methods, data, analyses | 2026, pp. 1-32 DOI: 10.25521/mda.567

Switching a Face-to-Face Survey to Self-Completion in Germany: Considering the Impact of Survey Mode Design on Response Rate, Sample Composition, and Response Patterns in the European Social Survey

Jan-Lucas Schanze[1], Caroline Hahn[1], Oshrat Hochman[1], Niklas Donth[2]

[1] *GESIS – Leibniz Institute for the Social Sciences*
[2] *University of Stuttgart*

Abstract

The European Social Survey (ESS) is a repeated cross-national survey program adhering to rigorous methodological standards, especially regarding comparability. The COVID-19 pandemic challenged some of these standards by preventing the implementation of the face-to-face mode of data collection in some of the participating countries. Germany and other countries switched to web- and paper-based self-completion modes, marking a stark deviation from previous ESS rounds. In this paper, we examine the impact of the mode switch on response rates, bias in the sample composition (e.g., age, education, migration background), and response behavior (e.g., item nonresponse, response patterns) by comparing the German self-completion ESS (Round 10) with the preceding and the succeeding face-to-face rounds. We found a much higher and more regionally balanced response rate in the self-completion ESS and small differences in bias in sample composition and response behavior between rounds. Within Round 10, respondents choosing the web mode differed in sociodemographic characteristics from those who completed the paper-based questionnaire.

Keywords: self-administered mixed-mode survey, face-to-face survey, mode switch, sample composition, response patterns

Face-to-face interviews have long been considered the gold standard of survey data (Rybak, 2023; Villar & Fitzgerald, 2017). In recent years, this gold standard has come under increasing pressure (Couper, 2017) as various face-to-face surveys in many countries, including Germany, have reported a sharp decline in response rates over time (Beullens et al., 2018; de Leeuw et al., 2018; Wolf et al., 2021). In an effort to counteract this trend, various measures have been implemented, for example, setting a minimum number of contact attempts for target persons (Meitinger et al., 2020; Stoop, 2005) or offering incentives (Pforr et al., 2015). Although these measures have resulted in increasing costs for face-to-face interviews (Wolf et al., 2021), they do not seem to offer a satisfactory and sustainable solution to counteract the trend of declining response rates (Luiten et al., 2020).

At the same time, technological advances, increasing internet penetration rates, and a cultural shift toward self-mediated activities have paved the way for alternative methods of data collection (Dillman, 2000). Due to the increasing dominance of smartphones as a means to use the internet (Eurostat, 2023), online surveys are literally in the palm of the respondents' hands. However, there are several drawbacks and challenges associated with the use of self-administered modes of data collection. Collecting data solely online implies a significant selection bias, because certain groups like the elderly, the less educated, or the less wealthy are less likely to "go online" than others due to lack of accessibility, skills, or trust (Biemer et al., 2018; Díaz de Rada, 2022; Dillman, 2017; Messer & Dillman, 2011). One way to avoid the exclusion of these population segments is to offer a combination of online and paper-based modes of data collection (Messer & Dillman, 2011).

Paper questionnaires present a different set of challenges and limitations that must be considered when evaluating the performance of a self-administered mixed-mode design. Compared with web surveys, they are more costly, take longer to process, and are more prone to data quality issues caused by incorrect routing or undesired response behavior (e.g., multiple ticks in single-tick items; e.g., Wolf et al., 2021). Thus, while paper questionnaires achieve a more balanced sample, they also pose more challenges for the data processing.

The use of smartphones to respond to web surveys also presents challenges. Studies have shown that it is associated with lower completion rates (Toepoel & Lugtig, 2014) and higher break-off rates (Couper et al., 2017) than other devices.

Acknowledgements
The authors would like to thank Selina Haaga and Anny Ngo for all their support in preparing this research article.

Direct correspondence to
Jan-Lucas Schanze, GESIS – Leibniz Institute for the Social Sciences, B6 4-5,
68159 Mannheim, Germany
E-mail: jan-lucas.schanze@gesis.org

In terms of response quality, previous findings seem to suggest a mixed picture, but where differences in quality across devices have been found, they seem to be fairly small (Lund Clement & Severin-Nielsen, 2020).

Self-administered mixed-mode surveys have long been an established method in the field of social survey research (de Leeuw, 2005; Couper, 2017). To investigate how well they can replace face-to-face surveys in large general population studies, survey methodologists have conducted experiments on online surveys and mode switching (Lugtig et al., 2022; Luijkx et al., 2021; Villar & Fitzgerald, 2017). Mode-switch experiments are particularly challenging as they require costly "parallel runs" with a random assignment of target persons to the face-to-face mode or the self-completion mode. Very few studies to date have managed this, e.g., the German General Social Survey in 2023 (Baumann et al., 2025) or the ESS in Finland and Great Britain (Lugtig, 2024a). Still in the context of the COVID-19 pandemic, face-to-face interviews presented new challenges like strict personal contact restrictions, thus forcing a mode-switch (Bautista, 2021; Gummer et al., 2020; Roßmann, 2022; Scherpenzeel et al., 2020).

In this study, we provide an overview of the first ever European Social Survey (ESS) conducted in Germany in a self-administered mixed-mode design. The entire operation was done in compliance with the ESS Core Scientific Team Specifications for a self-completion mixed-mode design via web interviews and paper questionnaires. The mode switch provided us with an opportunity to learn about whether and how well ESS data can be collected using a self-completion design. In this paper, we consider, first, response rates, second, bias in sample composition, and third, response patterns. We use this broad range of quality indicators to evaluate the performance of the ESS after the mode switch, comparing the results of Round 10 of the German ESS, conducted in 2021/22 using a mixed-mode (web and paper) design, with those of Round 9 and Round 11, conducted face-to-face in 2018/19 and 2023, respectively.

It is important to note that the present comparison of the mixed-mode self-completion design and the face-to-face design in the ESS in Germany is not based on an experimental test in which both designs were implemented in parallel. Instead, we compare three consecutive rounds. Although contextual factors, such as the impact of the COVID-19 pandemic during the Round 10 data collection, must be considered, we believe that our analysis provides important indications of the feasibility of implementing a self-administered design in the ESS in Germany and the potential consequences for key data quality indicators. In particular, the inclusion of Round 11 helps us understand whether changes from Round 9 to Round 10 are due to the mode switch or due to contextual changes or other unknown factors.

In the next section we will discuss the literature on self-administered mixed-mode designs and on mode switching in repeated cross-sectional surveys. In this context, we also present our guiding research questions for the present study.

In Section 3 we describe our data and methods, and in Section 4 we report our main results. We conclude by pointing out the main lessons we derive from the implementation of the self-completion ESS in Germany and its outcomes, and we suggest some ideas for future research.

State of Research

Response rates are a first meaningful quality measure in probability-based samples. The participation of a reasonably high number of people from the sample ensures that valid inferences can be made about the target population (Stoop, 2005). In addition, as Luiten et al. (2020) noted, "nonresponse results in a smaller realized sample and may result in a longer fieldwork period" (p. 469–470), both of which are consequential for the analytical potential of the data. Previous studies suggest that self-administered surveys may be less effective than interviewer-administered surveys in persuading target persons to participate (Groves et al., 2009; Villar & Fitzgerald, 2017). This may be because it is easier to overlook or ignore a letter than an interviewer at the door. Interviewers can also tailor their recruitment approach and address any concerns of target persons directly. At the same time, mixed-mode designs may help mitigate declining response rates by offering respondents more flexibility to complete the questionnaire at a time, pace, and place that suits them.

Although response rates in web-only surveys are typically 12 percentage points lower than in other survey modes, such as paper-based questionnaires or telephone surveys (Daikeler et al., 2020), recent studies have shown promising results when combining web and other modes. In an experiment conducted as part of the 2018 Generations and Gender Survey (GGS) across three countries, the response rates between a face-to-face condition and a push-to-web condition with a face-to-face follow-up varied significantly by country and design (Lugtig et al., 2022). In Germany, the face-to-face condition performed slightly better than the push-to-web design. Similarly, in the 2017 European Values Study (EVS) six countries participated in a mode experiment, in which target persons were randomly assigned to either a self-administered mixed-mode design offering a paper and a web questionnaire or a face-to-face design (Luijkx et al., 2021). The results were mixed regarding the response rates. Germany, Iceland, and the Netherlands reported higher response rates with the self-administered designs, whereas Denmark, Finland, and Switzerland reported higher response rates with the face-to-face design.

In Germany, self-administered mixed-mode designs may also reduce interviewer-related regional variation in cooperation rates and response rates. Because interviewers typically work in specific regions and sampling points, the regional response rates partly reflect the interviewer's performance. In

contrast, in the self-administered design, differences in response rates across municipalities can be explained by sociodemographic characteristics but are not increased by unequal recruiting effort.

The decision to participate in a survey is associated with specific sociodemographic characteristics. Nonrespondents are not missing at random in survey samples (Groves & Peytcheva, 2008; Groves et al., 2009). While interviewers can adapt doorstep behavior to address target persons' needs, such tailoring is more difficult in self-administered designs, as little is typically known about a target person in advance. Accordingly, empirical evidence shows that sample composition varies by mode (Díaz de Rada, 2022; Luijkx et al., 2021; Rybak, 2023; Villar & Fitzgerald, 2017; Wolf et al., 2021), and biases observed in face-to-face surveys are even more pronounced in self-administered surveys. In Germany, the latter suffer from a stronger overrepresentation of highly educated respondents (Stein et al., 2025) and underrepresent persons with a migration background (Luijkx et al., 2021; Lynn & Xena, 2023).

