Received: 7 October 2020 | Accepted: 16 December 2021

DOI: 10.1111/rssa.12809

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY | Series A Statistics in Society

Introducing Web in a mixed-mode establishment survey: Effects on nonresponse

Patrick Gleiser[1] | **Joseph W. Sakshaug**[1,2,3] | **Marieke Volkert**[1] | **Peter Ellguth**[1] | **Susanne Kohaut**[1] | **Iris Möller**[1]

[1]Institute for Employment Research (IAB) of the German Federal Employment Agency (BA), Nuremberg, Germany

[2]University of Mannheim, Mannheim, Germany

[3]Ludwig Maximilian University of Munich, Munich, Germany

Correspondence
Patrick Gleiser, Institute for Employment Research (IAB) of the German Federal Employment Agency (BA), Regensburger Strasse 104, D-90478 Nuremberg, Germany.
Email: patrick.gleiser@iab.de

Abstract

This study reports on a mode design experiment in which a Web starting mode was introduced for the first time in the Institute for Employment Research Establishment Panel. A cross-sectional sample of establishments was randomized to be interviewed via the traditional face-to-face procedure or a Web-first sequential mixed-mode design with face-to-face follow-ups. Extensive administrative data were used to estimate and compare nonresponse bias at multiple phases of the sequential mixed-mode design, and assess the relationship between mode design and establishment characteristics on the likelihood of response. We show that the final response rates and nonresponse bias were similar between both mode designs, but these results contrasted with the results at each phase of the sequential mode design. Larger establishments were significantly more likely to respond in the Web mode compared to the face-to-face mode. A moderate cost savings (of about 14% per respondent) was estimated for the Web-first sequential mode design.

KEYWORDS

administrative data, business survey, IAB Establishment Panel, nonresponse, Web survey

1 | INTRODUCTION

Collecting data about establishments is a crucial component of producing economic statistics. Such data are used to compile key figures on employment trends, production rates and revenues in different industries, which are used for policy-making in political and economic fields. As such, many establishments receive a large number of survey requests, including voluntary and mandatory requests. Not surprisingly, voluntary surveys typically have lower response rates than mandatory ones (Petroni et al., 2004). Moreover, large establishments are especially less likely to participate compared to smaller ones (Earp et al., 2018; Hartmann & Kohaut, 2000; Hecht et al., 2019; Janik & Kohaut, 2012; Phipps & Toth, 2012; König et al. 2021). This serves as a major threat to business data as large establishments have a disproportionate influence on survey estimates. Thus, nonresponse is a significant concern for the quality of economic data (Bavdaž et al., 2020).

One approach that has been proposed to counteract low response rates is to implement multiple modes of data collection. Mixed-mode designs have been shown to attract different types of respondents who have different propensities for responding in a particular mode (de Leeuw, 2005, 2018). Furthermore, survey organizations may implement multiple modes sequentially to save costs by reserving the most-expensive (usually interviewer-administered) modes, which have higher motivational power, for the most reluctant respondents who did not respond to earlier requests in the less-expensive (usually self-administered) modes.

Some of the most common modes used to collect survey data from establishments are face-to-face, telephone and paper-based methods (e.g. mail), but there has been a transition towards online data collection and mixing online and offline modes (Dillman et al., 2014). For instance, Statistics Sweden offers a Web option in all of their establishment surveys (Erikson, 2007), while Statistics Netherlands has adopted Web as their primary data collection mode. Introducing a Web mode in an existing establishment survey seems to be a promising option for eliciting response and minimizing respondent burden, as it may reduce internal paperwork, improve data quality and facilitate the response process, which is carried out at the establishment's convenience (Millar et al., 2018; Snijkers, 2008; Snijkers & Jones, 2013). Moreover, introducing a Web mode may reduce the costs of data collection (Dillman et al., 2014).

One survey that has recently transitioned from primarily face-to-face data collection to a mix of online and offline data collection is the Establishment Panel of the Institute for Employment Research (IAB) in Germany. Since 1993, the IAB Establishment Panel collects representative data from a panel of establishments each year. In addition, a cross-sectional sample of new establishments is recruited annually to adjust for structural changes in the economy. While the re-interview rate for establishments has remained steady at around 80%, the response rate for the annual cross-sectional samples has declined from around 62% in 1993 to only 20% in 2018 (Kantar, 2020). This trend is consistent with declining response rates worldwide (Dutwin & Lavrakas, 2016; Kennedy & Hartig, 2019; de Leeuw & de Heer, 2002; de Leeuw et al., 2018; Seiler, 2010), and motivated the introduction of Web in the Panel.

Although adding an online mode to establishment surveys is not necessarily new, there is little experimental evidence examining the impacts of introducing Web in a face-to-face establishment survey on response rates and nonresponse bias. Most establishment survey research that specifically addresses mode effects focuses on Web take-up rates or response rate differences between competing modes or mode designs without assessing nonresponse bias or the relationship between mode (design) and establishment characteristics (e.g. size, industry sector) on survey participation (Bremner, 2011; Erikson, 2007; Millar et al., 2018; Snijkers

& Jones, 2013). What is missing from this literature are assessments of response rates and nonresponse bias at each phase of a sequential mixed-mode design. Such assessments are useful for directly comparing the effects of individual starting modes (e.g. Web vs. face-to-face) and follow-up modes, as well as for answering questions about specific establishment subgroups. For example, are large establishments more (or less) likely to respond via Web than face-to-face?

The present study addresses these research gaps by reporting the results of a mode design experiment conducted in the 2018 IAB Establishment Panel, where a Web starting mode was introduced in the Panel for the first time. A new cross-sectional cohort of establishments was randomly allocated over two groups. The first group (control group) was assigned to the traditional procedure: face-to-face (FTF) interviewing with paper questionnaires and a nonresponse follow-up conducted via paper self-administered questionnaire (SAQ) that the interviewer left behind with the establishment for self-completion (Ellguth et al., 2014). The second group (experimental group) was assigned to a Web starting mode with nonresponse follow-ups conducted using the sequential FTF and SAQ modes. A unique strength of the present study is the use of detailed administrative data to estimate and evaluate nonresponse bias and assess interactions between the different mode designs and establishment characteristics on the likelihood of response. These analyses are performed separately for multiple phases of the mixed-mode designs to assess the cumulative effects of the starting modes and nonresponse follow-up modes. The results of these analyses will provide useful insights into the effects of introducing a Web starting mode on nonresponse in a traditionally face-to-face establishment survey.

Altogether, this study addresses the following research questions:

1. What are the effects of introducing a Web-first mixed-mode design on response rates in an establishment survey, relative to a FTF-first design?
2. Are there differences in nonresponse bias between the Web-first and FTF-first mixed-mode designs?
3. Are certain types of establishments (e.g. large establishments) more (or less) likely to respond to a Web-first or a FTF-first mixed-mode survey?

The remainder of the article is composed into five sections. Section 2 reviews different types of mixed-mode survey designs and summarizes the existing literature on Web mode effects and establishment survey participation. Section 3 describes the experimental design and data sources. Section 4 presents the evaluation methodology. Section 5 reports the results of the experiment, including cost analysis, and Section 6 summarizes the findings and discusses their implications for survey practice.

2 | BACKGROUND

2.1 | Mixed-mode survey designs

Two of the most common mixed-mode designs used in establishment surveys are concurrent and sequential (Snijkers & Jones, 2013). The concurrent approach offers multiple modes at the same time, allowing sampled units the opportunity to explicitly choose between them. With the sequential approach, there is no explicit mode choice and the data collection process is usually split into phases. In the first phase, sampled units are offered only one mode, the primary

or starting mode. First-phase nonrespondents are then followed up in the second phase with a different mode, the secondary mode. Nonrespondents in the second phase may be followed up using a tertiary mode and so on. The goal of this approach is to recruit a large share of respondents using the primary mode and reserve the follow-up mode(s) for the most reluctant units.

