper-based questionnaires) and

then added up to provide the overall postage costs. Next, we estimated the postage costs per complete case, and again we standardized the indicator using the concurrent group as the reference category.

Postage costs were chosen as an indicator for describing survey costs for several reasons. First, they are one of the main drivers of costs in self-administered surveys that invite participants via postal mail. Second, they are a good proxy for potential savings, as they are one of the most important factors determining differences in costs between mode-sequence designs. The third reason is comparability. Whereas the actual costs of printing, programming, and data collection vary by survey design and sample size, current postage rates are readily available to everyone and can be applied easily to other survey contexts.

4. RESULTS

4.1 Survey Outcome Rates

Table 1 displays the outcome rates with the respective confidence intervals for W1R and W1A. We started by comparing response rates in W1R between the experimental groups and the simulated group (top row).

We found that the concurrent design had the highest response rate (37.4 percent), the web-first-2 design (33.6 percent) and the web-first-3 design (35.1 percent) took intermediate positions, and the web-only design (30.4 percent) had the lowest response rate (all differences significant at $p < .01$). Although the web-only design achieved the lowest response rate, it was still surprisingly high, considering the simulated—and thus shortened—field period of 6 weeks. In the web-first-3 design, offering a paper-based questionnaire with an additional third reminder letter increased the response rate of the web-only group by 4.7 percentage points to 35.1 percent. The comparison of the web-first-3 with the web-first-2 design, by contrast, shows a difference in response rates of only 1.5 precentage points, which casts doubt on whether sending an additional reminder letter is justified. In figure A2 in the supplementary data online, we illustrate the response rates after a field period of 2, 4, 6, and 11 weeks.

To address our second research question on how mode-sequence designs are related to panel outcomes, we explored panel consent rates and recruitment rates in W1R, as well as cumulative response rates in W1A. Considering the panel consent rates in W1R (row 2 in table 1), the pattern observed for the response rates appears to be reversed, leading to a reduction in group differences. At 74.3 percent, the web-only design achieved the highest panel consent rate compared with all other groups (differences significant at $p < .01$). The web-first-2 design again took an intermediate position (71.0 percent), followed closely by the web-first-3 design with a panel consent rate of 70.3 percent, and

Table 1. Response Rates, Recruitment Rates, and Panel Consent Rates with 95 Percent Confidence Intervals

Rate	Concurrent Mean [95% CI]	Web-first-2 Mean [95% CI]	Web-first-3 Mean [95% CI]	Web-only Mean [95% CI]
AAPOR Response Rate 2 (W1R)	37.40[a,b,c] [36.83; 37.98]	33.60[a,d,e] [33.20; 34.00]	35.10[b,d,f] [34.54; 35.67]	30.36[c,e,f] [29.82; 30.91]
Panel consent rate (W1R)	68.80[a,c] [67.90; 69.71]	71.03[a,e] [70.37; 71.69]	70.26[f] [69.34; 71.18]	74.32[c,e,f] [73.38; 75.27]
Recruitment rate (W1R)	25.84[a,b,c] [25.31; 26.36]	24.04[a,d,e] [23.68; 24.40]	24.83[b,d,f] [24.31; 25.34]	22.71[c,e,f] [22.21; 23.21]
Cumulative AAPOR Response Rate 2 (W1A)	21.20[a,c] [20.71; 21.69]	19.70[a,d] [19.36; 20.03]	20.92[d,f] [20.44; 21.41]	19.60[c,f] [19.12; 20.07]

NOTE.—Statistical significance observed at the $p < .01$ level is based on independent sample *t*-tests for the following comparisons:
[a]Concurrent versus web-first-2.
[b]Concurrent versus web-first-3.
[c]Concurrent versus web-only.
[d]Web-first-2 versus web-first-3.
[e]Web-first-2 versus web-only.
[f]Paired *t*-tests were performed for the comparisons between web-first-3 versus web-only (reported at the $p < .01$ level).

the concurrent design with the lowest panel consent rate of 68.8 percent (differences between these three groups were not statistically significant, except for the web-first-2 design and the concurrent design, which are significant at $p < .01$). This pattern appears to be related to the mode of participation, as the panel consent rate was considerably higher in the web mode than in the paper mode, although the consent question was identical in both modes. Whereas across all groups, 74.0 percent of the respondents who participated via the web-based questionnaire gave their panel consent, only 55.3 percent of respondents who completed the paper-based questionnaire gave their panel consent (difference significant at $p < .01$; see also figure A3 in the supplementary data online).