In contrast, in the GGS experiment, Lugtig et al. (2022) found a lower nonresponse bias for gender, age, nationality, and urbanicity in the web mode with a small prepaid incentive (€5) than in the face-to-face mode. Indeed, even though nonrespondents in the push-to-web mode were approached by interviewers, this follow-up phase did not change nonresponse bias significantly. Similarly, in their meta-analysis, Cornesse and Bosnjak (2018) report that while web-only surveys are less representative than other single-mode designs, mixed-mode surveys tend to be more representative than single-mode surveys, implying that adding a paper option may mitigate nonresponse biases in web surveys. Consistent with this, in a meta-analysis of survey outcomes in the cross-national International Social Survey Programme (ISSP) and ESS, Rybak (2023) found higher response rates in face-to-face surveys and lower nonresponse bias in mail-only or self-administered mixed-mode surveys if they used a sampling frame of individuals. In the present study, we therefore also investigate whether there are differences in bias between samples from the self-completion and the face-to-face mode, comparing each sample with sociodemographic data from the German Microcensus.

Different modes of data collection also affect the response behavior of survey participants. In face-to-face surveys, interviewers can guide respondents through the questionnaire and provide clarifications, but this requires strict adherence to standardized scripts to limit interviewer effects. Interviewers may influence respondents' answers through their visible socio-demographic characteristics, suggestive probing (Beullens & Loosveldt, 2016; West & Blom, 2016), or by deviating from the specified instructions for the interview (Ackermann-Piek & Massing, 2014). Self-administered modes place greater navigational and cognitive demands on respondents, but they are typically less affected by social desirability bias (de Leeuw, 2005; Heerwegh, 2009; Lugtig, 2024b) because

respondents can complete the survey on their own and might perceive it as more anonymous.

During face-to-face interviews, interviewers can keep respondents focused and motivated, whereas in self-administered modes, respondents may lose focus and concentration when filling out the questionnaire on their own (Dillman, 2017). This is especially important for social surveys with their long, multifaceted questionnaires. With a duration of around 20 to 30 minutes, many self-completion surveys discussed in the literature (e.g., Biemer et al., 2018; Peytchev et al., 2019; Roberts et al., 2016; Toepoel & Lugtig, 2022) are shorter than the ESS questionnaire. It has been shown for longitudinal surveys that significantly longer survey durations of up to two hours are possible, even if this leads to more frequent pauses and interruptions during the response (Couper & McGonagle, 2019). This prompts the question of what impact the length of a self-administered questionnaire has on data quality. Andreadis and Kartsounidou (2020) observed higher response rates when a long questionnaire was divided into two parts, with minimal differences in response quality indicators. Cernat et al. (2022) found comparable response rates between a long self-completion questionnaire (55 minutes) and its shorter version (38 minutes), suggesting respondents can manage longer surveys.

Moreover, respondents are more likely to multitask during self-administered surveys, especially in web surveys. Self-administered surveys have thus been associated with greater non-differentiation, higher levels of straightlining, higher item nonresponse, and more midpoint responding than face-to-face surveys (Čehovin et al., 2023; Cernat et al., 2022; Heerwegh & Loosveldt, 2008; Wolf et al., 2021). However, results are inconsistent across studies, and effect sizes are modest. Self-administered surveys also allow respondents to control the pace of the interview, read questions more slowly and multiple times, and see all questions and answers in front of them, which may improve the quality of responses.

An additional factor influencing response behavior is the presentation of questions and answers either verbally in face-to-face interviews or visually in self-administered interviews. When questions and response categories are read aloud by the interviewer, respondents may disproportionately select the last applicable answer they remember due to memory capacity (de Leeuw, 2005). This is known as the recency effect. Using the EVS data, Cernat et al. (2022) observed slightly higher levels of recency effects in face-to-face than in self-completion interviews. Recency effects are expected to be rather small in ESS face-to-face interviews because interviewers must use showcards. Thus, visual stimuli were also used in the face-to-face rounds. By contrast, when response categories are presented visually in web or paper questionnaires, respondents tend to select the first answer that they feel applies to them, which leads to the primacy effect (de Leeuw, 2005). While there is empirical evidence supporting

the existence of primacy and recency effects (e.g., Galesic et al., 2008; Holbrook et al., 2007), results have not always been consistent (Dillman et al., 1995).

In the present study, we examine data from Round 10 of the ESS in Germany, which was collected using a self-administered mixed-mode design (web and paper questionnaires). The analyses focus on response rates, bias in sample composition, and response patterns, including differences in item nonresponse, the occurrence of recency and primacy effects, and the prevalence of straightlining. We compare the results from Round 10 with those of the preceding ESS Round 9 and the succeeding ESS Round 11 to infer how these indicators were affected by the switch from a face-to-face to a self-administered mixed-mode design. This comparison of three consecutive rounds of the ESS in Germany with two different mode designs allows us to address the following research questions:

RQ1: *What are the consequences of the switch from a face-to-face to a self-administered mixed-mode design for response rates?*

RQ2: *Are response rates more balanced across regions due to the absence of interviewer effects in the recruitment of respondents?*

RQ3: *Does the bias in sample composition differ across survey mode designs?*

RQ4: *Do we observe changes in response behavior when the survey mode design is changed?*

Data and Methods

The European Social Survey (ESS) started to collect data across Europe in 2002 and has been conducted every two years since then. Germany has participated in all ESS rounds and employed professional interviewers to collect data in the face-to-face mode up to and including Round 9 in 2018, which was carried out between late August 2018 and early March 2019. In contrast, Round 10 was implemented as a self-administered mixed-mode survey. In Round 11, carried out in 2023, the entire ESS switched back to the face-to-face mode. Table 1 summarizes the main design features of the three ESS rounds in Germany. In all rounds analyzed here (European Social Survey European Research Infrastructure [ESS ERIC], 2021, 2022, 2025), a random sample of the population living in private households was drawn from the decentralized population register. Fieldwork for ESS rounds 9 and 10 was carried out by the infas Institute for Applied Social Sciences. In Round 11 Verian Deutschland (formerly Kantar Public) was commissioned to conduct the ESS in Germany.

Table 1 Design features of the three rounds of the European Social Survey (ESS) analyzed in this study

	Round 9	Round 10	Round 11
Survey agency	infas Institute for Applied Social Sciences	infas Institute for Applied Social Sciences	Verian Deutschland (formerly Kantar Public)
Fieldwork period	08/2018–03/2019	10/2021–01/2022	05/2023–12/2023
Duration of fieldwork	27 weeks	13 weeks	32 weeks
Target population	General population 15 years or older	General population 15 years or older	General population 15 years or older
Sampling frame	Decentralized population register	Decentralized population register	Decentralized population register
Number of sampling points (number of municipalities)	198 (178)	200 (179)	200 (183)
Sampling design	Multi-stage, fully random	Multi-stage, fully random	Multi-stage, fully random
Gross sample size	8,695	23,997	9,200
Survey mode design	Face-to-face (CAPI)	Self-completion (CAWI/PAPI)	Face-to-face (CAPI)
Language of the questionnaire	German	German	German
Respondent incentives	€20/€30/€50 postpaid	€5 prepaid + €0/€10 postpaid	€5 prepaid + €10/€20 postpaid
Share of PAPI questionnaires	N.A.	34.4%	N.A.
Number of active interviewers	216	N.A.	163
Number of invitation letters	2	4	2
Duration of survey (median)	74 min	52 min	65 min

Notes: CAPI = Computer-assisted Personal Interview; CAWI = Computer-assisted Web Interview; PAPI = Paper-and-Pencil Interview; N.A. = not applicable.

For ESS Round 10, special standards were developed by the ESS Core Scientific Team (CST) to ensure that data collection in the self-completion modes would provide data of similar and sufficient quality across countries. One requirement was that a prepaid cash incentive is sent with the postal invitation. The web-based questionnaire was programmed to fit different devices (e.g., Maslovskaya et al., 2019). Further, the self-administered mixed-mode survey was about 10% shorter than the face-to-face interview.

In Germany, data collection for ESS Round 10 started in October 2021 and ended in early January 2022. As required by the central guidelines, a sequential push-to-web design was implemented for most of the gross sample. All target persons received a postal invitation to complete a 50-minute online questionnaire.

Among the target population older than 60 years, 50% were randomly assigned to the sequential push-to-web design, receiving the paper questionnaire once with the second reminder, while the other 50% received the paper questionnaire once together with the first invitation letter (concurrent design). Three reminders were sent in total to all target persons who had not responded or opted out after the first, third, and fifth week of fieldwork, respectively.[1] Interviewers or fieldworkers were not involved in the recruitment for ESS Round 10 in Germany.

Dependent Variables

RQ1 and RQ2: Response Rate

Comparing response rates across the three ESS rounds is less straightforward than it might first appear. In Rounds 9 and 11, interviewers visited every sample unit – as many as four times or more if required – and completed a contact form for every contact attempt. In Round 10, by contrast, there was exclusively postal contact with target persons, and the reasons for nonresponse, whether they were non-contacts, refusals, or ineligibles, were unknown. Therefore, although we might be able to compare the number of interviews with the number of returned questionnaires, other disposition codes are hardly comparable. For this reason, we use only AAPOR Response Rate 2 (RR2; American Association for Public Opinion Research [AAPOR], 2023), dividing the number of valid complete and partial interviews by the number of cases in the entire gross sample. Valid interviews are defined as those interviews that were published by the ESS following quality checks to ensure that interviews were conducted according to the specifications and that sociodemographic characteristics of the respondents matched the information on the selected target persons in the population register. Partial interviews exhibit between 50 and 75% of missing values but were still published by the ESS following quality checks.

In addition to analyzing the aggregated response rate (RQ1), we also examined the response rate achieved in each of the approximately 180 municipalities (RQ2).

RQ3: Bias in Sample Composition

Following the analysis of response rates, we compared the sample composition of each of the three rounds with administrative benchmark data derived from the German Microcensus collected in 2018, 2021, and 2023 (Statistisches Bundesamt, 2026). Here, we considered respondents' gender, age, formal level

[1] All letters (in German) can be downloaded from the ESS data portal (Country documentation: Germany): https://ess.sikt.no/en/study/172ac431-2a06-41df-9dab-c1fd8f3877e7/433).

of educational attainment,2 citizenship3 and migration background4, whether they lived in former eastern or western Germany; the urbanicity of their dwelling place (GKPOL7); and their labor force status (unemployed or employed, with all remaining inactive groups included in a residual category). For each category of those variables, we calculated the relative bias, dividing the share of the category by the respective share reported in the German Microcensus for the given year. The Microcensus is implemented annually by the federal and state statistical offices. It is a representative survey of 1% of households in Germany, and participation for selected persons is mandatory. Therefore, it is regarded as the most reliable source of sociodemographic information about the resident population in Germany. To test whether the differences in sample composition between each ESS round and the German Microcensus are statistically significant, we conducted two-sample proportion z-tests. For the visualization in graphs, we calculated the 95% confidence intervals for each share of a given ESS variable to determine whether the intervals include values provided by the Microcensus.