There are several practical advantages of mixed-mode surveys (relative to single-mode surveys) that have been reported in the literature. First, they can improve coverage by reaching different segments of the target population (de Leeuw, 2005, 2018). For instance, participating in a Web-only survey may not be possible for establishments due to technical limitations (Snijkers, 2008) or internal policies (Ellguth & Kohaut, 2014). Establishments without internet access or those with strict spam filters or firewalls may be unable to access the Web portal. This problem could be overcome, for example, by offering a paper questionnaire option or interviewer follow-up. Second, mixed-mode designs can reduce survey costs by encouraging the use of less-expensive (e.g. self-administered) modes (Ellguth & Kohaut, 2014). The sequential mixed-mode approach is considered to be more effective in this regard, especially when a less-expensive mode (e.g. Web) is implemented in the first phase and more expensive modes (e.g. face-to-face) are deployed only for the nonresponse follow-up phases. If a significant proportion of interviews are completed in the more economical starting mode, then cost savings might be achieved by minimizing the use of the more expensive follow-up modes.

Third, mixing modes can maintain (or even improve) response rates relative to single-mode designs (de Leeuw, 2005, 2018). Since multiple modes are utilized, there is greater opportunity for respondents to participate in one that is potentially less burdensome for them. For instance, some establishments may perceive interviewer-administered modes as cumbersome because they have to be prepared to complete the questionnaire 'on the spot' in one go. In contrast, offering the possibility of self-completion makes it easier for these establishments to participate at their convenience. This is important since the survey questions may require the look-up of records to provide accurate answers, which could be a rather burdensome task when an interviewer is present (Bavdaz et al., 2015; Edwards & Cantor, 1991). On the other hand, some establishments may prefer interviewer modes, especially when the questionnaire contains complex questions that would be difficult to answer without interviewer assistance. For these establishments an interviewer mode might prevent them from getting stuck on a difficult question and potentially breaking-off the survey before answering all of the essential questions.

However, while mixed-mode designs may benefit coverage, response rates and costs, they may also introduce differential measurement errors (or measurement mode effects) (Voogt & Saris, 2005). This occurs if respondents provide different answers depending on the mode they are interviewed in. For instance, respondents might provide different answers in modes with interviewer presence because of social desirability (Kreuter et al., 2009) or presentational effects (e.g. primacy, recency). Although measurement mode effects are an important and understudied topic in the establishment survey literature, we do not address them in this study. Rather our focus is on selection mode effects associated with introducing a Web-first mixed-mode design as an alternative to a traditional FTF-first design. FTF is a commonly used mode in establishment surveys because of its adaptability to the establishment's situation. For example, if establishments are not amenable to a face-to-face interview, then the interviewer can leave behind the paper questionnaire for self-completion (Schmucker et al., 2018).

2.2 | Mode effects on nonresponse in establishment surveys

Previous studies indicate that establishments are initially reluctant to adopt the Web mode, especially in concurrent mixed-mode designs. This is also reflected in overall low Web response rates (Bremner, 2011; Erikson, 2007; Marquette & Kornbau, 2013; Snijkers & Jones, 2013). Erikson (2007) summarizes findings on Web take-up rates for various establishment surveys from Statistics Sweden, reporting initially low Web take-up rates but a steady increase over time. For example, the take-up rate in a monthly short-term employment survey increased from 5% to 19% within 1 year. The authors speculate that establishments tend to familiarize themselves with the Web mode. General trends of increasing Web take-up rates have also been observed in the United States and Norway (Snijkers & Jones, 2013; Thompson et al., 2015).

Ellguth and Kohaut (2014) investigated respondent attitudes towards introducing a Web mode in the primarily face-to-face IAB Establishment Panel. Overall, about 38% of respondents expressed willingness to complete the survey via Web in a future wave of the panel. Large establishments reported greater willingness compared to smaller ones. Furthermore, establishments that refused the face-to-face interview, but agreed to complete the questionnaire via paper SAQ, reported greater willingness towards future Web participation compared to establishments interviewed face-to-face. The study also investigated reasons for non-willingness with the most common being data security concerns and a preference for interviewer assistance.

Bremner (2011) evaluated the results of a mode design experiment conducted within a UK establishment survey. Establishments were randomized into two mode design groups. One group consisted of a concurrent mixed-mode design in which establishments could choose between Web, paper self-completion and telephone data entry. The second group consisted of a single-mode Web-only design. The results showed that the mixed-mode design had a much higher response rate of 89% compared to the single-mode Web group of 59%. However, the sample excluded establishments with more than 300 employees, whom may possess a higher willingness to participate online.

Millar et al. (2018) compared response rates to the first two waves of the US Emergency Medical Services for Children Program's Performance Measures Survey. The first wave (2007–2008) was conducted via mail and the second wave (2010–2011) via a push-to-Web design with a paper questionnaire option offered during the nonresponse follow-up phase. The second wave yielded a significantly higher response rate (81%) compared to the first wave (57%), which the authors partly attributed to the implementation of the Web-push design.

2.3 | Mode design and establishment characteristics

Aside from qualitative studies, there is little experimental evidence about the interaction between mode (design) and establishment characteristics on survey participation. This is an important research gap as establishments differ in many aspects that likely influence their likelihood of participation under different mode designs. One important characteristic is establishment size (i.e. number of employees). Several studies have found that larger establishments are less likely to participate in voluntary surveys than smaller establishments (Earp et al., 2018; Hartmann & Kohaut, 2000; Hecht et al., 2019; Janik & Kohaut, 2012; Phipps & Toth, 2012), which could reflect differences in their organizational processes. Tomoaskovic-Devey et al. (1994) states that in small establishments the authority and capacity to respond is more centralized than in large establishments. Therefore, it is more likely that a single employee is able to complete

the survey, which makes the response process faster and straightforward. Large establishments, on the other hand, have a higher degree of differentiation and specialization of work areas. This is often reflected in the formation of branch offices or subsidiaries. As a result, internal responsibilities and guidelines on handling survey requests may be unclear, and the responding employee may have to coordinate with headquarters or other departments in advance. Additionally, the availability of information about the establishment can be more fragmented, in which case the designated respondent may lack access to the data sources that are necessary to complete the survey and may need to enlist additional personnel from other departments for assistance. All of these factors increase the amount of resources a large establishment may have to invest in order to participate in a survey, leading to a potentially high burden, especially for new survey requests.

Given the potentially higher burden imposed on larger establishments, these establishments may benefit the most from being interviewed in a self-administered mode, such as Web, as opposed to a face-to-face design. As noted, the structure of large establishments likely requires additional time to coordinate the response task between multiple departments and access the relevant data sources to procure the necessary information to answer the survey questions. In addition, these establishments are probably specialized and skilled enough through their previous survey experiences to complete the survey without interviewer assistance. In this situation, interviewer presence is likely to be unnecessary, cumbersome and potentially counterproductive. In contrast, smaller establishments may require more interviewer support due to having less survey experience and therefore be more amenable to a face-to-face interview. Moreover, the amount of staffing and resources needed to procure the required data for an 'on the spot' personal interview could be more manageable for smaller establishments (Hedlin et al., 2005).

3 | MODE EXPERIMENT AND DATA SOURCES

Two data sources are used to evaluate the mode design experiment: the cross-sectional sample of the 2018 IAB Establishment Panel, in which the experiment was conducted, and linked administrative data from the IAB Establishment History Panel, available for the entire sample. The following sections provide more detail regarding both data sources and the experimental design.

3.1 | Survey data: IAB establishment panel

The IAB Establishment Panel is a voluntary annual panel survey of establishments in Germany sponsored by the Institute for Employment Research (IAB) of the Federal Employment Agency (BA). It collects data on a wide range of topics, including general establishment characteristics, employment trends and technological capabilities. The first wave was carried out in 1993 in the former West Germany. Since 1996 it also covers regions of the former East Germany. The frame population comprises all establishments with at least one employee subject to social insurance contributions. The sampling process uses a complex design, leading to an intended overrepresentation of large establishments, small federal states and specific industry sectors (e.g. manufacturing) (Kantar, 2020). All sampled establishments have a unique identification number, which enables them to be linked to administrative data sources of the Federal Employment Agency.