A look at the recruitment rate in W1R (row 3 in table 1) reveals a similar pattern to the response rate. The concurrent design performed best, with a recruitment rate of 25.8 percent, followed closely by the web-first-3 design with 24.8 percent and the web-first-2 design with 24.0 percent, whereas the web-only design yielded the lowest recruitment rate with 22.7 percent (all differences significant at $p < .01$). Offering a paper-based questionnaire in a fourth contact in the web-first-3 design resulted in an increase in the

recruitment rate of 2.1 percentage points compared with the simulated web-only group.

Finally, if we look at the cumulative response rate in W1A (row 4 in table 1), we find that the initial advantage of the concurrent design has been diminished. The concurrent and web-first-3 designs yielded similar cumulative response rates of 21.2 percent and 20.9 percent, respectively, followed by the web-first-2 (19.7 percent) and the web-only design (19.6 percent). Differences between the concurrent and web-first-3 designs were both significantly different from the web-first-2 and web-only designs at $p < .01$. In the web-first-3 design, the paper-based questionnaire with a third reminder letter resulted in an increase of 1.3 percent in the cumulative response rate compared with the web-only design (difference significant at $p < .01$). Looking at the two mode-sequence designs with the largest initial differences in response rates in W1R, the initial superiority of the concurrent design over the web-only design in the response rate in W1R has shrunk from 7 percentage points to just 1.6 percentage points in the cumulative response rate in W1A. This reduction in the differences between the two mode-sequence designs was due primarily to the significantly higher consent rate in W1R in the web-only design. The convergence of cumulative response rates in W1A between the concurrent design on the one hand and the web-first-2 and web-first-3 designs on the other was similar.

4.2 Share of Paper-Based Questionnaires

We investigated whether and to what extent different mode-sequence-design strategies resulted in different shares of paper-based questionnaires in each group. We found that the concurrent design resulted in 46.7 percent of the participants completing the paper-based questionnaire, whereas the web-first-2 and the web-first-3 designs led to much lower shares of paper-based questionnaires (8.6 and 11.3 percent, respectively; differences significant at $p < .01$; see figure A4 in the supplementary data online). Based on this metric—and apart from a web-only design, where the share of paper-based questionnaires is 0 percent by design—a web-first-2 and a web-first-3 design are clearly preferable to a concurrent design when it comes to increasing the share of respondents who participate via a web-based questionnaire. On the other hand, compared with the web-first-2 design, our web-first-3 design pushed only a marginally larger percentage of respondents to the web-based questionnaire (+2.6 percentage points; difference significant at $p < .01$).

4.3 Data Quality

Overall, the share of no answers in W1R was low, which may also reflect our deliberate effort to provide a short, easy, and interesting questionnaire in

W1R. However, important differences were observed between the groups: the concurrent design yielded the highest share of no answers (2.4 percent), followed by the web-first-2 (1.4 percent) and the web-first-3 designs (1.3 percent). At 0.9 percent, the web-only design achieved the lowest share of no answers (all differences significant at $p < .01$, except in the web-first-2 and web-first-3 designs; see figure A5 in the supplementary data online with the average share of no answers across all substantive variables included in W1R). The share of no answers was generally larger for paper-based questionnaires (4.5 percent) than for web-based questionnaires (0.9 percent; differences significant at $p < .01$; see figure A6 in the supplementary data online). As described above, the concurrent design had the highest proportion of respondents who completed the paper-based questionnaire. Therefore, the higher share of no answers in the concurrent design is likely due to this high proportion of no answers in the paper-based questionnaires. Thus, paper-based questionnaires have a lower data quality as measured by the share of no answers, which also has a negative effect on the data quality of the concurrent design.

4.4 Sample Composition

We assessed the magnitude of differences between the sample composition of the mode-sequence-design groups and reference distributions for a set of key demographic variables (see table 2). For this purpose, we calculated Duncan's dissimilarity index. We found only tiny differences in sample composition between the mode-sequence designs for employment status, marital status, and urbanity. We found small differences for gender, with the concurrent design attracting slightly more female participants, thus moving the proportion farther from the benchmark (see table A4 in the supplementary data online). We also found small differences for age, with the concurrent design showing smaller deviations from the reference distribution for those aged 35–40 and 45–50 years (see table A5 in the supplementary data online). Similarly, we found a small difference in education, with the concurrent design representing International Standard Classification of Education 2011 Levels 3 and 4 slightly better than the other mode-sequence designs (see table A6 in the supplementary data online). Regarding household income, the web-first-3 design performed marginally better compared with the other designs, as deviations from the reference distribution were smaller in high- and low-income categories (see table A8 in the supplementary data online). Regarding household size, one-person households were better represented in the web-first-3 and web-only designs than in the other mode-sequence designs (see table A10 in the supplementary data online). Finally, a small difference appears to exist for nationality between the concurrent design and the other designs. However, a comparison of the eight nationality categories in table A11 in the supplementary data online reveals no clear picture and only minimal fluctuations.