RQ4: Response Patterns

Complementing the analyses of sample composition, we also compared indicators of data quality across the two survey mode designs in Rounds 9, 10, and 11 and between modes within Round 10. We calculated the share of item nonresponse, and we looked at response patterns – namely, primacy and recency effects and straightlining.

The level of item nonresponse is strongly influenced by the survey design and the questionnaire design. Regarding questionnaire design, the three ESS rounds analyzed here used hidden codes for refusals. In the face-to-face mode, interviewers were instructed not to read out refusal codes, such as “Don’t know” or “Prefer not to answer”, and not to prompt such responses. The ESS Round 10 questionnaires did not include “Don’t know” and “Refusal” options. Instead, at the beginning of the questionnaire, respondents were instructed to leave questions blank if they did not know the answer or did not want to answer. In the paper questionnaire, respondents sometimes ticked more than one response option even if they were required to select a single option. For most items, those inconsistencies were coded as missing values, as the response could not be interpreted and processed in a meaningful way. This ESS-wide decision for data pro-

---

2 The category “low education” includes all qualifications up to lower secondary schooling (ISCED Levels 0–2); “intermediate education” comprises upper secondary and non-academic vocational qualifications (ISCED Levels 3–4); “high education” refers to all successfully completed tertiary education programs (ISCED Levels 5–8).

3 Whether they have German citizenship or not.

4 A combination of citizenship (German/non-German) and immigration experience (immigrant/born in Germany).

cessing increased the share of item nonresponse only for paper questionnaires, as incorrect multiple responses were prevented by the CAWI programming. We restricted our analysis of item nonresponse to the core questionnaire that was fielded in all three rounds of the ESS. This left us with 182 core items in Round 9, 179 core items in Round 10, and 178 core items in Round 11.[5]

Primacy and recency effects were analyzed in 39 attitudinal items. For each respondent, we counted how often either the first, uppermost, or rightmost (primacy effect) or the last, lowest, or leftmost (recency effect) option of a response scale was selected. This value was divided by the individual count of substantive responses. Thus, if a respondent did not answer five out of the 39 items, the number of first and last ticks was divided by 34.

Straightlining was analyzed in three batteries of items with at least 4 items fielded in all three rounds. The item batteries were the trust battery (7 items with an 11-point scale; e.g., trust in the UN[6]), the satisfaction battery (6 items with an 11-point scale; e.g., satisfaction with the national government), and the social equality statement battery (4 items with a 5-point scale; e.g., ashamed if close family member gay or lesbian). If respondents differentiated their replies to those items, they were coded as 0; if they used straightlining and answered all items within an item battery with the same answer option, they were coded as 1. Results from the three batteries were then summarized in a binary variable capturing whether respondents used straightlining in any of the three item batteries (coded as 1) or varied their replies (coded as 0). Respondents who did not answer any of the items (100% missing values) were not defined as straightliners.

To assess whether differences across modes are statistically significant, we used OLS regressions for metric dependent variables (share of item nonresponse, share of ticks to first or last response category) and logistic regressions for binary dependent variables (occurrence of straightlining). In all multivariate models analyzing response patterns, we included various sociodemographic variables in order to disentangle whether differences between face-to-face, paper-based, and web-based responses were caused by the mode or by underlying differences in the sample composition. The following set of variables should at least partially explain self-selection into modes: age group, sex, level of educational attainment (in three groups based on the International Standard Classification of Education [ISCED] 2011), migration background, living in former eastern or western Germany, being in paid work, being retired, frequency of internet usage, and self-rated urbanicity of the place of residence. In all our analyses, we weighted the data with the design weights provided by the ESS.

---

[5] The small difference can be explained with further minor changes in the questionnaire. For instance, the number of items measuring level of education was reduced from three to two in Germany.

[6] Trust in scientists was excluded because it was fielded only in Round 10.

Results

In this section, we present the results of our analyses of the impact of the mode switch in the German ESS on overall and regional response rates, sample composition, and response patterns.

RQ1: Response Rates

In ESS Round 9, the gross sample comprised 8,695 persons aged 15 years or older and living in private households across Germany. By the end of the fieldwork period, 2,358 respondents had participated. In 10 interviews, respondents broke off the interview before reaching its end. We labeled these interviews as "partial." Dividing the number of complete and partial interviews by the number of people in the gross sample resulted in an AAPOR Response Rate 2 (RR2) response rate of 27.2%. A little more than half of all target persons refused to participate (51.9%), and a few cases were noncontacts (1.8%, Ghirelli et al., 2022: 61).

In Round 10, a sample comprising 23,997 persons aged 15 years or older and living in private households in Germany was drawn from the population register. Of these, 8,376 completed the survey interview. In addition, 349 respondents (1.5% of the gross sample) broke off the web interview before reaching the end but still answered at least 50% of the questions and were therefore included in the data release. Dividing those 8,725 cases by the number of cases in the gross sample resulted in an RR2 of 36.4%.

In Round 11, 9,200 persons aged 15 years or older and living in private households in Germany were invited to participate in the ESS. At the end of fieldwork, 2,420 respondents followed the invitation and conducted a valid interview with their interviewer, which amounts to an RR2 of 26.3%. Compared to the previous face-to-face round, the share of target persons that were defined as noncontacts increased to 9.7%; 46.6% of the gross sample explicitly refused to participate (ESS ERIC, 2025).

Thus, Round 10 of the ESS in Germany achieved a 10 percentage points higher response rate in the self-completion design than the previous and the following face-to-face round, which answers our first research question (RQ1). While the German response rate increased significantly in ESS Round 10 compared to Round 9, the reported response rates decreased in all other ESS countries that switched from a face-to-face mode to self-completion modes (Lynn & Xena, 2023). The response rates indicate that people in Germany are willing to participate in a rather long self-completion survey. In fact, more people respond to this type of survey than to a face-to-face survey. Two-thirds of respondents participated online (66.6%); the remaining 33% did so on paper. In the web mode, most respondents used a personal computer to participate (70.2%), followed by smartphones (25.2%) and tablets (4.6%).

Importantly, whereas in the two face-to-face rounds, roughly 30 interviews (0.3% of the gross sample) were classified as invalid following quality checks and hence excluded, this number was much higher in the self-administered modes in Round 10, reaching 588 interviews (2.5% of the gross sample). In most cases, this was due to the "wrong" respondent answering the questionnaire on behalf of the target person within the household. We identified correct respondents by comparing information on sex and age derived from the population register with the corresponding information provided in the questionnaire.

RQ2: Regional Differences in Response Rates

Analyzing the response rates at the level of municipalities (Figure 1) revealed another disadvantage of face-to-face general population surveys, namely, that they rely on having highly qualified interviewers in all regions who have enough time and resources to contact and persuade their cases with equal effort. In a large country like Germany, with a shrinking interviewer workforce, this has become more difficult to achieve. Comparing municipal response rates in Rounds 9, 10, and 11, the range was much larger in the two face-to-face rounds (R9: 4.5% to 63.6%; R11: 0.0% to 56.5%) than in the self-completion round (R10: 13.3% to 51.6%).[7] Larger differences in municipal response rates could result in larger deviations in the sample composition.

Differences in response rates could also be detected at a higher level of regional division – the 16 German federal states (Laender). As can be seen in Table A1 in the Appendix, the spread of response rates was a little larger in Round 9 (14.8 percentage points, ranging from 18.2% to 33%) and a lot larger in Round 11 (27.2 percentage points, ranging from 3.3% to 30.5%) compared with Round 10 (12.6 percentage points, ranging from 28.5% to 41.1%). In all federal states, the response rate was higher when the self-completion mode was used (Lynn & Xena, 2023), even though the change was not always statistically significant due to small sample sizes in some of the federal states. Interestingly, the differences in response rates between German federal states of the former eastern and western German federal states were nearly identical in ESS rounds 9 and 10. In both parts of Germany, it increased from 27% to 36% when the mode was switched. In Round 11, the response rate decreased to 27% in western federal states and 24% in eastern federal states.

[7] Note that different municipalities are sampled in each ESS round. Whereas large municipalities are always included in the sample, smaller municipalities are randomly drawn as the first stage in the multi-stage sample design.

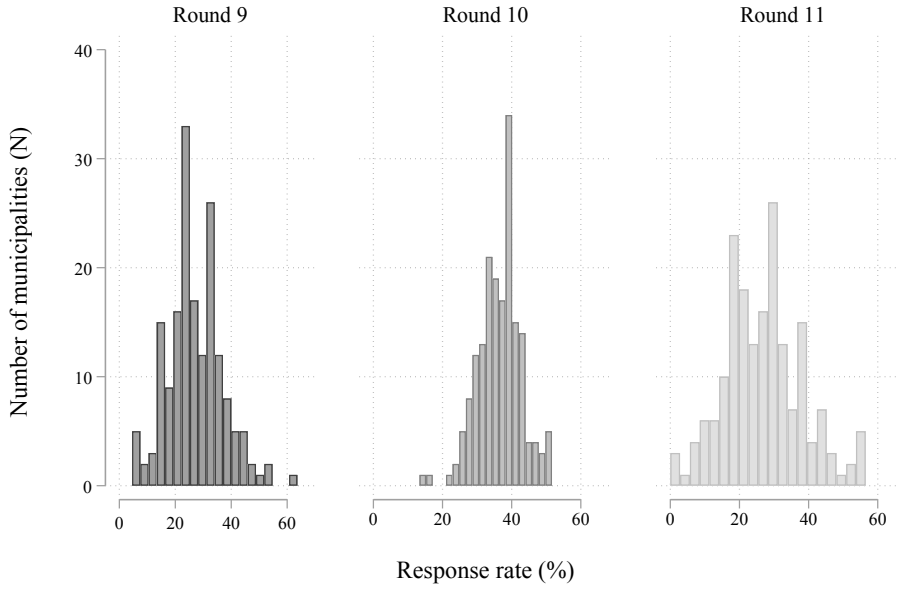

Figure 1 Density plot of response rates at the level of German municipalities in ESS rounds 9, 10 and 11

Municipalities can also be classified by their number of inhabitants. Figure A1 in the Appendix shows response rates for the three rounds in rural and urban regions. In Round 9, the statistically significant difference between the smallest and largest municipalities amounted to 5 percentage points ($p < .05$) and 8 percentage points in Round 11 ($p < .01$). This difference cannot be attributed entirely to the use of interviewers, as decreasing rates with increasing municipality size were also observed in the self-administered mode, albeit to a lesser extent (difference of about 3 percentage points, $p < .05$).