Traditionally the survey is conducted face-to-face with paper questionnaires. In the case of refusal, interviewers offer establishments the possibility of self-completion via the paper questionnaire, which the interviewer leaves behind and picks-up at a later date. All sampled establishments receive an advance letter highlighting the importance of participation along with confidentiality and data protection assurances. It also announces that an interviewer will soon show up at the establishment to attempt an interview. Each year around 16,000 interviews are successfully completed. The sample consists of two parts. First, there is the reinterview sample, which comprises all establishments that participated in at least one of the previous two waves. Then there is a new cross-sectional sample of establishments that are recruited into the panel. The present study is based solely on the new sample of establishments recruited in year 2018 (or the 26th wave) of the IAB Establishment Panel.

3.2 | Mode design experiment

A mode design experiment was conducted in the 2018 cross-sectional sample of the IAB Establishment Panel to evaluate the impacts of introducing a Web starting mode instead of the traditional FTF starting mode. A total of 13,151 establishments were randomly allocated to either the standard FTF-first mode design (control group) or a Web-first mode design (experimental group). The field period started on the 28th of June and lasted until the 30th of November.

Both mode design groups are depicted in Table 1. The control group used the traditional mode design with FTF as the primary mode of data collection. The questionnaire is intended to be completed face-to-face and interview appointments are made, if necessary. However, for establishments that refuse or are unable to participate face-to-face (e.g. due to internal policies or privacy concerns), then interviewers offer establishments the option of self-completion by leaving the paper questionnaire behind. Thus, strictly speaking, the control group is a sequential mixed-mode design with a face-to-face starting mode and nonresponse follow-ups conducted via SAQ. In total, 5235 establishments were assigned to the two-phase (FTF–SAQ) control group.

The experimental group used a sequential mixed-mode design with Web as the primary mode, followed by FTF as the secondary mode for Web nonrespondents, and SAQs used as the final mode for all remaining nonrespondents. All establishments within the experimental group received a letter of invitation including login information to the online questionnaire. Web logins were only possible for the first 33 days of the field period and up to two reminder letters were sent to nonrespondents during this period. The last reminder included an announcement that an interviewer will show up for a personal interview if a Web response was not received. The administration of the FTF and SAQ modes in the experimental group was carried out identically as in the control group. A total of 6190 establishments were assigned to the three-phase (Web–FTF–SAQ) experimental group.

TABLE 1 The mode design experiment

	Primary mode	Secondary mode	Tertiary mode
Control group	FTF	SAQ	–
Experimental group	Web	FTF	SAQ

3.3 | Administrative data: Establishment History Panel (BHP)

The Establishment History Panel (BHP) is an annual, cross-sectional administrative database containing all establishments in Germany with at least one employee liable to social insurance contributions since 1975 (since 1992 for establishments in East Germany) (Schmucker et al., 2018). These data have been used in prior methodological research analysing nonresponse bias in the IAB Establishment Panel and other establishment surveys (Bossler et al., 2018; Janik, 2011; Küfner et al., 2020; Sakshaug & Vicari, 2018).

The BHP includes basic information about establishments (e.g. size, industry sector, aggregate employee statistics). The reference date for each annual cross-section is the 30th of June. All establishments in the BHP have an identification number that enables direct linkage to the full sample of the IAB Establishment Panel. As the 2018 cross-sectional sample of the IAB Establishment Panel was drawn 1 year earlier, in 2017 (Bechmann et al., 2019), the BHP administrative data from the 2017 cross-section are linked to the sample for analysing nonresponse bias. Linkage was unsuccessful for 249 (Control group: 100; Experimental group: 149) establishments because of ID changes. ID changes can occur due to recent changes in ownership or legal form, or because of mergers or splits. These establishments are removed from the nonresponse bias analysis, but are retained in the response rate analysis.

A total of 25 BHP variables are used for the nonresponse bias analysis. These variables are classified into four substantive groups. The first variable group, basic properties, contains characteristics about the establishment's size (1–9, 10–49, 50–249 and 250+ employees), geographical location (East or West Germany), industry sector (Agriculture/Production, Service and Public/Education/Health/Arts; see Table A.1 in the Online Supplementary Materials for the recoding scheme) and year of foundation which is based on the first time the establishment appears in the BHP administrative data (Before or After 2010).

The second variable group, employee structure, describes the demographic distribution of the establishment's employees, including the proportion of female employees (0, >0-0.5, >0.5-<1, 1), average age (in years) based on quartiles (16-37.67, >37.67-42.41, >42.41-46.5, >46.5), proportion of German employees (0-<1, 1), proportions of low- and high-qualified employees (0, >0-1) and the proportion of medium-qualified employees (0, >0-<0.5, 0.5-<1, 1).

The third variable group, employment type, covers the distribution of employee contract types within the establishment. The first variable measures the proportion of regular workers (0-0.5, >0.5-<1, 1). The term 'regular' refers primarily to employees subject to social insurance contributions. Non-regular workers include trainees, employees in partial retirement, short-term employees and casual workers (Schmucker et al., 2018, 52). Then there are two variables measuring the proportions of employees working part-time and full-time (0, >0-0.5, >0.5-<1, 1).

The fourth variable group, occupation type, refers to distributions of occupations held by employees within the establishment. For instance, one variable measures the proportion of employees in simple manual occupations. Occupational types are defined per the Blossfeld classification (Blossfeld, 1987; Schmucker et al., 2018). The Blossfeld classification distinguishes between 12 occupational groups which are based on level and task requirements for the held job. All 12 variables in this group are binary (0, >0–1) and therefore only indicate whether an establishment has any employees who work in the occupational class or not.

All BHP variables are categorized based on inspection of their original distributions and level of skewness to ensure sufficient sample sizes within cells. Tables A.2–A.5 of the Online Supplementary Materials provide information about the frequency distributions for all variables by mode design group. For further details about each variable reference is made to Schmucker et al. (2018).

4 | METHODOLOGY

The analysis of the mode design experiment consists of three parts. First, response rates are compared between the two mode design groups: experimental and control. Then, nonresponse bias is estimated across the 25 BHP variables (47 estimates) in the four variable groups described in Section 3.3 and compared between the two mode design groups. The final part of the analysis applies logistic regression to assess interactions between the different mode designs and establishment characteristics on the likelihood of survey participation.

4.1 | Sequential mode scenarios

To enable a more detailed evaluation of the mode design experiment and assess the cumulative effects of introducing each sequential mode on the response rate, nonresponse bias and the likelihood of survey participation, up to four different scenarios are examined. Three of the scenarios compare relevant subsets of the full mode sequences, beginning with the starting modes and treating respondents in subsequent modes as nonrespondents. Specifically, the first scenario considers only the starting modes from both the experimental and control groups: Web only versus FTF only, where interviews conducted in the subsequent follow-up modes are treated as nonrespondents. This scenario provides a direct comparison of self- and interviewer-administration on nonresponse prior to any mode switch. The second scenario compares Web only versus FTF–SAQ. This comparison provides insights into whether differences between the Web and FTF starting modes are tempered after the SAQ follow-up mode is offered and self-administration is ensured in both mode designs. The third scenario compares Web–FTF versus FTF only. This is the main comparison for assessing whether introducing a Web starting mode into an otherwise FTF-only design affects the outcomes of interest. Lastly, the fourth scenario compares the full mode sequences in both groups: Web–FTF–SAQ versus FTF–SAQ. This comparison, similar to the previous comparison, assesses the overall effect of using a Web starting mode in the traditional FTF–SAQ design of the IAB Establishment Panel. We note that some of these scenarios are simulated as we do not know definitively what would have happened if FTF was actually implemented with no SAQ follow-up. In this case, we can only simulate the results by ignoring the SAQ follow-up data. Tables A.6–A.9 report the frequencies of respondents and nonrespondents in each mode design group for all four scenarios.