Table 2. Sample Composition of the Experimental Groups and the Simulated Group for Key Demographic Variables Calculated Using Duncan's Dissimilarity Index

Variable	Concurrent	Web-first-2	Web-first-3	Web-only
Gender	0.065	0.058	0.057	0.052
Age group	0.019	0.025	0.023	0.028
Education level	0.161	0.175	0.171	0.178
Employment status	0.052	0.051	0.054	0.054
Household income	0.055	0.050	0.039	0.052
Marital status	0.040	0.042	0.041	0.036
Household size	0.064	0.063	0.057	0.056
Nationality	0.123	0.115	0.111	0.111
Urbanity	0.040	0.036	0.039	0.039

In summary, we conclude that no substantial differences exist between the experimental groups across a range of sociodemographic variables. Moreover, the simulated web-only design performed equally well, which indicates that offering an additional paper-based questionnaire did not improve the sample composition for our relatively young sample (aged 18–49 years). A weighted analysis of the sample composition led to the same conclusions (see table A13 in the supplementary data online).

4.5 Substantive Outcomes

Turning to the substantive outcomes in W1R, we compared mean values and variances between the experimental groups and the simulated group (see figures 2 and 3; detailed values are provided in tables A15 and A16 in the supplementary data online). Regarding means, we found no substantial differences between the groups, even though a range of variables were covered in the analysis (i.e., life satisfaction, relationship satisfaction, satisfaction with employment status, experiences during the COVID-19 pandemic, family values, gender values, and care values). The application of the design weights led to the same results (see figure A7 in the supplementary data online). Similarly, we found no substantial differences in variances between groups. Consequently, the different mode-sequence designs did not affect the means and variances of the variables analyzed.

As a further robustness test, we also investigated differences in mean values between the web and paper modes across the experimental groups and the simulated group (see figure A8 in the supplementary data online). Although there were small differences on a few variables, point estimates were still very close. Hence, researchers would likely come to the same conclusions, regardless of survey mode. Turning to variances in the variables of interest (see figure A9 in the

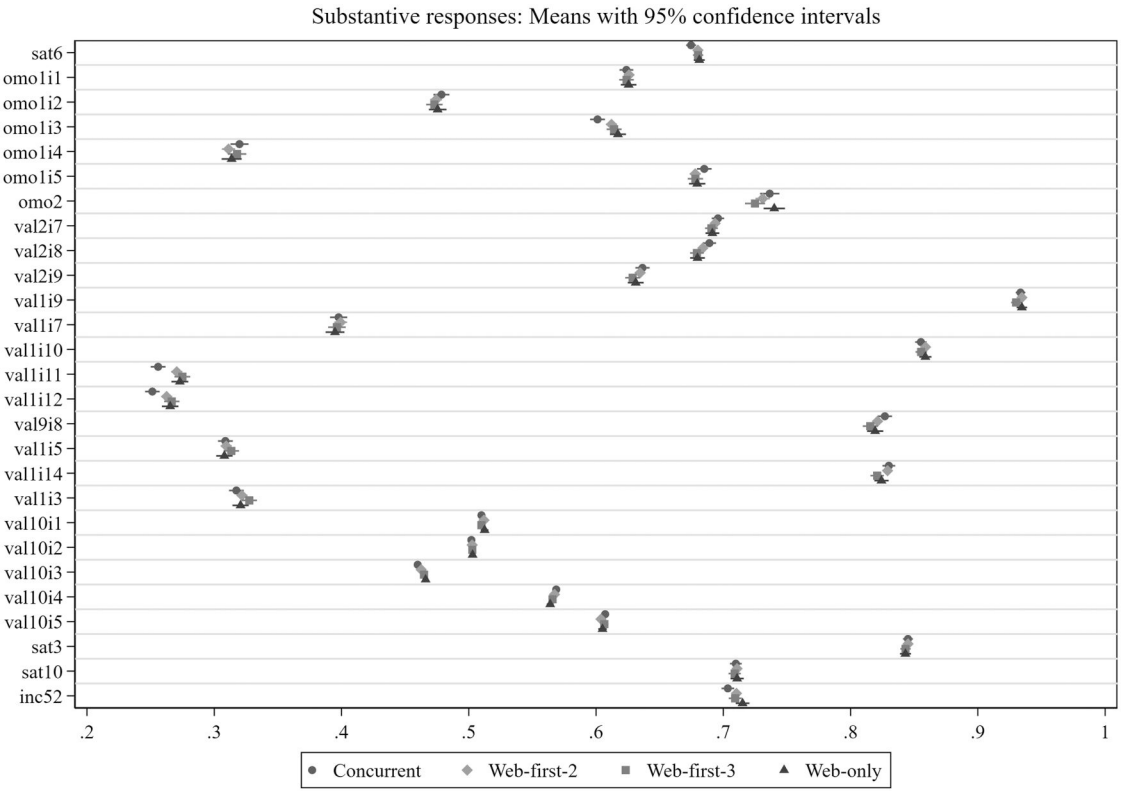

Figure 2. Item Means with 95 percent Confidence Intervals in the Experimental Groups and the Simulated Group.