To summarize the analysis of regional differences (RQ2), ESS Round 10, which was conducted in self-administered modes, achieved a regionally more balanced response rate at the level of individual municipalities than ESS rounds 9 and 11 did. We believe this finding is a result of the fact that Round 10 did not use interviewers and hence was unaffected by different interviewer assignments and fieldwork efforts as in Round 9 and Round 11.

RQ3: Bias in Sample Composition

Turning to the analysis of bias in the sample composition, the age distribution was expected to differ across survey modes, as target persons might have had different mode preferences depending on their age. The two groups of bars on the left and the bars for Round 11 on the right in Figure 2 show only small dif-

ferences across the three rounds. Whereas the youngest age group (15–24 years) was significantly underrepresented in Round 9, this was no longer the case in rounds 10 and 11. Patterns of underrepresentation for young (25–34 years) and early middle-age (35–44 years) groups bore a great resemblance irrespective of whether interviewers were used or not. In the two face-to-face rounds, we observed a stronger overrepresentation of respondents aged 65–74 years, whereas the overrepresentation of the oldest age group (75 years and older) was less pronounced than in Round 10. In Round 11, the oldest age group was even slightly underrepresented. To summarize, this first analysis does not indicate that either the face-to-face mode or the self-administered modes had a clear edge in representing all age groups equally well.

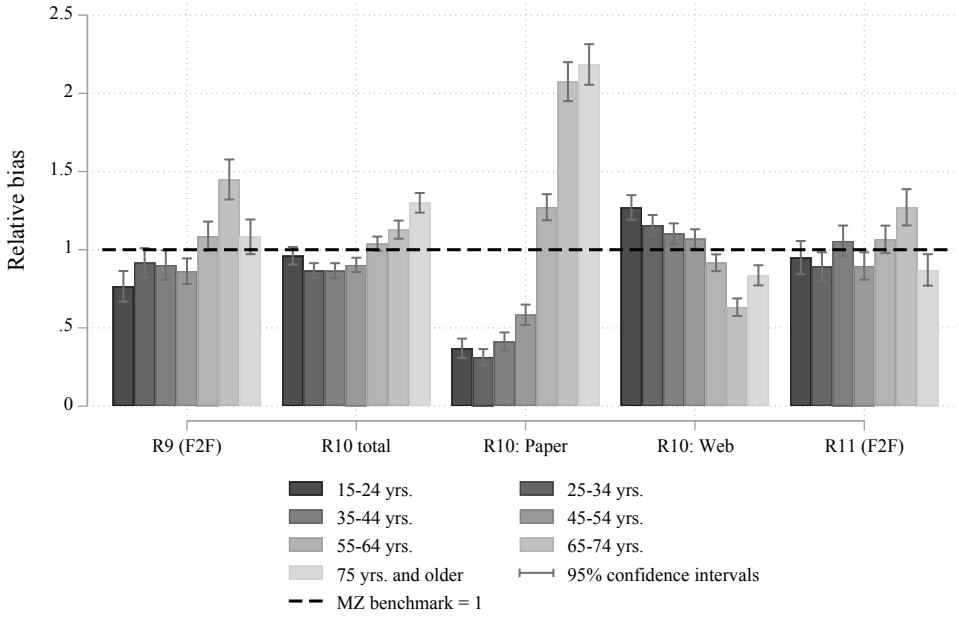

Figure 2 Comparison of the relative bias in the age distribution in ESS rounds 9, 10, and 11 with the German Microcensus (MZ) benchmark (= 1)

Looking at Round 10 data revealed strong differences in respondents' self-selection into modes (paper or web) depending on their age. Younger age groups were slightly overrepresented in the web mode but strongly underrepresented in the paper mode. The opposite was true for older respondents, who predominantly chose to answer the questionnaire on paper. Only the combination of both modes balanced the overall age distribution.

Round 10 achieved a more balanced sex distribution than Round 9 (see Figure A2 in the Appendix). In Round 9, women were underrepresented compared with

men. The data collected in Round 10 show a balanced sex distribution, as women were overrepresented in the paper questionnaires, whereas they were underrepresented in the web mode. However, the better representation of the two sexes cannot be attributed to the mode switch, as the face-to-face Round 11 also led to a more equal sex distribution than in Round 9.

Surveys in Germany and elsewhere struggle to represent all levels of education equally, and in all three rounds of the ESS, we accordingly observed a strong underrepresentation of respondents with a low level of education (see Figure 3). In Round 9 the size of the group of respondents with a low level of education was 61% of the required size according to the administrative benchmark (Microcensus), and it was even smaller in Round 10 (41%) and Round 11 (44%). While the comparison of Round 10 with Round 9 could lead to the conclusion that the move away from interviewers appears to have exacerbated the underrepresentation, the comparison with the latest ESS face-to-face data does not support this conclusion. Apparently, the mode switch alone does not change the representation of people with a low level of education. Moreover, respondents with an intermediate level of education were well represented in Rounds 9, 10, and 11, and respondents with a high level of education were highly overrepresented by a factor of more than 1.5 in Rounds 9 and 10 and 1.4 in Round 11.

Within Round 10, we found differences in education between respondents participating online or on paper. Target persons with a low or intermediate level of education participated more often on paper, while target persons with a high level of education more often preferred the web.

Foreign citizens and people with a migration background are another group that is usually underrepresented in social surveys. In Round 9, the relative size of the group without German citizenship was only 57% of the respective size in the administrative benchmark data, the German Microcensus (Figure 4). The self-administered modes in Round 10 led to a further reduction in the share of persons without a German citizenship to 41% of the respective size in the Microcensus. In Round 11, the underrepresentation was slightly mitigated, although it was still very pronounced at 51% of the required group size. This development indicates that the stronger underrepresentation in Round 10 is probably a result of the mode switch. The underrepresentation of non-German citizens was greater in the paper-based than in the web-based mode. Table A2 shows the breakdown of the respondents by migration background. For some categories, the difference between the three rounds was very small, showing at most a slight advantage for the interviewer-administered rounds 9 and 11, which better covered foreign citizens with migration background and German citizens with migration background.

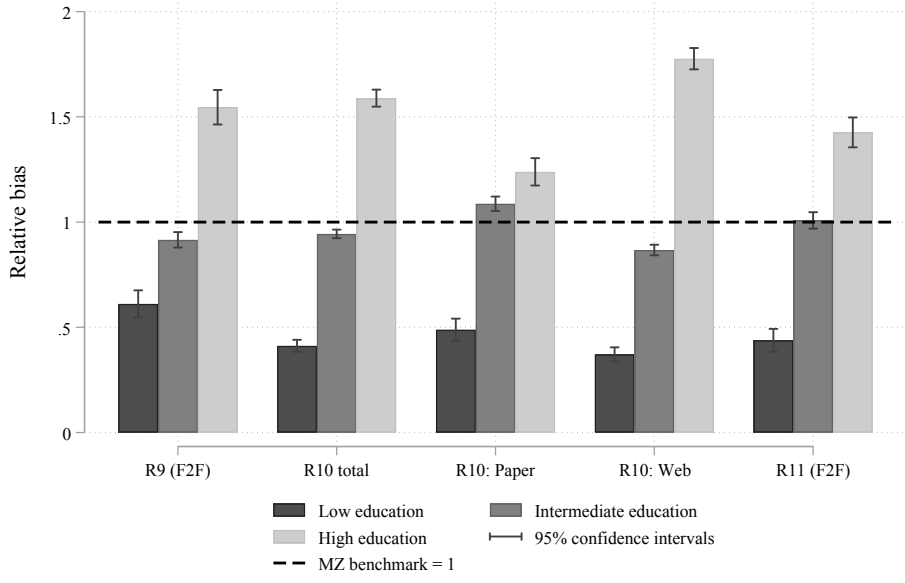

Figure 3 Comparison of the relative bias in respondents' level of educational attainment in ESS rounds 9, 10, and 11 with the German Microcensus (MZ) benchmark (= 1)

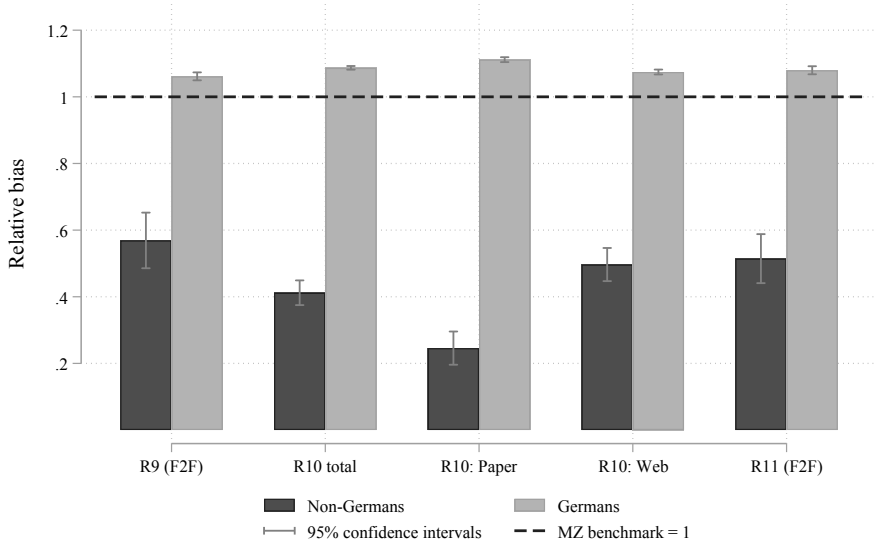

Figure 4 Comparison of the relative bias in respondents' citizenship in ESS rounds 9, 10, and 11 with the German Microcensus (MZ) benchmark (=1)

Roughly 2% of the German population were unemployed and actively looking for a job between 2018 and 2023. In Round 9, this small, disadvantaged group was slightly underrepresented. However, the 95% confidence interval included the "true" value derived from administrative statistics (see Figure A3 in the Appendix). In Round 10, and to a lesser extent also in Round 11, the underrepresentation of the unemployed population was more pronounced. By contrast, the population in paid work was slightly overrepresented in all three rounds. Within Round 10, there is an overrepresentation of this population in the web-based mode and an underrepresentation in the paper-based mode. The lower share of employed persons in the paper-based mode was also driven by the higher share of older respondents (see Figure 2), most of whom are retired.