Not all comparisons are considered for each of the three analyses (response rates, nonresponse bias and modelling participation). For example, the Web-only comparisons are not considered in the nonresponse bias analysis due to small cell sizes of Web respondents for many BHP variables. The Web-only comparisons are, however, considered in the regression analysis which utilizes a subset of the BHP variables.

4.2 | Response rate definition

To address the first research question, we define a completed interview (or respondent) as having answered two essential questions in the IAB Establishment Panel. The first question relates to the number of employees, broken down by employee groups, and the second question collects detailed information on the employee structure. These questions appear in the beginning and middle parts of the questionnaire, respectively, and refer to general information that could most

likely be answered by any establishment. Partial interviews are all cases that do not meet the criteria for a completed interview but answered at least one question (essential or non-essential).

For non-interviews, a distinction is made between refusals and non-contacts using the final disposition codes (AAPOR, 2016). Refusals comprise all establishments that had been successfully contacted but declined to participate in the survey for any reason. Non-contacts include all establishments for which no contact could be made. Final disposition codes, available from contact history records, are used to categorize each non-interview as refusal or non-contact (see Table A.10). We note that partial interviews were manually coded to the refusal category 'invalid interview' because although there was contact, the interview was not sufficiently completed. Final disposition codes were missing for some cases. These cases were inspected to see whether any interviewer contact was made during the field period using intermediate field codes. If an interviewer contact was made, then these cases were classified as refusals. If there was no evidence of an interviewer contact, then non-contact was assumed.

The response rate is calculated using the Response Rate 1 definition proposed by (AAPOR 2016, 61), which essentially divides the number of completed interviews by the full sample.

$$\text{Response Rate} = \frac{I}{(I+P)+(R+NC)} \tag{1}$$

where I refers to the number of completed interviews, P the number of partial interviews, R the number of refusals and NC the number of non-contacts.

4.3 | Estimating nonresponse bias

To address the second research question, estimates of nonresponse bias are calculated for each BHP variable in each mode design group. The BHP variables are used as a proxy for nonresponse bias, given their likely correlation with the actual survey variables.

Nonresponse bias is calculated based on the estimated proportion of each BHP variable category, $c = (1, 2, \ldots, C)$. Specifically, nonresponse bias is computed as the difference between the estimated proportion of a variable category among the responding cases $\overline{Y}_{c,r}$ and the proportion of the same category in the drawn sample $\overline{Y}_{c,n}$:

$$\text{Nonresponse Bias}\left(\overline{Y}_c\right) = \overline{Y}_{c,r} - \overline{Y}_{c,n}. \tag{2}$$

As the magnitude of the estimated nonresponse bias is difficult to interpret across categories of different variables, we also report the absolute relative nonresponse bias (Groves, 2006), by calculating the ratio of the nonresponse bias of a category c to its proportion in the sample $\overline{Y}_{c,n}$. This measure can be interpreted as the magnitude of nonresponse bias relative to the sample-based estimate:

$$\text{Absolute Relative Nonresponse Bias}\left(\overline{Y}_c\right) = \left|\frac{\overline{Y}_{c,r} - \overline{Y}_{c,n}}{\overline{Y}_{c,n}}\right|. \tag{3}$$

To facilitate the comparison of nonresponse bias between both mode design groups, measures of the average absolute nonresponse bias (AAB) and the average absolute relative nonresponse bias (AARB) are computed for each variable group proposed in Section 3.3 and across all

BHP variables. These aggregate measures are calculated either by taking the sums of the absolute nonresponse bias estimates or the absolute relative nonresponse bias estimates across all variable categories of the respective variable group, and dividing by the total number of categories C in that variable group. For example, the AARB is calculated using the following expression:

$$\text{Average Absolute Relative Nonresponse Bias} = \frac{\sum_{c=1}^{C} \left| \frac{\overline{Y}_{c,r} - \overline{Y}_{c,n}}{\overline{Y}_{c,n}} \right|}{C}. \tag{4}$$

We note that there are other options for using administrative data to assess sample representativeness or the balance of respondent characteristics with respect to the characteristics of the full sample, such as the R-indicator and the coefficient of variation of the response propensities (Moore et al., 2018; Schouten et al., 2009).

4.4 | Modelling survey participation

To investigate the interaction between the competing mode designs and establishment characteristics on the likelihood of survey participation (i.e. the third research question), a regression analysis is performed. Logistic regression is applied to the response indicator, which is a vector of binary indicators $R = (R_1, R_2, \ldots, R_n)'$ with $R_i = 1$ for response and $R_i = 0$ for nonresponse.

A parsimonious model was fitted using a subset of the BHP variables described in Section 3.3. The selection of variables was decided based on a review of the literature, discussions with the IAB Establishment Panel's survey team, and cell size considerations. Additionally, one variable was added to measure the general economic situation of an establishment. This variable measures whether an establishment had an increase, decrease or no change in employment size compared to the previous year.

To assess the influence of mode design, an indicator variable is included $X = (X_1, X_2, \ldots, X_n)'$ with $X_i = 1$ for establishments assigned to the experimental group (i.e. Web–FTF–SAQ) and $X_i = 0$ for establishments assigned to the control group (i.e. FTF–SAQ). The mode design indicator variable is interacted with all establishment characteristics.

The model can be expressed as follows:

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + X_i'\beta + \gamma Z_i' + \eta X_i' Z_i' \tag{5}$$

where p_i is the conditional probability of establishment i to respond, α is the model intercept, β is the coefficient for the mode design indicator variable X, γ is a vector of coefficients for the selected establishment variables Z and η is a vector of coefficients for the interaction effects between the mode design indicator X and the establishment variables Z. The coefficients are calculated using the *glm* function in R (R Core Team, 2017).

All analyses are performed unweighted in order to isolate the impact of mode (design) on unadjusted nonresponse. That is, our main interest lies in randomization rather than representation, and our focus is on assessing the relative effects of the experimental mode (design) manipulations rather than the level of nonresponse bias in the population.

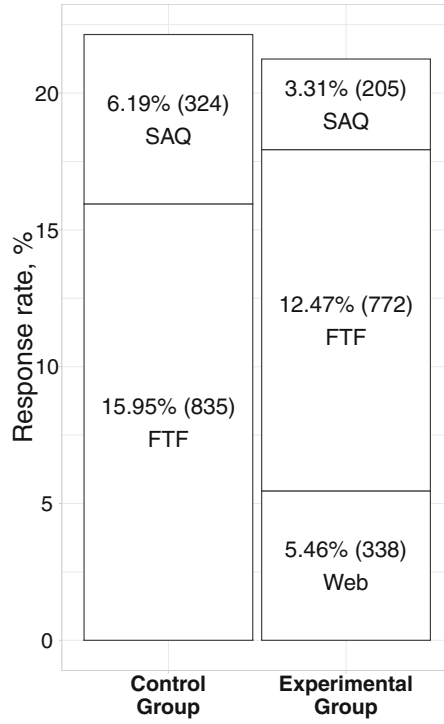

FIGURE 1 Response rate and number of completed interviews (in parentheses) for control and experimental mode design groups

5 | RESULTS

5.1 | Response rates

This section addresses the first research question by comparing response rates between the Web-first experimental group and the FTF-first control group. Out of 11,425 sampled establishments, a total of 2474 completed the survey for an overall response rate of 21.65%. Figure 1 shows the response rates for both mode design groups. There are three key observations that can be made about the mode design groups and the cumulative effects of each deployed mode on response. First, 338 (or 5.46% of) establishments assigned to the experimental group completed the survey in the Web starting mode and thus avoided an interviewer visit. This is substantially lower than the response rate of the FTF starting mode of the control group (15.95%); a statistically significant difference ($p < 0.01$). The lower Web response rate is not particularly surprising given that self-administered modes are known to achieve lower response rates than interviewer-administered modes, and the duration of the Web-only phase was much shorter, about 1 month, compared to the 5-month FTF field period of the control group.