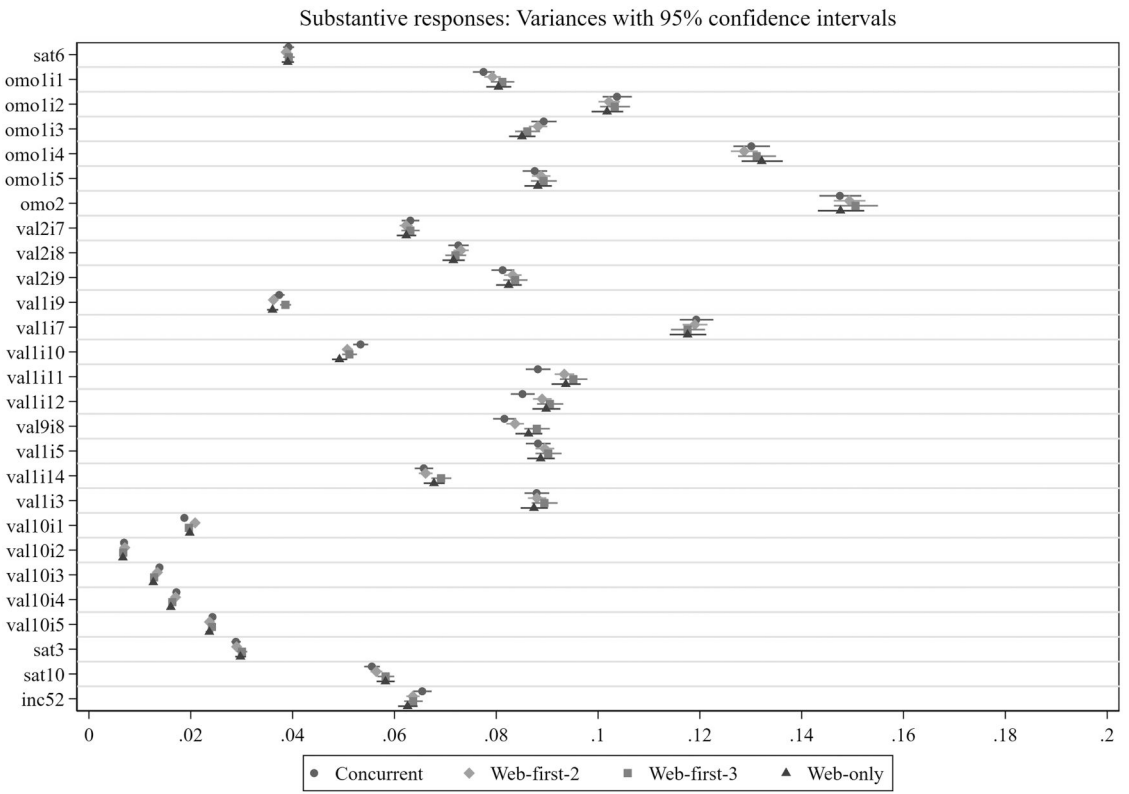

Figure 3. Item Variances with 95 percent Confidence Intervals in the Experimental Groups and the Simulated Group.

supplementary data online), we found slightly more variation in the paper-based questionnaires compared with the web-based questionnaires. However, when considering the direction of means, we found no systematic pattern as to which variables showed differences between the web mode and the paper mode. Here too, the application of design weights led to the same results regarding the item mean values (see figure A10 in the supplementary data online).

4.6 Survey Costs and Fieldwork Effort

The different mode-sequence designs used in FReDA recruitment wave W1R resulted in different survey costs and fieldwork effort. As stated earlier, postage costs were a main cost driver. They varied systematically by group: first, they varied substantially depending on whether a paper-based questionnaire was enclosed. Sending or receiving a paper-based questionnaire required a large letter (€1.55), which is almost twice as expensive in terms of postage as sending a standard letter (€0.80). Second, the number of reminder letters differed between the groups. Because the number of reminder letters affects the number of postal mailings, the fieldwork effort also varied depending on the mode-sequence design.