Regarding RQ3, about the consequences of the mode switch to sample composition, the results are mixed. Although the sample composition in Rounds 9, 10, and 11 was similar in some key characteristics, such as age, the self-administered design further exacerbated the underrepresentation of certain marginalized groups, particularly unemployed persons and those without German citizenship or those who migrated to Germany. Comparing the two face-to-face rounds, we observe a declining quality of the sample composition over time.

RQ4: Response Patterns

Figure 5 depicts the distribution of item nonresponse rates in the ESS core questionnaires in ESS Rounds 9, 10, and 11. More questionnaires were completed without any missing responses in Round 9 (57.6%) and Round 11 (55.4%) than in Round 10 (38.7%). The interviewer-administered implementation prevented interviews with a particularly large number of missing values (see Figure 5). In Rounds 9 and 11, less than 4% of questionnaires exhibited more than 5% missing values on ESS core items, whereas 20.6% of questionnaires did so in Round 10. On average, the share of missing values was 3.4 percentage points higher in Round 10 than in Round 9 and Round 11 (bivariate OLS regression, $p < .001$). Round 9, which was conducted face-to-face, yielded a lower share of missing values than both the web (2.9 percentage points) and the paper mode (4.3 percentage points) in Round 10. This mode effect was driven partly by differences in the sample composition. Controlling for sociodemographic characteristics, such as age, level of educational attainment, and migration background, the web mode had 1.2% more missing values than the face-to-face mode in Rounds 9 and 11, and the paper mode had 2.5% more missing values (see Figure 6). Both effects were statistically significant ($p < .001$).

We also compared the share of missing values on questions that might be perceived as sensitive, namely the questions on voting decisions and the net household income. Regarding voting decisions, we tested two questions, asking respondents which candidate they had given their first vote to and which party

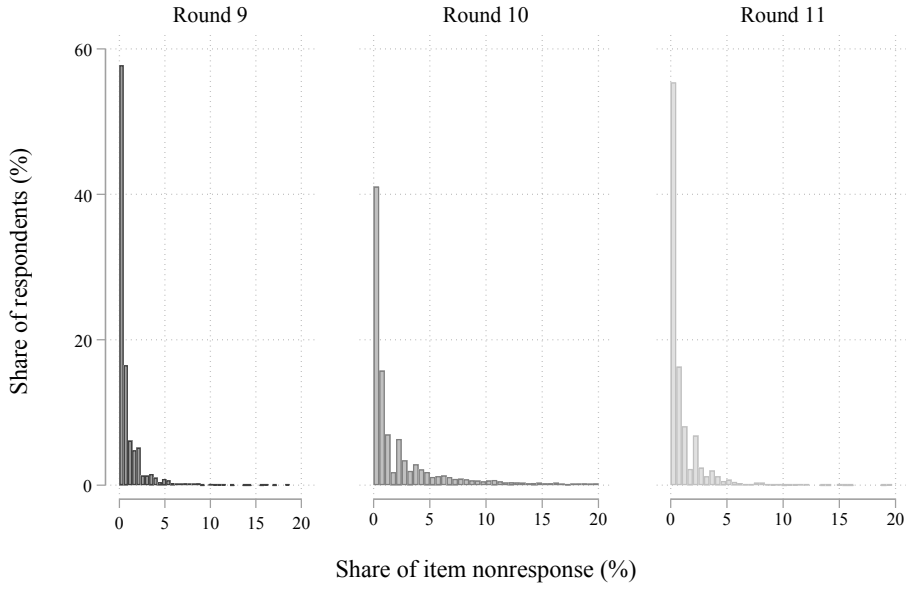

Figure 5 Density plot of the share of item nonresponse in ESS core questionnaire in Rounds 9, 10, and 11

they had given their second vote to in the last federal election. While the rate of nonresponse on the question about the first vote was equal in Round 9 and Round 10 (12.2%), it significantly increased to 13.3% in Round 11 ($p < .001$). In contrast, the share of respondents who did not answer the question about the second vote was higher in the self-administered round (12.3%) than in the interviewer-administered Rounds 9 (9.9%) and 11 (11.2%). Given that survey mode was not statistically significant once sociodemographic variables were controlled for, we conclude that the difference in the rate of nonresponse on the question about the second vote was not driven by the mode switch.

The level of nonresponse on the question about net household income was even higher, even though the ESS asks only about income brackets. In Round 9, 11.5% of respondents did not answer this question; in Round 11, this was the case for 10.4% of all respondents. In Round 10, this share increased to 23.4%, a statistically significant increase compared to the two face-to-face rounds ($p < .001$, not controlling for sociodemographic variables). The difference also remains statistically significant when controlling for differences in the sample composition ($p < .001$).

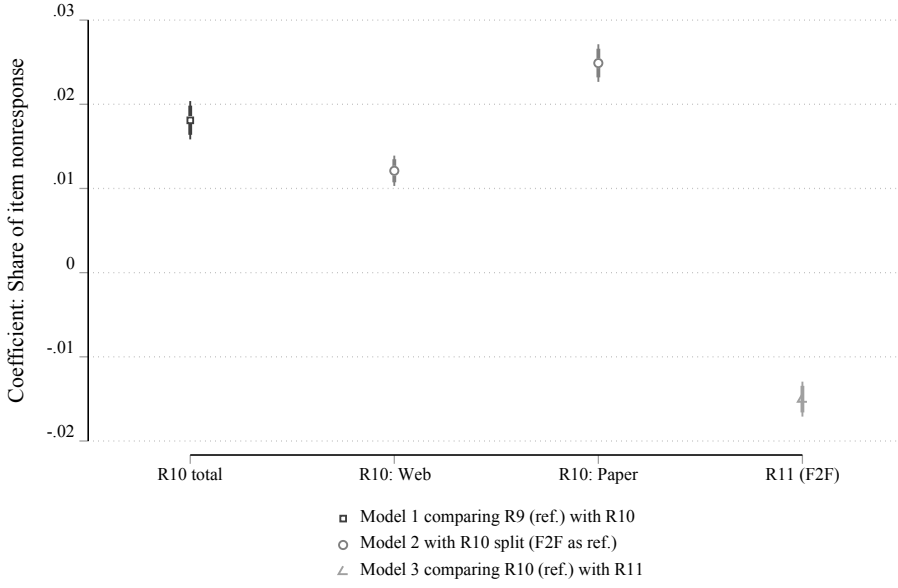

Figure 6 Multivariate OLS regression coefficients for the share of missing values on ESS core questionnaires across ESS rounds 9, 10, and 11 and within Round 10

Notes: Each model controls for sociodemographic characteristics (age group, sex, level of educational attainment, migration background, living in eastern or western Germany, being in paid work, being retired, frequency of internet usage, and self-rated urbanicity of the place of residence)

Looking at patterns of response on 39 ESS questions fielded in all three rounds, respondents tended to select the first response option more often in the self-administered modes (Round 10) than in the face-to-face mode (Rounds 9 and 11; see Figure 7). This mode effect was not driven by differences in the sociodemographic composition of the samples, and it did not differ between the web-based and paper-based modes. However, the share of ticks on the last response option (recency effect) did not differ significantly between the ESS Rounds 9 and 10, and in Round 11 the last response option was even selected slightly less often than in the previous self-completion round ($p < .05$). This suggests that the more visual mode – self-completion – did not lead to a strong adaptation of the response behavior. As mentioned above, this result might also be driven by the ESS face-to-face design, where interviewers must use many showcards during the interview.