Second, following up the Web nonrespondents with the interviewer-administered FTF mode more than tripled the response rate to 17.93%, which is statistically significantly higher than the FTF-only control group ($p < 0.01$). Lastly, offering the SAQ mode to the remaining nonrespondents in both groups produced similar final response rates: 22.14% in the FTF–SAQ

TABLE 2 Average absolute and average absolute relative nonresponse bias for each variable group and overall, by FTF-only and Web–FTF groups

	FTF only		Web–FTF	
Variable group	Avg. abs. NR bias %	Avg. abs. relative NR bias %	Avg. abs. NR bias %	Avg. abs. relative NR bias %
Basic properties	7.51	25.91	5.11	18.33
Employee structure	6.89	27.75	5.21	21.33
Employment type	6.52	28.81	5.43	23.46
Occupation type	9.56	28.33	7.14	20.17
Overall	7.60	27.79	5.73	20.93

group and 21.24% in the Web–FTF–SAQ group; a statistically insignificant difference ($p = 0.25$). Furthermore, the refusal rate in the Web–FTF–SAQ sequence (63.05%) was similar to that of the FTF–SAQ control group (64.57%), indicating that introducing the Web-first design did not increase the overall rate of refusals. Thus, we may conclude that introducing the Web starting mode did not affect the overall response rate (or refusal rate) of the survey, compared to the traditional FTF-first design, and saved a significant number (approximately 1246) of interviewer visits, as about 26% of all interviews were completed online. Potential cost savings are discussed later in Section 5.4.

5.2 | Nonresponse bias

This section addresses the second research question by reporting the results of the nonresponse bias analysis as described in Section 4.3. As a reminder, the bias analysis is conducted under exclusion of the 249 non-matched cases between the survey and administrative data (see Section 3.3). The matched sample produces nearly identical response rates to those described in the previous section and does not change the results of the response rate comparisons.

The first comparison, between the Web–FTF and FTF-only groups, is shown in Table 2. The table reports the average absolute nonresponse bias (AAB) and the average absolute relative nonresponse bias (AARB) for each of the four substantive BHP variable groups and across all BHP variables. Overall, both aggregate nonresponse bias measures are lower for the Web–FTF group (AAB: 5.73%; AARB: 20.93%) than for the FTF-only group (AAB: 7.60%; AARB: 27.79%). The same pattern holds for each of the four variable groups, indicating that the Web–FTF sequence did a slightly better job of minimizing aggregate nonresponse bias compared to the FTF-only design.

Table 3 reports the aggregate nonresponse bias measures for the full mode sequences in both groups, that is, after the final SAQ mode was offered to all remaining nonrespondents. Offering the SAQ mode reduced the AAB and AARB for both mode design groups, but the bias reduction was more dramatic in the FTF–SAQ group, where self-administration was utilized for the first time. Both mode design groups are comparable with regard to nonresponse bias, with a slightly lower overall aggregate nonresponse bias for the FTF–SAQ group (AAB: 3.77%; AARB: 13.91%) than for the Web–FTF–SAQ group (AAB: 4.27%; AARB: 15.77%). A similar pattern is observed for each variable group.

TABLE 3 Average absolute and average absolute relative nonresponse bias for each variable group and overall, by FTF–SAQ and Web–FTF–SAQ groups

	FTF–SAQ		Web–FTF–SAQ	
Variable group	**Avg. abs. NR bias %**	**Avg. abs. relative NR bias %**	**Avg. abs. NR bias %**	**Avg. abs. relative NR bias %**
Basic properties	3.74	13.71	3.98	14.51
Employee structure	3.69	14.82	3.84	15.77
Employment type	2.99	13.05	3.89	16.89
Occupation type	4.62	13.69	5.36	15.57
Overall	3.77	13.91	4.27	15.77

Individual bias estimates for each of the 47 variable categories are reported in Table A.11 for the Web–FTF versus FTF-only comparison and Table A.12 for the Web–FTF–SAQ versus FTF–SAQ comparison. Both results reveal a high variance on the degree of bias between the variables. Some variables (e.g. industry sector, geographical location) are more or less unaffected by nonresponse bias, while other categories (e.g. establishment size, employee composition of establishment) have large nonresponse biases, often exceeding 20% absolute relative nonresponse bias. The largest nonresponse biases can be seen for categories of the establishment size variable, including establishments with more than 250 employees or between 50-249 employees, which are both largely underrepresented in all mode design groups. Likewise, establishments with between 1-9 employees are severely underrepresented. Out of the four mode design scenarios, the FTF-only group produces the largest bias estimates across all establishment size categories, which suggests that there is a clear benefit to introducing a self-administered mode, either as a starting mode or as a follow-up mode, to reduce nonresponse bias in this variable.

5.3 | Predictors of survey participation

This section addresses the last research question by presenting the regression results for each of the mode design comparison scenarios described in Section 4.1. For each comparison, two models are fitted: a main effects model, which includes the experimental mode design indicator and selected establishment characteristics, and an interaction model with the experimental mode design indicator interacted with each establishment characteristic. The interaction model is used to assess whether the likelihood of participation for certain types of establishments is differentially affected by the mode design.

5.3.1 | Comparison of Web only versus FTF only and FTF–SAQ

The first set of regressions examines the likelihood of participation among establishments in the Web-only experimental group (EG) and the FTF-only control group before any subsequent mode switching occurred. These models provide a direct comparison of the effects of self- and interviewer administration on the likelihood of participation. The fitted model is presented in Table 4. The main-effects model shows that establishments assigned to the initial Web mode were less likely to participate in the survey compared to establishments assigned to the initial FTF mode,

TABLE 4 Logit regression results: Web only versus FTF only and FTF–SAQ (N=11, 176)

Covariates	Web only versus FTF only		Web only versus FTF–SAQ	
	Main effects (β)	Interaction effects (β)	Main effects (β)	Interaction effects (β)
(Intercept)	−1.277**	−1.140**	−0.945**	−0.840**
EG (Ref. Control Group)	−1.203**	−1.799**	−1.610**	−2.098**
Establishment size				
(Ref. 1–9)				
10–49	−0.079	−0.126	−0.046	−0.083
50–249	−0.664**	−0.959**	−0.408**	−0.526**
250+	−1.230**	−1.692**	−0.983**	−1.177**
Location				
(Ref. West)				
East	0.145*	0.219**	0.097	0.134
Year of foundation				
(Ref. 2010 or before)				
2010 or later	−0.022	0.039	−0.088	−0.065
Industry sector				
(Ref. Agriculture/Production)				
Service	−0.199*	−0.161	−0.214*	−0.195
Public/Education/Health/Arts	−0.051	−0.080	0.027	0.026
Prop. high education				
(Ref. [0])				
(0–1]	−0.108	−0.152	−0.115	−0.159
Avg. age				
(Ref. [16–37.67])				
(37.67–42.41]	0.067	0.021	0.056	0.021
(42.41–46.5]	0.128	0.015	0.192*	0.015
(46.5–89]	0.162	0.153	0.185	0.184
Prop. regular worker				
(Ref. [0–0.5])				
(0.5–1)	−0.120	−0.089	−0.022	0.047
[1]	−0.224	−0.261	−0.222*	−0.243
Prop. full time				
(Ref. [0])				
(0–0.5]	0.100	−0.005	0.058	−0.059
(0.5–1)	0.015	−0.151	0.007	−0.141
[1]	0.160	0.107	0.165	0.106

(Continues)

TABLE 4 (Continued)