Looking at the postage costs per completed case in table 3, the concurrent design was the most costly, as a paper-based questionnaire was sent to invitees twice using large letters—with the initial invitation letter and the second reminder letter—depending on participation behavior. Furthermore, the concurrent design also led to the highest number of returned paper-based questionnaires, which further increased postage costs. In the web-first-2 design—where the paper-based questionnaire was sent only with the second reminder, and where we received back substantially less letters with paper-based questionnaires—postage costs per completed case were about 18 percent lower than in the concurrent design.

Interestingly, due to the additional third reminder letter with the paper-based questionnaire, postage costs per completed case in the web-first-3 design were similar to those in the concurrent design (92 percent of the postage costs in the concurrent design). The web-only design was the least expensive—postage costs per completed case were 71 percent of those in the concurrent design. However, the web-only design, and even more so the web-first-3 design, required the highest number of postal mailings per completed case, especially compared with the concurrent design.

5. CONCLUSIONS

With the present study, we set out to investigate how best to recruit a self-administered mixed-mode panel study. In an experiment implemented in the recruitment wave of the German panel study FReDA (W1R), we tested three

Table 3. Calculation of Relative Postage Costs and Relative Fieldwork Effort (Number of Postal Mailings)

	Concurrent (reference)		Web-first-2		Web-first-3		Web-only	
	Standard letter	Large letter	Standard letter	Large letter	Standard letter	Large letter	Standard letter	Large letter
Postage costs[a]	€0.80	€1.55	€0.80	€1.55	€0.80	€1.55	€0.80	€1.55
Invitation letter		27,039	54,041		27,176		27,176	
1st reminder letter	20,947		42,090		21,210		21,210	
2nd reminder letter		18,260		37,852	18,484		18,484	
3rd reminder letter						17,581		
Total number of letters sent		66,246		133,983		84,451		66,870
Received paper questionnaires		4,733		2,053		825		
Number of completed cases		10,075		18,081		9,509		8,010
Total number of letters sent per completed case		6.58		7.41		8.88		8.35
Relative number of letters sent per completed case		1.00		1.13		1.35		1.27
Overall postage costs		€94,307.20		€138,757.55		€82,025.30		€53,496.00
Postage costs per completed case		€9.36		€7.67		€8.63		€6.68
Relative postage costs per completed case		1.00		0.82		0.92		0.71

NOTE.—[a]The postage rates charged by the German postal service, Deutsche Post, in 2021 were used to compute postage costs. Standard letters were used for the initial invitation letter and reminder letters without an enclosed paper-based questionnaire. Large letters were always used when sending or receiving back paper-based questionnaires. The total and relative number of letters sent per completed case do not include letters received back with paper-based questionnaires. Overall postage costs, postage costs per completed case, and relative postage costs per completed case include postage costs for paper-based questionnaires received back.

different mode-sequence designs—concurrent, web-first including a paper-based questionnaire with the second reminder (web-first-2), and web-first including a paper-based questionnaire with the third reminder (web-first-3)—as well as a fourth, simulated web-only design. We examined the effects of different mode-sequence designs in W1R and the first regular wave (W1A). To evaluate the performance of the different mode-sequence designs, we analyzed a comprehensive set of key performance indicators of relevance when setting up panel studies. Concerning sample composition and substantive outcomes, we found only small deviations from the reference distributions across a range of sociodemographic variables, and no differences (in means and variances) in distributions of substantive responses, which can be considered the parameters that matter most to users of panel data.

As expected from previous findings, the web-first-2 design performed well. However, when all key performance indicators were considered, the web-only design proved to be an equally viable recruitment strategy for the FReDA target group (18–49-year olds). Both designs yielded comparatively good response and recruitment rates at reasonable survey costs.

By contrast, the concurrent design was less cost-efficient but resulted in more cases participating in W1R, making it potentially more attractive for cross-sectional surveys when judged by the response rate. However, this substantial initial advantage in terms of participation (especially compared with the web-only design) largely diminished in the context of the FReDA panel study because of the lower panel consent rate in the concurrent design (especially among respondents who used the paper-based questionnaire) and a sharper decline in participation rates in W1A. This puts the concurrent design on par with the web-first-3 design with respect to the cumulative response rate in W1A, while still being ahead of the web-first-2 and web-only designs by a small margin. However, the concurrent design resulted in a huge share of paper-based questionnaires, which we found to be associated with lower data quality in terms of no answers. This significantly limits the recommendation of the concurrent design. Furthermore, based on our findings, we prefer the web-first-2 design over the web-first-3 design, as the small gains in participation in the latter do not justify the substantial increase in survey costs and fieldwork effort.