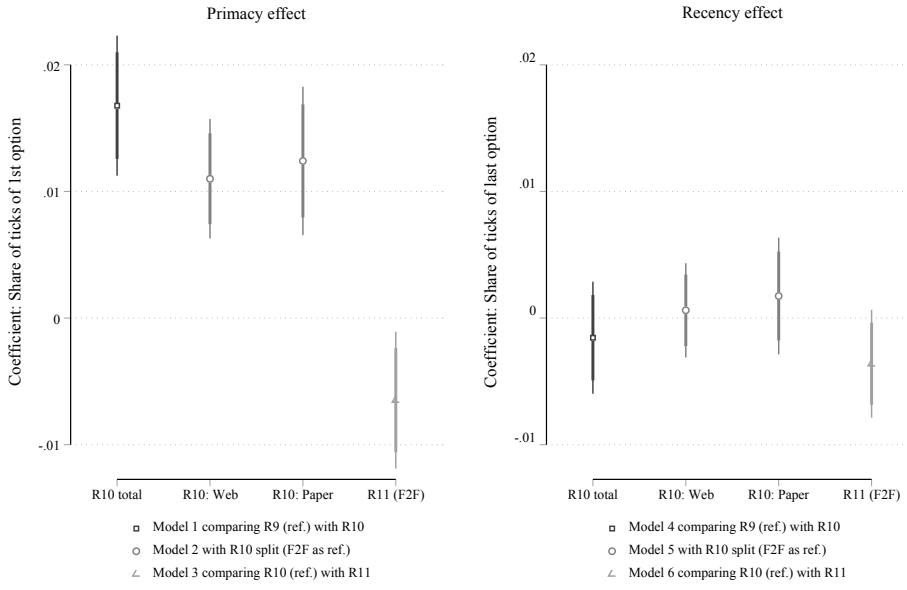

Figure 7 Multivariate OLS Linear regression coefficients comparing patterns of response on 39 ESS questions across ESS rounds 9, 10, and 11 and within Round 10

Notes: Each model controls for sociodemographic characteristics (age group, sex, level of educational attainment, migration background, living in eastern or western Germany, being in paid work, being retired, frequency of internet usage, and self-rated urbanicity of the place of residence)

Finally, we compared the odds for straightlining in any of three ESS item batteries fielded in all three ESS rounds. Straightlining occurred more often in the self-administered survey (Round 10) than in the interviewer-administered surveys (Rounds 9 and 11). In Round 9, 1.4% of all respondents used straightlining in any of the three item batteries. In Round 10, this share increased to 3.7%, which was statistically significant even when sociodemographic differences in the sample composition were controlled for ($p < .001$). In Round 11, the share of straightlining declined significantly to 1.6% ($p < .001$). Within Round 10, straightlining was less pronounced in the web-based questionnaire (2.7%) than in the paper-based questionnaire (5.7%; see Figure 8). This mode effect was statistically significant even when sample composition was controlled for ($p < .001$).

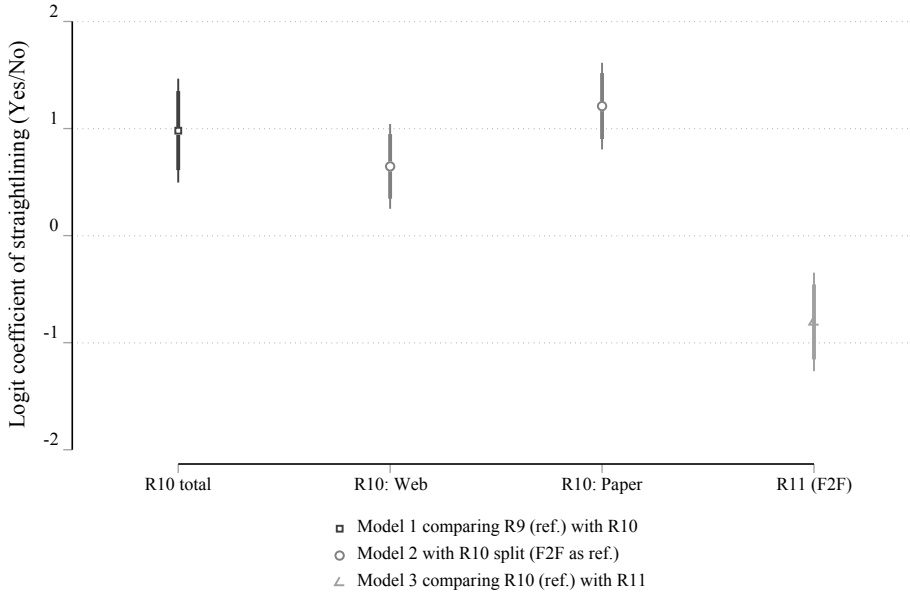

Figure 8 Multivariate logit model for binary straightlining indicator (straightlining in any of three ESS item batteries = 1) in ESS rounds 9, 10, and 11 and within Round 10

Notes: Each model controlling for sociodemographic characteristics (age group, sex, level of educational attainment, migration background, living in eastern or western Germany, being in paid work, being retired, frequency of internet usage, and self-rated urbanicity of the place of residence)

Addressing our research question on whether the mode of data collection affects response behavior (RQ4), we found a higher share of item nonresponse, a slightly higher share of ticks on the first answer option (primacy effect), and more straightlining in the self-administered modes (Round 10). However, despite a higher share of undesired response patterns, most Round 10 respondents did not use straightlining. In other words, we consider the mode differences to be rather small.

Discussion

In the present paper, we analyzed the implications of a mode switch on data quality in a large general population survey. For this purpose, we compared the German ESS Round 10 data collected in self-administered modes (web/paper) in 2021/22 with German ESS data collected in the traditional face-to-face mode in 2018/19 and 2023. We examined the overall and regional response rates; bias

in the sample composition and how it differed compared with official statistics (German Microcensus); and several indicators of response patterns.

Answering our first research question (RQ1), the German ESS achieved a much higher response rate in Round 10 without interviewers than in the previous and subsequent rounds with interviewers. Response rates differed at the level of municipalities in all three rounds; the spread of response rates across municipalities was greater in the interviewer-administered mode (RQ2). With an ever-smaller pool of interviewers, survey agencies have increasing problems covering all regions in Germany equally. This effect disappears when target persons can fill out the questionnaire themselves.

Examining bias in the sample composition to answer our third research question (RQ3), we found small differences across the three rounds. The youngest age group was slightly better represented in the self-administered modes than in the face-to-face mode, and the underrepresentation of marginalized groups – namely people with a low level of educational attainment, people without German citizenship, and unemployed persons – increased when respondents were asked to complete the survey themselves. However, we also observe that the underrepresentation of people with a low level of education did not improve in the face-to-face Round 11 after the mode switch. We assume that this can be explained with a declining fieldwork quality, namely a declining response rate, higher noncontact rate, and unequal regional response rates. People with a lower level of education seem to require a higher interviewer effort, which cannot be fully maintained with the given interviewer capacity in Germany. Overall, the differences in bias across rounds were small. In line with previous research, web-based respondents were younger, more educated, and more often with a migration background than those respondents choosing the paper-based mode in Round 10.

As a last indicator of data quality, we compared response patterns (RQ4). Like earlier research, we found a higher level of item nonresponse in the self-completion round than in the face-to-face rounds. In the self-administered mode (Round 10), levels of primacy effects, and straightlining were slightly higher than in the face-to-face mode (Round 9 and Round 11). Again, those findings mirror those of an earlier study (Cernat et al., 2022). As in the case of differences in the sample composition, the overall mode effect was rather modest, especially considering the radical change in the survey design in Round 10. Considering the quality indicators analyzed in this article self-completion surveys can meet the high-quality standards of face-to-face surveys or – in the case of response rates – even outperform interviewer-administered surveys. Therefore, self-completion surveys have the potential to replace face-to-face surveys in Germany.

Our analyses have several limitations. First and foremost, our mode comparison was not based on a parallel run – that is, the self-completion data and face-to-face data were not collected at the same time. Therefore, any changes

in response rates and sample composition might also have been caused by the eventful time that passed between the three ESS rounds, even though a comparative analysis of ESS data suggested that – compared to mode differences – "time and the Covid pandemic had minimal effects on measurement" (Lugtig, 2024b, p. 19). We can provide further evidence for this by analyzing face-to-face Round 11 two years after the mode switch. Since most of the results of this face-to-face round are similar to those six years earlier, the differences in Round 10 appear to be attributable to the mode switch.

Second, it goes without saying that survey data are the result of many major and minor decisions made during the design phase, and those decisions hamper the internal validity of our conclusions. For instance, for ESS Round 10, the questionnaire was shortened by about 10%, highly salient questions on the COVID-19 pandemic were added, and all target persons received a prepaid cash incentive. Internal validity is strengthened by comparing the self-administered round with two face-to-face rounds. Even though the two face-to-face rounds differ slightly in design (e.g., different survey agency, different incentive scheme, see Table 1), we draw similar conclusions from the comparisons with the self-administered round. The fact that every survey program has its own unique features also limits the external validity of our results. Therefore, similar analyses in other countries and with other survey data would be highly welcome. In 2025/26 (ESS Round 12), the ESS is collecting data in both face-to-face and self-administered modes in all participating countries, hopefully encouraging further research and verifying and complementing the findings of this study.

When drawing a conclusion on whether the face-to-face mode or the self-administered modes have an edge in the context of ESS in Germany, we also need to consider differences in the costs and the process of data collection. Fieldwork ran for more than 25 weeks in Round 9 and more than 30 weeks in Round 11, whereas it took only half as long in Round 10 and was concluded after 13 weeks. Moreover, costs differed greatly: In Round 10, a valid interview was 78% less expensive than in Round 9 and 76% less expensive than in Round 11, including all fixed and variable costs. Thus, if we consider the significantly higher response rate and the small differences in sample composition and response behavior, the return on investment of the self-administered ESS (Round 10) was significantly higher. On the other hand, however, marginalized groups were even less well represented in Round 10 than in Rounds 9 and 11, where interviewers likely helped to lower the barriers to participation for these groups. Future studies should therefore investigate ways to improve the representation of these groups in self-completion surveys.

Data and code availability statement

The data analyzed in this research article is publicly available and can be downloaded from the ESS Data Portal (European Social Survey European Research Infrastructure [ESS ERIC], 2021, 2022, 2025) and the database of the Federal Statistical Office of Germany (Statistisches Bundesamt, 2026). The Stata code used to prepare and analyze ESS and Microcensus data is available as part of the supplementary material published alongside this article.

Conflicts of interest

The authors declare no conflicts of interest.

Funding information

The authors did not receive additional funding for their work on this research article.