Interaction effects	Main effects (β)	Interaction effects (β)	Main effects (β)	Interaction effects (β)
Prop. female employees				
(Ref. [0])				
(0–0.5]	0.177	0.180	0.102	0.092
(0.5–1)	−0.010	0.111	−0.066	0.017
[1]	0.088	0.109	0.010	0.001
Economic situation				
(Ref. [no change])				
Employment decreased	0.004	0.088	0.091	0.180
Employment increased	0.006	−0.011	0.037	0.035
Establishment size × EG				
(Ref. 1–9)				
10–49	–	0.237	–	0.195
50–249	–	0.905**	–	0.472
250+	–	1.251**	–	0.736*
Location × EG				
(Ref. West)				
East	–	−0.228	–	−0.143
Year of foundation × EG				
(Ref. 2010 or before)				
2010 or later	–	−0.207	–	−0.103
Industry sector × EG				
(Ref. Agriculture/Production)				
Service	–	−0.076	–	−0.042
Public/Education/Health/Arts	–	0.161	–	0.055
Prop. high education × EG				
(Ref. [0])				
(0–1]	–	0.160	–	0.166
Avg. age × EG				
(Ref. [16–37.67])				
(37.67–42.41]	–	0.141	–	0.141
(42.41–46.5]	–	0.351	–	0.232
(46.5–89]	–	0.052	–	0.022
Prop. regular worker × EG				
(Ref. [0–0.5])				

(Continues)

TABLE 4 (Continued)

Interaction effects	Main effects (β)	Interaction effects (β)	Main effects (β)	Interaction effects (β)
(0.5–1)	–	−0.159	–	−0.295
[1]	–	0.067	–	0.049
Prop. full time × EG				
(Ref. [0])				
(0–0.5]	–	0.476	–	0.530*
(0.5–1)	–	0.816*	–	0.627*
[1]	–	0.637*	–	0.312
Prop. female employees × EG				
(Ref. [0])				
(0–0.5]	–	−0.102	–	−0.013
(0.5–1)	–	−0.485	–	−0.391
[1]	–	−0.097	–	0.012
Economic situation × EG				
(Ref. [no change])				
Employment decreased	–	−0.206	–	−0.297
Employment increased	–	0.068	–	0.021
AIC:	6920.5	6869.8	7905.3	7905.4
Pseudo R-squared:	0.098	0.114	0.135	0.141
Likelihood-ratio test (χ^2):	541.79	634.43	845.87	887.75
Likelihood-ratio test (p-value):	0.000	0.000	0.000	0.000

Signif. level: 0.01**; 0.05*

which is consistent with the significantly lower response rate observed for the Web-only group (see Section 5.1). Other significant main effects include establishment size (−), East Germany (+), and service industry sector (−). The interactions model in Table 4 reveals that establishments with 50 or more employees in the Web-only group are more likely to participate compared to those in the FTF-only group. This suggests that larger establishments were generally more amenable to self-administration than interviewer administration. The only other statistically significant interaction is the proportion of full-time workers: establishments with a workforce of more than 50% full-time workers were more likely to respond via the Web mode than the FTF mode.

Next, we assess the interaction effects for the Web-only design compared to the FTF–SAQ design (Table 4). Even though the FTF nonrespondents were offered the chance to complete a self-administered interview via the SAQ mode, the results still show that larger establishments, particularly those with more than 250 employees, participated at a higher rate in the Web-only design compared to the FTF–SAQ design. The effect size of the interaction term is, however, smaller than that observed in the Web-only versus FTF-only comparison (Table 4), indicating a benefit of offering the SAQ mode in recruiting large establishments. Thus, we may conclude

more generally that both self-administered modes were beneficial in terms of recruiting large establishments. Offering the SAQ mode to FTF nonrespondents also tempers the previously observed positive interaction effect of the Web-only design and establishments employing more than 50% full-time workers, and slightly increases the positive interaction between the Web-only design and establishments with up to 50% full-time workers (excluding zero). All other mode design interactions are statistically insignificant.

5.3.2 | Comparison of Web–FTF versus FTF only

In this section, we assess the participation effects of the sequential Web–FTF design compared to the FTF-only design. The regression results are reported in Table 5. Starting with the main-effects analysis, we find that establishments assigned to the Web–FTF experimental group were more likely to participate compared to the FTF-only control group, which corresponds to the higher response rate previously reported for the Web–FTF group (see Section 5.1). Other statistically significant main effects include: establishment size (−), service industry sector (−), employing all regular workers (−), located in East Germany (+), employing an older workforce (average 46.5+ years) (+), all full-time employees (+) and up to 50% female employees (excluding zero) (+). The interaction model shows that all types of establishments had a similar likelihood of participating in either the Web–FTF or FTF-only mode designs. That is, all interaction terms between the mode design and establishment characteristics are statistically insignificant.

5.3.3 | Comparison of Web–FTF–SAQ versus FTF–SAQ

The final set of regression models compares the effects of the full mode sequences, Web–FTF–SAQ and FTF–SAQ, on the likelihood of participation. The results, reported in Table 5, show that there is no longer a statistically significant main effect of mode design—both mode sequences performed similarly well in recruiting different types of establishments. Again, this is in line with the response rate comparisons from Section 5.1. It is also evident from the interactions model that the different mode designs did not differentially influence the likelihood of participation for establishment subgroups. That is, none of the interactions between mode design and establishment characteristics are statistically significant. Therefore, we conclude that the likelihood of participating in the IAB Establishment Panel was independent of introducing the Web starting mode.

5.4 | Cost analysis

As noted earlier, a key motivation for introducing a Web starting mode in an otherwise face-to-face survey is to reduce costs. Here, we provide an indication of the potential cost savings of introducing Web in the IAB Establishment Panel by examining the cost differential between the two mode designs. We can only provide an indication of savings based on assumed costs as exact cost data are unavailable and subject to contractual arrangements. Furthermore, we cannot account for costs associated with setting up the Web survey or post-processing of the data. The expenses of the Web starting mode phase included printing and mailing up to two postal reminders (0.85 EUR each).

TABLE 5 Logit regression results: Web–FTF versus FTF only and Web–FTF–SAQ versus FTF–SAQ ($N = 11176$)

	Web–FTF versus FTF only		Web–FTF–SAQ versus FTF–SAQ	
Covariates	**Main effects (β)**	**Interaction effects (β)**	**Main effects (β)**	**Interaction effects (β)**
(Intercept)	−1.145**	−1.140**	−0.891**	−0.840**
EG (Ref. Control Group)	0.150**	0.135	−0.051	−0.142
Establishment size				
(Ref. 1–9)				
10–49	−0.246**	−0.126	−0.167*	−0.083
50–249	−0.920**	−0.959**	−0.655**	−0.526**
250+	−1.569**	−1.692**	−1.183**	−1.177**
Location				
(Ref. West)				
East	0.104	0.219**	0.066	0.134
Year of foundation				
(Ref. 2010 or before)				
2010 or later	0.012	0.039	−0.057	−0.065
Industry sector				
(Ref. Agriculture/Production)				
Service	−0.161*	−0.161	−0.195**	−0.195
Public/Education/Health/Arts	−0.016	−0.080	0.041	0.026
Prop. high education				
(Ref. [0])				
(0–1]	−0.118	−0.152	−0.139*	−0.159
Avg. age				
(Ref. [16–37.67])				
(37.67–42.41]	0.044	0.021	0.044	0.021
(42.41–46.5]	0.111	0.015	0.162*	0.135
(46.5–89]	0.141*	0.153	0.161*	0.184
Prop. regular worker				
(Ref. [0–0.5])				
(0.5–1)	−0.109	−0.089	−0.026	0.047
[1]	−0.248**	−0.261	−0.213*	−0.243
Prop. full time				

(Continues)

TABLE 5 (Continued)