Our study has some clear messages for survey-based demographic research and the recruitment of participants for a self-administered mixed-mode panel study using web-based and paper-based questionnaires. First, given how well the web-first-2, web-first-3, and web-only designs performed, it seems worthwhile to step up efforts to adapt existing survey programs to web-based surveying and to benefit from the use of web-based questionnaires. In contrast to paper-based questionnaires, web-based questionnaires are more flexible in terms of (i) the inclusion of audio-visual content; (ii) the use of dependent interviewing to reduce response burden and increase data quality; (iii) the implementation of (survey) experiments to facilitate causal analyses; and (iv)

the use of complex filters and branching to tailor the questionnaire to subgroups (Couper and Bosnjak 2010).

Second, our study highlights the fact that designing a survey—especially a complex panel study—involves trade-offs between different aspects deemed relevant by the researchers. Notably, we found that the performance of the mode-sequence designs differed profoundly in terms of survey outcome rates, survey costs, and fieldwork effort.

Third, survey costs remain understudied in survey research (Olson et al. 2021), which is unfortunate as costs are an important constraint when it comes to making design decisions. This lack of information is probably also due to a lack of comparability of cost structures across different settings. The calculation of postage costs provided in the present study can easily be applied to other survey contexts and should serve as a good indication for overall differences between the designs. However, we were not able to neatly distinguish other costs. Nonetheless, it can be safely said that by adopting a web-only design, costs can be further reduced, as there is no need to create a paper-based questionnaire, no need to input the returned questionnaires, and no need to harmonize and integrate data sets from web-based and paper-based data collections.

Our study has several limitations. First, we would caution that our results are probably not generalizable to population-wide surveys, as we targeted only people aged 18–49 years. There is evidence for Germany that when offered a choice between a web-based and a paper-based questionnaire, older age groups are more likely to choose the paper-based option (Wolf et al. 2021). In our view, it would be worthwhile studying how different mode-choice designs can be combined for different (age) groups. A more widespread implementation of the adaptive survey design paradigm (Wagner 2008; Schouten et al. 2018) would be an important step forward.

Second, the fact that our findings are based only on German data might impair their generalizability to other contexts. We would welcome replication studies in other countries to gather insights into structural and contextual differences (e.g., internet penetration rate or smartphone coverage rates) that facilitate the use of specific mode-sequence designs.

Third, we compared only a limited number of mode-sequence designs, and the employed designs differed both in terms of fieldwork duration and the number of reminder letters sent. On the one hand, the simulated web-only group was evaluated based on a shortened field period of 6 weeks, so slightly higher response and recruitment rates can be expected with a longer fieldwork duration. On the other hand, the web-first-3 group received a larger number of mailings than the other groups. Thus, increasing the number of reminder letters sent in the concurrent and the web-only designs to three might increase response and recruitment rates. Ideally, our experiment would have also included a paper-only group, but we dismissed this option as too expensive. Due to the possibility of respondents self-selecting into the different modes,

mode effects cannot be clearly distinguished from composition effects. For this reason, we compared only the mail mode with the web mode in the supporting analysis.

Fourth, we focused on the recruitment process for a family demography panel study and also evaluated participation in the first regular wave after recruitment. Panel studies are designed for long-term data collection, and it seems worthwhile to conduct further analyses that cover longer periods of time. Because FReDA has adopted a "tailored" design—thereby reacting to respondents' mode preferences and altering offered mode choices accordingly—an experimental comparison across multiple waves was not possible.

Supplementary Materials

Supplementary materials are available online at academic.oup.com/jssam.

REFERENCES

American Association for Public Opinion Research (AAPOR) (2023), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (10th ed.), Alexandria, VA: AAPOR. Available at https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf.

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, 391–408.

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, 103–115.

Brick, J. M., and Williams, D. (2013), "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys," *Annals of the American Academy of Political and Social Science*, 645, 36–59.

Burton, J., Lynn, P., and Benzeval, M. (2020), "How Understanding Society: The UK Household Longitudinal Study Adapted to the COVID-19 Pandemic," *Survey Research Methods*, 14, 235–239.

Bujard, M., Gummer, T., Hank, K., Neyer, F. J., Pollak, R., Schneider, N. F., Spieß, C. K., Wolf, C., Bauer, I., Börlin, S., Bretschi, D., Brüggemann, K., Christmann, P., Frembs, L., Groß, K., Jost, C., Kunz, T., Lenke, R., Lines, E., Lück, D., Naderi, R., Naumann, E., Nutz, T., Oehrlein, A.-S., Oellers, V., Ruckdeschel, K., Schmid, L., Schumann, A., Schumann, N., Stein, A., Thönnissen, C., Ullrich, E., von den Driesch, E., and Weih, U. (2022), *FReDA—The German Family Demography Panel Study* (Study No. ZA7777; Data File Version 1.0.0). GESIS–Leibniz Institute for the Social Sciences. https://doi.org/10.4232/1.13745.