References

American Association for Public Opinion Research (AAPOR). (2023). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (10th ed.). https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf

Ackermann-Piek, D.& Massing, N. (2014). Interviewer behavior and interviewer characteristics in PIAAC Germany. *Methods, Data, Analyses, 8*(2), 199–222. https://doi.org/10.12758/mda.2014.008

Andreadis, I., & Kartsounidou, E. (2020). The impact of splitting a long online questionnaire on data quality. *Survey Research Methods*, *14*(1), 31–42. https://doi.org/10.18148/srm/2020.v14i1.7294

Baumann, H., Quoß, F., Siegers, P., Kathage, L., Engelking, L. (2025). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUScompact 2023. GESIS-Variable Reports 2025|01, GESIS – Leibniz-Institut für Sozialwissenschaften, Köln. https://access.gesis.org/dbk/78534

Bautista, R. (2021). Conducting the General Social Survey (GSS) during a pandemic—Interview with René Bautista. *Survey Practice*. https://doi.org/10.29115/SP-2021-0007

Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). Response rates in the European Social Survey: Increasing, decreasing, or a matter of fieldwork efforts? *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2018-00003

Beullens, K. & Loosveldt, G. (2016). Interviewer effects in the European Social Survey. *Survey Research Methods, 10*(2), 103–118. https://doi.org/10.18148/srm/2016.v10i2.6261

Biemer, P. P., Murphy, J., Zimmer, S., Berry, C., Deng, G., & Lewis, K. (2018). Using bonus monetary incentives to encourage web response in mixed-mode household surveys. *Journal of Survey Statistics and Methodology, 6*(2), 240–261. https://doi.org/10.1093/JSSAM/SMX015

Čehovin, G., Bosnjak, M., & Lozar Manfreda, K. (2023). Item nonresponse in web versus other survey modes: A systematic review and meta-analysis. *Social Science Computer Review, 41*(3), 926–945. https://doi.org/10.1177/08944393211056229

Cernat, A., Sakshaug, J., Christmann, P., & Gummer, T. (2022). The impact of survey mode design and questionnaire length on measurement quality. *Sociological Methods and Research, 53(4), 1873-1904*. https://doi.org/10.1177/00491241221140139

Cornesse, C., & Bosnjak, M. (2018). Is there an association between survey characteristics and representativeness? A meta-analysis. *Survey Research Methods, 12*(1), 1–13. https://doi.org/10.18148/srm/2018.v12i1.7205

Couper, M. P. (2017). New developments in survey data collection. *Annual Review of Sociology, 43*, 121–145. https://doi.org/10.1146/annurev-soc-060116-053613

Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. In P. B. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, C. Tucker, & B. T. West (Eds.), *Total survey error in practice* (pp. 133–154). John Wiley & Sons Inc. https://doi.org/10.1002/9781119041702.CH7

Couper, M. P. & McGonagle, K. (2019). Recent developments in web-based data collection for longitudinal studies. *Panel Study of Income Dynamics Technical Paper*, 19-03. http://dx.doi.org/10.2139/ssrn.6077968

Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2020). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology, 8*(3), 513–539. https://doi.org/10.1093/JSSAM/SMZ008

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., Hox, J., & Luiten, A. (2018). International nonresponse trends across countries and years: An analysis of 36 years of Labour Force Survey data. *Survey Insights: Methods from the Field*. https://doi.org/10.13094/SMIF-2018-00008

Díaz de Rada, V. (2022). Concurrent mixed modes: Response quality, speed, and cost. *Field Methods, 34*(3), 191–205. https://doi.org/10.1177/1525822X221105917

Dillman, D., Brown, T., Carlson, J., Carpenter, E., Lorenz, F., Mason, R., Saltiel, J. Songster, R. Effects of category order on answers in mail and telephone surveys. *Rural Sociology, 60*(4), 674–687. https://doi.org/10.1111/j.1549-0831.1995.tb00600.x

Dillman, D. A. (2000). *Mail and internet surveys: The Tailored Design Method* (2nd ed). John Wiley and Sons.

Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology, 43*(1), 3–30. https://www150.statcan.gc.ca/n1/pub/12-001-x/2017001/article/14836-eng.pdf

European Social Survey European Research Infrastructure (ESS ERIC). (2021). *European Social Survey: ESS9 – Integrated data file, edition 3.1*. Sikt – Norwegian Agency for Shared Services in Education and Research. https:/doi.org/10.21338/ess9e03_2

European Social Survey European Research Infrastructure (ESS ERIC). (2022). *European Social Survey: ESS10 – Integrated data file, edition 1.0*. Sikt – Norwegian Agency for Shared Services in Education and Research. https://doi.org/https://doi.org/10.21338/ess10sce01_0

European Social Survey European Research Infrastructure (ESS ERIC). (2025). *ESS11 - integrated file, edition 4.0*. Sikt – Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess11e04_0

European Social Survey European Research Infrastructure (ESS ERIC). (2025). *ESS11 - 2023 Country Documentation Report, edition 4.0*. Sikt – Norwegian Agency for Shared Services in Education and Research. https://stessrelpubprodwe.blob.core.windows.net/data/round11/survey/ESS11_country_documentation_report_e04.pdf

Eurostat (2023). *Digital economy and society: ICT usage in households and by individuals*. Retrieved April 20, 2024 from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_households_and_individuals#Devices_used_to_connect_to_the_internet

Galesic, M., Tourangeau, R., Couper, M., Conrad, F. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly, 72*(5), 892–913. https://doi.org/10.1093/poq/nfn059

Ghirelli, N., Lynn, P., Dorer, B., Schwarz, H., Kappelhof, J., van de Maat, J., Kessler, G., Briceno-Rosas, R., & Rød, L-M. (2022). ESS9 overall fieldwork and data quality report, GESIS.

Groves, R., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Groves, R. M., Fowler, F. J., Couper, M. P., Jr., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley

Gummer, T., Schmiedeberg, C., Bujard, M., Christmann, P., Hank, K., Kunz, T., Lück, D., & Neyer, F. J. (2020). The impact of Covid-19 on fieldwork efforts and planning in pairfam and FReDA-GGS. *Survey Research Methods, 14*(2), 223–227. https://doi.org/10.18148/SRM/2020.V14I2.7740

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, *21*(1), 111–121. https://doi.org/10.1093/IJPOR/EDN054

Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public Opinion Quarterly*, *72*(5), 836–846. https://doi.org/10.1093/POQ/NFN045

Holbrook, A., Krosnick, J., Moore, D., Tourangeau, R. (2007) Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes. *Public Opinion Quarterly, 71*(3), 325–348, https://doi.org/10.1093/poq/nfm024

Lugtig, P., Toepoel, V., Emery, T., Cabaço, S. L. F., Bujard, M., Naderi, R., Schumann, A., & Lück, D. (2022). *Can we successfully move a cross-national survey online? Results from a large three-country experiment in the Gender and Generations Programme survey.* OSF. https://doi.org/10.31235/OSF.IO/MU8JY

Lugtig, P. (2024a). ESS round 10 mode experiments in Great Britain and Finland. Findings on mode effects. ESS ERIC. Online: https://www.europeansocialsurvey.org/sites/default/files/2024-10/round-10-experimental-comparison-final.pdf

Lugtig, P. (2024b). Comparison of ESS rounds 9 and 10: mode differences. ESS ERIC. Online: https://www.europeansocialsurvey.org/sites/default/files/2024-10/round-9-10-comparison-final.pdf

Luijkx, R., Jónsdóttir, G. A., Gummer, T., Ernst Stähli, M., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S. P., Hjaltason, Á. B., Joye, D., Lomazzi, V., Maineri, A. M. , Milbert, P., Ochsner, M., Pollien, A., Sapin,

M., Solanes, I., ... Wolf, C. (2021). The European Values Study 2017: On the way to the future using mixed-modes. *European Sociological Review, 37*(2), 330–346. https://doi.org/10.1093/ESR/JCAA049

Luiten, A., Hox, J., & De Leeuw, E. (2020). Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys. *Journal of Official Statistics, 36*(3), 469–487. https://doi.org/10.2478/jos-2020-0025

Lund Clement, S., & Severin-Nielsen, M. K. (2020). Device effects on survey response quality. A comparison of smartphone, tablet and PC responses on a cross sectional probability sample. *Survey Methods: Insights from the Field, Special Issue: "Advancements in Online and Mobile Survey Methods."* https://doi.org/10.13094/SMIF-2020-00020

Lynn, P. & Xena, C. (2023). A comparison of non-response patterns over time in countries that switched to self-completion at Round 10. ESS ERIC. Online: https://www.europeansocialsurvey.org/sites/default/files/2024-09/ESS_Sample_Composition_report_sc_Round_10.pdf

Maslovskaya, O., Durrant, G. B., Smith, P. W., Hanson, T., & Villar, A. (2019). What are the characteristics of respondents using different devices in mixed-device online surveys? Evidence from six UK surveys. *International Statistical Review, 87*(2), 326–346. http://dx.doi.org/10.1111/insr.12311

Meitinger, K., Stadtmüller, S., Silber H., et al. (2020). Fieldwork monitoring in practice: Insights from 17 large-scale social science surveys in Germany. *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2020-00006

Messer, B. L., & Dillman, D. A. (2011). Surveying the general public over the internet using address-based sampling and mail contact procedures. *Public Opinion Quarterly, 75*(3), 429–457. https://doi.org/10.1093/POQ/NFR021

Peytchev, A., Peytcheva, E., Conzelmann, J.G., Wilson, A., & Wine, J. (2019), Modular survey design: Experimental manipulation of survey length and monetary incentive structure. *Journal of Survey Statistics and Methodology, 8*(2), 370-384. https://doi.org/10.1093/jssam/smz006

Pforr, K., Blohm, M., Blom, A. G., Erdel, B., Felderer, B., Fräßdorf, M., Hajek, K., Helmschrott, S., Kleinert, C., Koch, A., Krieger, U., Kroh, M., Martin, S., Saßenroth, D., Schmiedeberg, C., Trüdinger, E. M., & Rammstedt, B. (2015). Are incentive effects on response rates and nonresponse bias in large-scale, face-to-face surveys generalizable to Germany? Evidence from ten experiments. *Public Opinion Quarterly, 79*(3), 740–768. https://doi.org/10.1093/POQ/NFV014

Roberts, C., Joye, D., & Ernst Stähli, M. (2016). *Mixing modes of data collection in Swiss social surveys: Methodological report of the LIVES-FORS mixed mode experiment* (LIVES Working Papers 2016/48). Swiss National Centre of Competence in Research. https://doi.org/10.12682/lives.2296-1658.2016.48

Roßmann, J. (2022). Data Brief: Die German Longitudinal Election Study (GLES) zur Bundestagswahl 2021. *Easy Social Sciences, 67*, 72–77. https://doi.org/https://doi.org/10.15464/easy.2022.09