Covariates	Web–FTF versus FTF only		Web–FTF–SAQ versus FTF–SAQ	
	Main effects (β)	Interaction effects (β)	Main effects (β)	Interaction effects (β)
(Ref. [0])				
(0–0.5]	0.078	−0.005	0.042	−0.059
(0.5–1)	−0.047	−0.151	−0.007	0.141
[1]	0.186**	0.107	0.217 *	0.106
Prop. female employees				
(Ref. [0])				
(0–0.5]	0.254**	0.180	0.233 *	0.092
(0.5–1)	0.120	0.111	0.135	0.017
[1]	0.108	0.109	0.116	0.001
Economic situation				
(Ref. [no change])				
Employment decreased	−0.079	0.088	0.033	0.180
Employment increased	−0.064	−0.011	−0.008	0.035
Establishment size × EG				
(Ref. 1–9)				
10–49	–	−0.209	–	−0.159
50–249	–	0.068	–	−0.242
250+	–	0.194	–	−0.018
Location × EG				
(Ref. West)				
East	–	−0.207	–	−0.133
Year of foundation × EG				
(Ref. 2010 or before)				
2010 or later	–	−0.047	–	0.017
Industry sector × EG				
(Ref. Agriculture/Production)				
Service	–	−0.005	–	−0.009
Public/Education/Health/Arts	–	0.108	–	0.020
Prop. high education × EG				

(Continues)

TABLE 5 (Continued)

Interaction effects	Main effects (β)	Interaction effects (β)	Main effects (β)	Interaction effects (β)
(Ref. [0])				
(0–1]	–	0.060	–	0.029
Avg. age × EG				
(Ref. [16–37.67])				
(37.67–42.41]	–	0.038	–	0.048
(42.41–46.5]	–	0.168	–	0.055
(46.5–89]	–	−0.024	–	−0.045
Prop. regular worker × EG				
(Ref. [0–0.5])				
(0.5–1)	–	−0.043	–	−0.133
[1]	–	0.025	–	0.067
Prop. full time × EG				
(Ref. [0])				
(0–0.5]	–	0.159	–	0.195
(0.5–1)	–	0.184	–	0.238
[1]	–	0.146	–	0.192
Prop. female employees × EG				
(Ref. [0])				
(0–0.5]	–	0.128	–	0.267
(0.5–1)	–	0.017	–	0.221
[1]	–	−0.002	–	0.209
Economic situation × EG				
(Ref. [no change])				
Employment decreased	–	−0.286	–	−0.394
Employment increased	–	−0.089	–	−0.074
AIC:	9757.5	9780.4	11412	11435
Pseudo R-squared:	0.071	0.074	0.044	0.046
Likelihood-ratio test (χ^2):	485.06	504.14	320.36	337.41
Likelihood-ratio test (p-value):	0.000	0.000	0.000	0.000

Signif. level: 0.01**; 0.05*

The FTF/SAQ expenses included interviewer travel (14 EUR per personal visit) and interviewer payment for a completed interview, which varied depending on the size of the establishment.

Based on our hypothetical cost data, the estimated total cost of the traditional FTF–SAQ design was about 2-percentage points higher than that of the Web–FTF–SAQ design. Although the total costs were very similar between both mode designs, the cost per respondent in the Web–FTF–SAQ design was about 14-percentage points lower than the FTF–SAQ design. Thus, we can conclude that introducing the Web starting mode resulted in a moderate potential cost savings per respondent based on our assumed variable costs.

6 | DISCUSSION

This study experimentally investigated the effects of introducing a Web starting mode on nonresponse in a large-scale, primarily face-to-face voluntary establishment survey in Germany—the IAB Establishment Panel. A cross-sectional cohort of new establishments was randomly assigned to either of two sequential mixed-mode design groups: a Web–FTF–SAQ experimental group or the standard FTF–SAQ (control) group. The mode design experiment was evaluated on three primary outcomes: response rates, nonresponse bias and the interaction of mode design and establishment characteristics on the likelihood of response. These outcomes were evaluated at multiple phases of the mixed-mode sequences in order to compare the cumulative effects of deploying each subsequent mode. The use of an experimental mode design, extensive administrative data to estimate nonresponse bias, and interaction models to study cumulative mode effects on survey participation are rare in the establishment survey literature.

The response rate analysis revealed that about 5.5% of establishments in the experimental Web-first design completed the survey online, which equated to about 26% of all respondents in the full Web–FTF–SAQ sequence. This was significantly lower than the response rate of the FTF-only phase of the control group (16%). However, following up the Web nonrespondents with the FTF mode increased the response rate threefold to about 17.9%, eclipsing that of the FTF-only control group. Following up the remaining nonrespondents with the SAQ mode in both mode design groups produced similar final response rates (Web–FTF–SAQ: 21.2%; FTF–SAQ: 22.1%) and refusal rates, implying no effect of the Web starting mode on participation rates. Based on hypothetical cost data, we estimated a savings of about 1246 interviewer visits and a cost savings of around 14% per respondent in the Web-first design.

The nonresponse bias analysis showed some notable differences between the two mode design groups. Specifically, the Web–FTF sequence produced lower levels of aggregate nonresponse bias compared to the FTF-only phase of the control group. Offering the SAQ follow-up mode in both mode design groups reduced this discrepancy, resulting in a mostly comparable but slightly lower aggregate bias for the FTF–SAQ group relative to the Web–FTF–SAQ group. The largest nonresponse biases were observed for establishment size: large establishments (250+ employees) were substantially underrepresented in all mode design scenarios, but especially in the FTF-only scenario before the SAQ follow-up mode was offered. This finding underscores the importance of including a self-administered mode when recruiting large establishments in a mixed-mode design.

The regression analysis confirmed that larger establishments were more likely to respond via Web than face-to-face, despite the much shorter Web-only phase of the field period. Moreover, the likelihood of response among large establishments remained higher in the Web-only phase than in the full FTF–SAQ sequence of the control group. Thus, the Web mode seems to be more attractive to large establishments than a sequence of interviewer administration followed by

paper-based self-administration. No significant interactions between mode design and establishment characteristics were observed for either of the Web–FTF versus FTF only or Web–FTF–SAQ versus FTF–SAQ comparisons, signifying that the Web starting mode had no influence on the likelihood of participation across establishment subgroups.

Taken together, this study did not show any serious drawbacks of introducing a Web starting mode on response rates and nonresponse bias in the IAB Establishment Panel. Rather it highlighted the benefits of using a Web mode for the recruitment of large establishments and the consequences of relying solely on an interviewer-administered (FTF-only) design. In addition, while the take-up rate of the Web starting mode was not particularly high, more than a thousand interviewer visits were avoided which indicated a potential cost savings. Future work will assess the long-term effects of introducing the Web starting mode on attrition in future waves of the IAB Establishment Panel, as has been studied in the household context (Lynn, 2013). Some research has shown that Web take-up rates have increased over time as establishments become more familiar with online data collection (Erikson, 2007; Millar et al., 2018; Snijkers & Jones, 2013; Thompson et al., 2015). Thus, it will be interesting to see whether the share of establishments participating online will increase in subsequent waves of the IAB Establishment Panel.

In conclusion, the study findings are encouraging for the IAB Establishment Panel and for other voluntary establishment surveys more generally. The introduction of a Web-first sequential mixed-mode design, as a replacement for a traditional FTF-first design, does not appear to negatively affect response rates or nonresponse bias, and shows some promise in recruiting larger establishments. This is direct evidence that nonresponse effects are not worsened by swapping out a relatively expensive interviewer-administered starting mode with a self-administered one and that some cost savings may result from this transition.

ORCID

Patrick Gleiser https://orcid.org/0000-0002-5731-4874

REFERENCES

AAPOR. (2016) *Standard definitions: final dispositions of case codes and outcome rates for surveys*, 9th edn. The American Association for Public Opinion Research.

Bavdaz, M., Giesen, D., Cerne, S.K., Löfgren, T. & Raymond-Blaess, V. (2015) Response burden in official business surveys: measurement and reduction practices of national statistical institutes. *Journal of Official Statistics*, 31(4), 559–588.

Bavdaž, M., Snijkers, G., Sakshaug, J.W., Brand, T., Haraldsen, G., Kurban, B. et al. (2020) Business data collection methodology: current state and future outlook. *Statistical Journal of the IAOS*, 36(3), 741–756.