Čehovin, G., Bosnjak, M., and Lozar Manfreda, K. (2023), "Item Nonresponse in Web Versus Other Survey Modes: A Systematic Review and Meta-Analysis," *Social Science Computer Review*, 41, 926–945.

Cernat, A., Couper, M. P., and Ofstedal, M. B. (2016), "Estimation of Mode Effects in the Health and Retirement Study Using Measurement Models," *Journal of Survey Statistics and Methodology*, 4, 501–524.

Cernat, A., and Revilla, M. (2021), "Moving from Face-to-Face to a Web Panel: Impacts on Measurement Quality," *Journal of Survey Statistics and Methodology*, 9, 745–763.

Cernat, A., Sakshaug, J., Christmann, P., and Gummer, T. (2022), "The Impact of Survey Mode Design and Questionnaire Length on Measurement Quality," *Sociological Methods & Research*. https://doi.org/10.1177/00491241221140139.

Cornesse, C., Felderer, B., Fikel, M., Krieger, U., and Blom, A. G. (2022), "Recruiting a Probability-Based Online Panel via Postal Mail: Experimental Evidence," *Social Science Computer Review*, 40, 1259–1284.

Couper, M. P., and Bosnjak, M. (2010), "Internet Surveys," in *Handbook of Survey Research*, (2nd ed.), eds. P. V. Marsden and J. D. Wright, Bingley, UK: Emerald Group, pp. 527–550.

de Leeuw, E. D. (2005), "To Mix or Not to Mix Data Collection Modes in Surveys," *Journal of Official Statistics*, 21, 233–255.

———. (2018), "Mixed-Mode: Past, Present, and Future," *Survey Research Methods*, 12, 75–89.

de Leeuw, E. D., and de Heer, W. (2002), "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison," in *Survey Nonresponse*, eds. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, New York: John Wiley & Sons, pp. 41–54.

de Leeuw, E. D., Hox, J., and Luiten, A. (2018), "International Nonresponse Trends across Countries and Years: An Analysis of 36 Years of Labour Force Survey Data," *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2018-00008.

Dillman, D. A. (2017), "The Promise and Challenge of Pushing Respondents to the Web in Mixed-Mode Surveys," *Survey Methodology*, 43, 3–30.

Duncan, O. D., and Duncan, B. (1955), "A Methodological Analysis of Segregation Indexes," *American Sociological Review*, 20, 210–217.

Groves, R. M. (2006), "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly*, 70, 646–675.

Groves, R. M., and Peytcheva, E. (2008), "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis," *Public Opinion Quarterly*, 72, 167–189.

Groves, R. M., Fowler, F. J. Jr, Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2009), *Survey Methodology*, Hoboken, NJ: Wiley.

Gummer, T. (2019), "Assessing Trends and Decomposing Change in Nonresponse Bias: The Case of Bias in Cohort Distributions," *Sociological Methods & Research*, 48, 92–115.

Gummer, T., and Roßmann, J. (2019), "The Effects of Propensity Score Weighting on Attrition Biases in Attitudinal, Behavioral, and Socio-Demographic Variables in a Short-Term Web-Based Panel Survey," *International Journal of Social Research Methodology*, 22, 81–95.

Gummer, T., Christmann, P., Verhoeven, S., and Wolf, C. (2022), "Using a Responsive Survey Design to Innovate Self-Administered Mixed-Mode Surveys," *Journal of the Royal Statistical Society A*, 185, 916–932.

Holmberg, A., Lorenc, B., and Werner, P. (2010), "Contact Strategies to Improve Participation via the Web in a Mixed-Mode Mail and Web Survey," *Journal of Official Statistics*, 26, 465–480.

Klausch, T., Hox, J. J., and Schouten, B. (2013), "Measurement Effects of Survey Mode on the Equivalence of Attitudinal Rating Scale Questions," *Sociological Methods & Research*, 42, 227–263.

Klausch, T., Schouten, B., and Hox, J. J. (2017), "Evaluating Bias of Sequential Mixed-Mode Designs against Benchmark Surveys," *Sociological Methods & Research*, 46, 456–489.

Legleye, S., and Charrance, G. (2021), "Sequential and Concurrent Internet-Telephone Mixed-Mode Designs in Sexual Health Behavior Research," *Journal of Survey Statistics and Methodology*, 11, 75–99.

Luijkx, R., Jónsdóttir, G. A., Gummer, T., Ernst Stähli, M., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S., Hjaltason, B. A., Joye, D., Lomazzi, V., Maineri, A. M., Milbert, P., Ochsner, M., Pollien, A., Sapin, M., Solanes, I., Verhoeven, S., and Wolf, C. (2021), "The European Values Study 2017: On the Way to the Future Using Mixed-Modes," *European Sociological Review*, 37, 330–346.