Rybak, A. (2023). Survey mode and nonresponse bias: A meta-analysis based on the data from the international social survey programme waves 1996-2018 and the European social survey rounds 1 to 9. *PLoS ONE, 18*(3), e0283092. https://doi.org/10.1371/journal.pone.0283092

Scherpenzeel, A., Axt, K., Bergmann, M., Douhou, S., Oepen, A., Sand, G., Schuller, K., Stuck, S., Wagner, M., & Börsch-Supan, A. (2020). Collecting survey data among the 50+ population during the COVID-19 outbreak: The Survey of Health, Ageing and

Retirement in Europe (SHARE). *Survey Research Methods, 14*(2), 217–221. https://doi.org/10.18148/SRM/2020.V14I2.7738

Statistisches Bundesamt (2026). GENESIS-online, the Database of the Federal Statistical Office: Microcensus (Code 12211). Online: https://www-genesis.destatis.de/

Stein, A., Gummer, T., Naumann, E., Rohr, B., Silber, H., Auriga, R., Bergmann, M., Bethmann, A., Blohm, M., Cornesse, C., Christmann, P., Coban, M., Décieux, J., Gauly, B., Hahn, C., Helmschrott, S., Hochman, O., Lemcke, J., Naber, D., Pötzschke, S., Roßmann, J., Schanze, J., Schmidt, T., Schneider, S., Spangenberg, H., Rettig, T., Trappmann, M., Weinhardt, M., & Weiß, B. (2025). Education bias in probability-based surveys in Germany: Evidence and possible solutions. International Journal of Social Research Methodology, *29*(2), 197-214. https://doi.org/10.1080/13645579.2025.2508889.

Stoop, I. (2005). *The hunt for the last respondent: Nonresponse in sample surveys*. Social and Cultural Planning Office of the Netherlands.

Toepoel, V., & Lugtig, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of Internet users. *Social Science Computer Review, 32*(4), 544–560. https://doi.org/10.1177/0894439313510482

Toepoel, V., & Lugtig, P. (2022). Modularization in an era of mobile web: Investigating the effects of cutting a survey into smaller pieces on data quality. *Social Science Computer Review, 40*(1), 150–164. https://doi.org/10.1177/0894439318784882

Villar, A., & Fitzgerald, R. (2017). Using mixed modes in survey data research: Results from six experiments. In M. J. Breen (Ed.), *Values and identities in Europe: Evidence from the European Social Survey* (pp. 273–310). Routledge.

West, B. & Blom, A. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175–211. https://doi.org/10.1093/jssam/smw024

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting general social surveys as self-administered mixed-mode surveys. *Public Opinion Quarterly, 85*(2), 623–648. https://doi.org/10.1093/POQ/NFAB039

Appendix

Table A1 Response rates (AAPOR RR2) at the level of federal states

	Round 9 (%)	Round 10 (%)	Round 11 (%)
Baden-Württemberg	24.2	37.3***	26.3***
Bavaria	29.5	37.5***	28.0***
Berlin	23.0	35.7***	20.9***
Brandenburg	25.0	35.8**	22.4***
Bremen	18.2	39.2*	3.3***
Hamburg	30.7	37.9	19.6***
Hesse	23.5	34.7***	26.7***
Lower Saxony	33.0	39.3**	22.2***
Mecklenburg-Vorpommern	30.7	40.6*	24.6***
North Rhine-Westphalia	25.8	33.5***	30.5
Rhineland-Palatinate	25.8	37.7***	24.9***
Saarland	30.7	35.8	18.8***
Saxony	31.4	41.1***	26.7***
Saxony-Anhalt	25.4	28.5	27.4
Schleswig Holstein	27.6	40.0***	27.2***
Thuringia	27.7	35.0*	25.7*
SD	3.9	3.1	6.3

Notes: Tests of statistical significance conducted with two-sample tests of proportions. Round 10 with Round 9 as reference, Round 11 with Round 10 as reference. *$p < .05$; **$p < .01$; ***$p < .001$.

Table A2 Migration background across ESS rounds 9, 10, and 11 and across modes within Round 10

	Microcensus 2021 (%)	Round 9 (%)	Round 10 (%)	Paper (%)	Web (%)	Round 11 (%)
Foreign citizenship with own migration	11.4	6.3***	4.6***	2.9***	5.5***	6.5***
Foreign citizenship without own migration	1.4	0.8*	0.7***	0.3***	0.9***	0.8*
German citizenship with own migration	7.1	5.7**	5.2***	4.9***	5.3***	5.6**
Parents born abroad, German citizenship without own migration	5.2	8.9***	9.1***	6.4**	10.5***	12.3***
No migration background	74.8	78.2***	80.5***	85.6***	77.6***	74.9
N	70,636,000	2,338	8,463	2,857	5,606	2,405

Notes: Tests of statistical significance conducted with two-sample tests of proportions $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

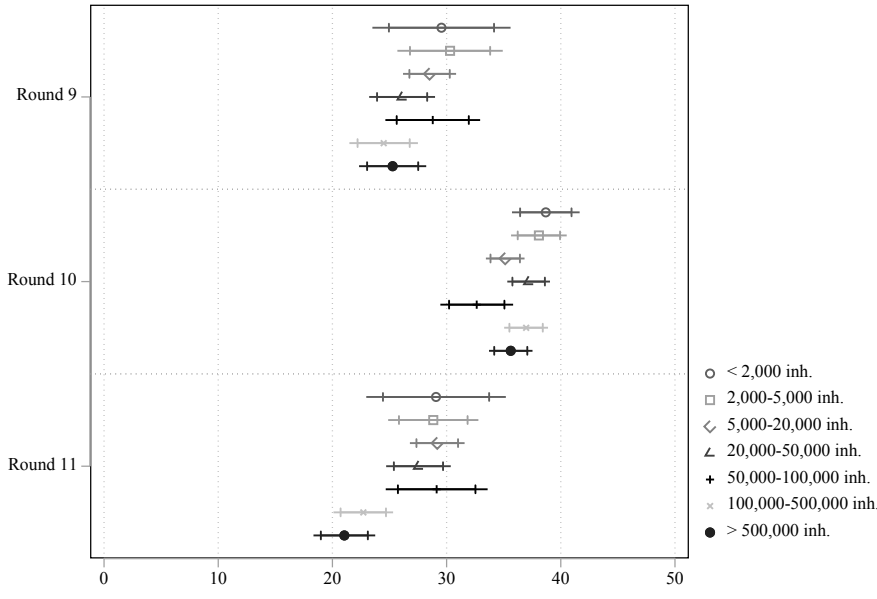

Figure A1 Comparison of response rates by size of municipalities in ESS rounds 9, 10, and 11

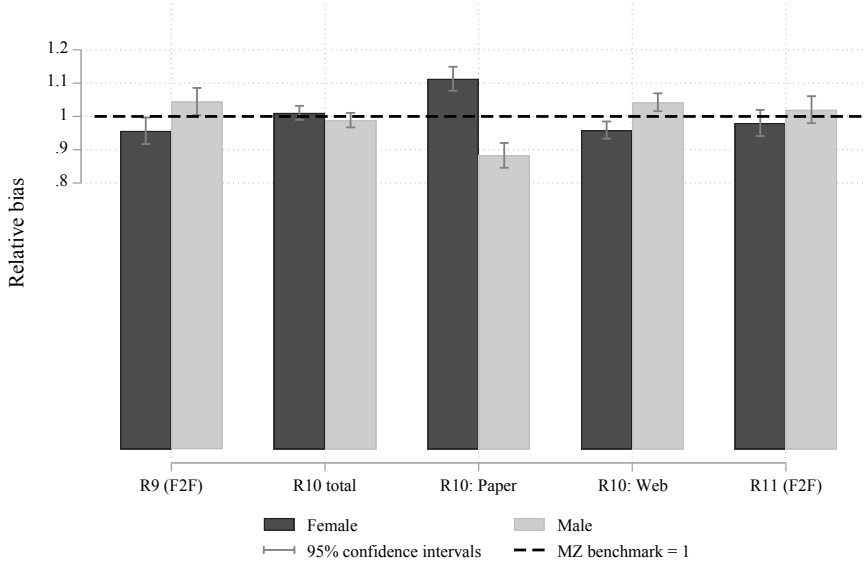

Figure A2 Comparison of the relative bias in respondents' sex in ESS rounds 9, 10, and 11 with the German Microcensus (MZ) benchmark (=1)

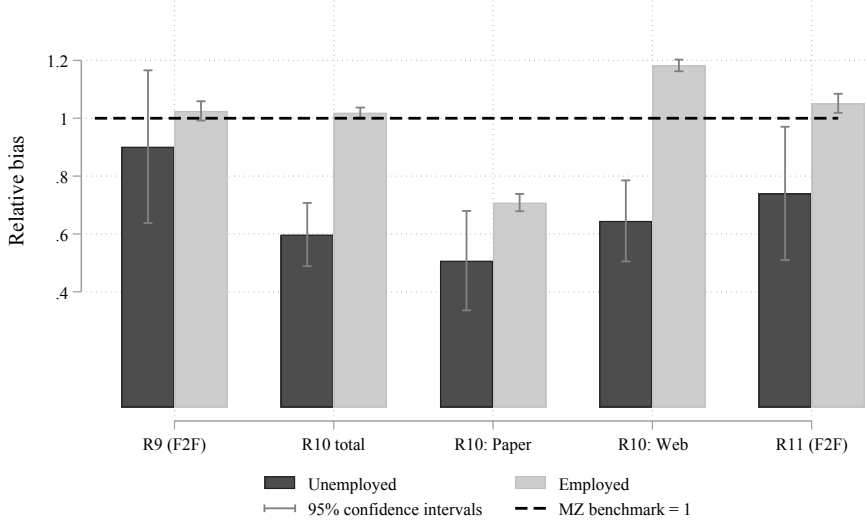

Figure A3 Comparison of the relative bias in respondents' labor force status in ESS rounds 9, 10, and 11 with the German Microcensus 2021 (MZ) benchmark (=1)

Online First publication date: 2026-07-14