Bechmann, S., Tschersich, N., Ellguth, P., Kohaut, S. & Baier, E. (2019) Technical report on the IAB Establishment panel * wave 25 (2017). (FDZ-Methodenreport, 01/2019 (en)).

Blossfeld, H.P. (1987) Labor market entry and the sexual segregation of careers in the Federal Republic of Germany. *The American Journal of Sociology*, 93(1), 89–118.

Bossler, M., Geis, G. & Stegmaier, J. (2018) Comparing survey data with an official administrative population: assessing sample-selectivity in the IAB establishment panel. *Quality & Quantity*, 52(2), 899–920.

Bremner, C. (2011) An investigation into the use of mixed mode data collection methods for UK business surveys. In: Deirdre, G. & Mojca, B. (Eds.) *Proceedings of the BLUE-ETS conference on burden and motivation in official business surveys, Heerlen, Netherlands, 22–23 March*, Statistics Netherlands, pp. 217–220.

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

Dutwin, D. & Lavrakas, P. (2016) Trends in telephone outcomes, 2008–2015, *Survey Practice*, 9(3), 1.

Earp, M., Toth, D., Phipps, P. & Oslund, C. (2018) Assesing nonresponse in a longitudinal establishment survey using regression trees. *Journal of Offical Statistics*, 34(2), 463–481.

Edwards, W.S. & Cantor, D. (1991) Towards a response model in establishment surveys. In: Biermer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. & Sudman, S. (Eds.) *Measurement errors in surveys*. New York: Wiley, pp. 211–233.

Ellguth, P. & Kohaut, S. (2014) Are they willing to use the web? * First results of a possible switch from PAPI to CAPI/CAWI in an establishment survey. In: Statistics Canada (Ed.) *Beyond traditional survey taking: adapting to a changing world. Proceedings of statistics Canada symposium 2014, Quebec*, pp. 1–6.

Ellguth, P., Kohaut, S. & Möller, I. (2014) The IAB establishment panel—methodological essentials and data quality. *Journal for Labour Market Research*, 47(1–2), 27–41.

Erikson, J. (2007) Effects of offering web questionnaires as an option in enterprise surveys: the Swedish experience (Paper presented at the ICES-III, Montreal June 18–21, pp. 825–829).

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

Hartmann, J. & Kohaut, S. (2000) Analysen zu Ausfällen (Unit-Nonresponse) im IAB-Betriebspanel. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 33(4), 609–618.

Hecht, V., Litzel, N. & Schäffler, J. (2019) Unit nonresponse at the firm level: a cross-border analysis using the IAB-ReLoc data. *Journal for Labour Market Research*, 53(2). Available from: http://doi.org/10.1186/s12651-018-0251-y

Hedlin, D., Dale, T. & Haraldsen, G. (2005) Developing methods for assessing perceived response burden (Tech. Rep.). Eurostat.

Janik, F. (2011) Unit non-response in establishments surveyed for the first time in the IAB establishment panel (Tech. Rep.). Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for …

Janik, F. & Kohaut, S. (2012) Why don't they answer? Unit non-response in the IAB establishment panel. *Quality & Quantity*, 46(3), 917–934.

Kantar. (2020) IAB-Betriebspanel, Arbeitgeberbefragung 2018. Methodenbericht (Kantar Group).

Kennedy, C. & Hartig, H. (2019) Response rates in telephone surveys have resumed their decline. *Pew research center*. Available from: https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-Kreuter

König, C., Sakshaug, J.W., Stegmaier, J., & Kohaut, S. (2021) Trends in establishment survey nonresponse rates and nonresponse bias: evidence from the 2001-2017 IAB Establishment Panel. *Journal of Official Statistics*, 37(4), 931–953.

Kreuter, F., Presser, S. & Tourangeau, R. (2009) Social desirability bias in CATI, IVR, and Web surveys: the effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865.

Küfner, B., Sakshaug, J.W. & Zins, S. (2020) Using administrative data and machine learning to address nonresponse bias in establishment surveys (Tech. Rep.). Paper presented at the BigSurv20 Conference (Big Data Meets Survey Science).

de Leeuw, E. (2005) to mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21(2), 233–255.

de Leeuw, E. (2018) Mixed-mode: past, present, and future. *Survey Research Methods*, 12(2), 75–89.

de Leeuw, E. & de Heer, W. (2002) Trends in household survey nonresponse: a longitudinal and international comparison. In: Groves, R.M., Dillman, D.A., Eltinge, J.L. & Little, R.J.A. (Eds.) *Survey nonresponse*. Hoboken: Wiley, pp. 41–54.

de Leeuw, E., Hox, J. & 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*, 1–11. https://surveyinsights.org/?p=10452

Lynn, P. (2013) Alternative sequential mixed-mode designs: effects on attrition rates, attrition bias, and costs. *Journal of Survey Statistics and Methodology*, 1(2), 183–205.

Marquette, E. & Kornbau, M. (2013) Testing contact and response strategies to improve response in the 2012 economic census. *Proceedings of the Federal Committee on Statistical Methodology*.

Millar, M.M., Schmuhl, P., Page, K., Genovesi, A.L., Ely, M., Hemingway, C. et al. (2018) Improving response to an establishment survey through the use of web-push data collection methods. *Mathematical Population Studies*, 25(3), 168–179.

Moore, J.C., Durrant, G.B. & Smith, P.W. (2018) Data set representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(1), 229–248.

Petroni, R., Sigman, R., Willimack, D., Cohen, S. & Tucker, C. (2004) Response rates and nonresponse in establishment surveys–BLS and Census Bureau. *Federal Economic Statistics Advisory Committee*, 1–50.

Phipps, P. & Toth, D. (2012) Analyzing establishment nonresponse using an interpretable regression tree model with linked administrativ data. *The Annals of Applied Statistics*, 6(2), 722–794.

R Core Team. (2017) R: a language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from: https://www.R-project.org/

Sakshaug, J.W. & Vicari, B.J. (2018) Obtaining record linkage consent from establishments: the impact of question placement on consent rates and bias. *Journal of Survey Statistics and Methodology*, 6(1), 46–71.

Schmucker, A., Ganzer, A., Stegmaier, J. & Wolter, S. (2018) Establishment history panel 1975–2017. FDZ-Datenreport, 09/2018 (en), Nuremberg. Available from: http://10.5164/IAB.FDZD.1809.en.v1

Schouten, B., Cobben, F., Bethlehem, J. (2009) Indicators for the representativeness of survey response. *Survey Methodology*, 35(1), 101–113.

Seiler, C. (2010) Dynamic modelling of nonresponse in business surveys (IFO Working Paper Series No. 93). IFO Institute - Leibniz Institute for Economic Research at the University of Munich.

Snijkers, G. (2008) Getting data for business statistics: a response model (Paper presented at the 4th European Conference on Quality in Official Statistics, Rome July 8–11).

Snijkers, G. & Jones, J. (2013) Business survey communication. In: Snijkers, G., Haraldsen, G., Jones, J. & Willimack, D.K. (Eds.) *Designing and conducting business surveys*. Hoboken: Wiley, pp. 359–430.

Thompson, K.J., Oliver, B. & Beck, J. (2015) An analysis of the mixed collection modes for two business surveys conducted by the US Census Bureau. *Public Opinion Quarterly*, 79(3), 769–789.

Tomoaskovic-Devey, D., Leiter, J. & Thompson, S. (1994) Organizational survey nonresponse. *Administrative Science Quarterly*, 39 (31), 439–457.

Voogt, R.J. & Saris, W.E. (2005) Mixed mode designs: finding the balance between nonresponse bias and mode effects. *Journal of Official Statistics*, 21(3), 67–387.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Gleiser, P., Sakshaug, J.W., Volkert, M., Ellguth, P., Kohaut, S. & Möller, I. (2022) Introducing Web in a mixed-mode establishment survey: Effects on nonresponse. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3), 891–915. Available from: https://doi.org/10.1111/rssa.12809