Lynn, P. (2020), "Evaluating Push-to-Web Methodology for Mixed-Mode Surveys using Address-Based Samples," *Survey Research Methods*, 14, 19–30.

Mauz, E., von der Lippe, E., Allen, J., Schilling, R., Müters, S., Hoebel, J., Schmich, P., Wetzstein, M., Kamtsiuris, P., and Lange, C. (2018), "Mixing Modes in a Population-Based Interview Survey: Comparison of a Sequential and a Concurrent Mixed-Mode Design for Public Health Research," *Archives of Public Health*, 76, 8–17.

Olson, K., Smyth, J. D., and Ganshert, A. (2018), "The Effects of Respondent and Question Characteristics on Respondent Answering Behaviors in Telephone Interviews," *Journal of Survey Statistics and Methodology*, 7, 275–308.

Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N. A., McCarthy, J. S., O'Brien, E., Opsomer, J. D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z. T., Turakhia, C., and Wagner, J. (2021), "Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys: AAPOR Task Force Report," *Journal of Survey Statistics and Methodology*, 9, 381–411.

Olson, K., Wagner, J., and Anderson, R. (2021), "Survey Costs: Where Are We and What Is the Way Forward? " *Journal of Survey Statistics and Methodology*, 9, 921–942.

Patrick, M. E., Couper, M. P., Laetz, V. B., Schulenberg, J. E., O'Malley, P. M., Johnston, L. D., and Miech, R. A. (2018), "A Sequential Mixed-Mode Experiment in the U.S. National Monitoring the Future Study," *Journal of Survey Statistics and Methodology*, 6, 72–97.

Piccitto, G., Liefbroer, A. C., and Emery, T. (2022), "Does the Survey Mode Affect the Association between Subjective Well-Being and Its Determinants? An Experimental Comparison between Face-to-Face and Web Mode," *Journal of Happiness Studies*, 23, 3441–3461.

Poushter, J. (2016), *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies*, Washington, DC: Pew Research Center.

Rao, K., Kaminska, O., and McCutcheon, A. L. (2010), "Recruiting Probability Samples for a Multi-Mode Research Panel with Internet and Mail Components," *Public Opinion Quarterly*, 74, 68–84.

Sakshaug, J. W., Beste, J., Coban, M., Fendel, T., Haas, G.-C., Hülle, S., Kosyakova, Y., König, C., Kreuter, F., Küfner, B., Müller, B., Osiander, C., Schwanhäuser, S., Stephan, G., Vallizadeh, E., Volkert, M., Wenzig, C., Westermeier, C., Zabel, C., and Zins, S. (2020), "Impacts of the COVID-19 Pandemic on Labor Market Surveys at the German Institute for Employment Research," *Survey Research Methods*, 14, 229–233.

Sakshaug, J. W., Cernat, A., and Raghunathan, T. E. (2019), "Do Sequential Mixed-Mode Surveys Decrease Nonresponse Bias, Measurement Error Bias, and Total Bias? An Experimental Study," *Journal of Survey Statistics and Methodology*, 7, 545–571.

Schneider, N. F., Bujard, M., Wolf, C., Gummer, T., Hank, K., and Neyer, F. J. (2021), Family Research and Demographic Analysis (FReDA): Evolution, Framework, Objectives, and Design of "The German Family Demography Panel Study," *Comparative Population Studies*, 46, 149–186.

Schouten, B., Cobben, F., and Bethlehem, J. (2009), "Indicators for the Representativeness of Survey Response," *Survey Methodology*, 35, 101–113.

Schouten, B., Peytchev, A., and Wagner, J. (2018), *Adaptive Survey Design*, Boca Raton, FL: CRC Press.

Seidenberg, A. B., Moser, R. P., and West, B. T. (2023), "Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)," *Journal of Survey Statistics and Methodology*, 11, 743–757.

Silber, H., Roßmann, J., Gummer, T., Zins, S., and Weyandt, K. W. (2021), "The Effects of Question, Respondent, and Interviewer Characteristics on Two Types of Item Nonresponse," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184, 1052–1069.

Taylor, K., and Silver, L. (2019), *Smartphone Ownership Is Growing Rapidly around the World, but Not Always Equally*, Washington, DC: Pew Research Center.

Wagner, J. (2008), "Adaptive Survey Design to Reduce Nonresponse Bias," doctoral dissertation, University of Michigan.

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., and Verhoeven, S. (2021), "Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys," *Public Opinion Quarterly*, 85, 623–648.