

Secondary Publication



Prücklmair, Franz; Rendtel, Ulrich

Alternative selection mechanisms in online surveys

Date of secondary publication: 12.02.2026

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-113134x

Primary publication

Prücklmair, Franz; Rendtel, Ulrich (2025): Alternative selection mechanisms in online surveys, in: Wirtschafts- und sozialstatistisches Archiv : ASTA ; eine Zeitschrift der Deutschen Statistischen Gesellschaft, Berlin ; Heidelberg: Springer, Vol. 19, No. 3–4, pp. 165–193, doi: 10.1007/s11943-025-00362-8.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>



Alternative selection mechanisms in online surveys

Franz Prücklmair  · Ulrich Rendtel

Received: 17 February 2025 / Accepted: 7 November 2025 / Published online: 3 December 2025
© The Author(s) 2025

Abstract The rapid changes in internet accessibility and the evolving ways it is used could raise doubts about whether earlier findings on selection effects in internet surveys still hold true. We address the following questions: 1) Is the selection criterion by mere internet access still a reasonable self-selection criterion? 2) If this is no longer true, how should alternative self-selection processes be modeled? 3) How can these selection processes be controlled? Are demographic control variables sufficient to establish the Missing at Random (MAR) condition, or is it possible to establish the MAR condition with more powerful control variables? 4) To what extent do the weighting procedures correct a self-selection bias? We investigate these questions in the setting of a simulation study where we assume four different selection models. These involve the length of internet use, posting behavior, and interest in politics and are based on theoretical considerations. We use the European Social Survey (ESS) as a simulation environment, which contains these variables and demographic background variables. It also includes our outcome variable, the vote in the 2017 Bundestag election in Germany. In order to judge the differences of the non-probability results and the simulated universe we compare the results of ESS estimates of the 2017 Bundestags elections the real election results.

Keywords Online surveys · River Sampling · Non-probability samples · Quasi-Randomization approach · Calibration · European Social Survey

✉ Franz Prücklmair
Otto-Friedrich-Universität Bamberg, Bamberg, Germany
E-Mail: franz.pruecklmair@uni-bamberg.de

Ulrich Rendtel
Freie Universität Berlin, Berlin, Germany
E-Mail: ulrich.rendtel@fu-berlin.de

Alternative Auswahlmechanismen in Online-Umfragen

Zusammenfassung Die rasanten Veränderungen beim Internetzugang und in der Art seiner Nutzung werfen die Frage auf, ob frühere Erkenntnisse zu Selektionseffekten in Internetbefragungen weiterhin gültig sind. In diesem Zusammenhang untersuchen wir die folgenden Fragen: 1) Ist das Auswahlkriterium des bloßen Internetzugangs noch ein sinnvolles Selbstselektionskriterium? 2) Wenn nicht, wie sollten alternative Selbstselektionsprozesse modelliert werden? 3) Wie können diese Selektionsprozesse kontrolliert werden? Sind demografische Kontrollvariablen ausreichend, um die „Missing at Random“ (MAR)-Bedingung zu erfüllen, oder sollten stärkere Kontrollvariablen einbezogen werden? 4) In welchem Maße korrigieren Gewichtungungsverfahren einen Selbstselektionsbias? Diese Fragen werden im Rahmen einer Simulationsstudie untersucht, in der vier unterschiedliche Selektionsmodelle angenommen werden. Diese beruhen auf theoretischen Überlegungen und beziehen sich auf die Dauer der Internetnutzung, das Posting-Verhalten und das politische Interesse. Als Simulationsumgebung dient der European Social Survey (ESS), der diese Informationen, demografische Hintergrundvariablen sowie die Zielvariable – die Wahlentscheidung bei der Bundestagswahl 2017 – enthält. Zur Bewertung der Abweichungen zwischen den nicht-probabilistischen Ergebnissen und dem simulierten Universum vergleichen wir die ESS-Schätzungen der Bundestagswahl 2017 mit den tatsächlichen Wahlergebnissen.

Schlüsselwörter Online-Umfragen · River Sampling · Nicht-Zufallsstichproben · Quasi-Randomisierung · Kalibrierung · European Social Survey

1 Introduction

Online samples can take various forms, but they all share the common feature that participants are contacted, or responses are collected, via the internet at some point, see Callegaro et al. (2015) and Vehovar et al. (2016). Persons may first be selected through a probability sample and are then asked to complete an online questionnaire. See Amarov and Rendtel (2013), Enderle et al. (2013), and Rendtel and Amarov (2014) for applications within the framework of the German Microcensus, or consider the German Internet Panel of the University of Mannheim (Blom et al. 2015). However, there are approaches where internet users are asked by some widgets to participate in some voting experiment, e.g. how they judge an actual event in an online newspaper. Those who participate in the voting experiment are then asked for further participation in a survey. This kind of selection is called “River Sampling” (AAPOR 2010).

Despite uncontrolled self-selection effects (Bethlehem 2010), online surveys via River Sampling have become popular survey instruments as they deliver cheap and fast results. Moreover Elliot and Valliant (2017) have produced the framework of Quasi-Randomization (QR), which under regularity conditions, allows an unbiased estimation of population parameters from non-probability surveys. For a given set of variables, this approach compares the distribution of a non-probability sample

with the distribution of a probability sample. If these variables explain the self-selection mechanism of the non-probability sample, the weighted results from the QR-approach deliver valid population estimates. The draw-back of this approach is the unobserved selection process, as non-respondents are not recorded in the River Sampling procedure. In practice, one uses demographic variables, like age groups, gender, education and regional indicators which are known for the non-probability and the probability survey. This makes the computation of QR-weights feasible.

In order to evaluate the approach, one has to use a simulation strategy where the self-selection from the population is repeatedly simulated. Hence, we need a survey which contains information on internet use, on demographic variables and on the outcome variables which are of interest. By using variables on internet use we aim to generate a plausible selection process, which yields an artificial self-selected sample. Finally, we compare the results of the outcome variable from the original sample and the self-selected sample. Differences indicate how well the QR-weights reduce possible self-selection effects.

Schnell et al. (2017) were the first authors who used the European Social Survey (ESS) for such an evaluation. Their main outcome variable is a reported health status. For the self-selection process they generated a dummy variable, which assumes self-selection if the person has access to the internet and uses it. They used ESS data from 2010. They conclude:

“Weighting by calibration on age, gender, ethnic background, urban residence, education and household income does not eliminate the observed health differences. Therefore, the underlying missing data mechanism might be considered as an example of MNAR¹ If this holds, no weighting strategy will be able to eliminate health bias in web surveys”. (Schnell et al. 2017).

However, internet access and its use has dramatically changed since 2010. Now, it is present in almost every household in Germany (Qualitätsbericht EVS 2018; Statistisches Bundesamt 2022)² and its use is no longer restricted to PC's. Easy-to-use apps on smartphones can be utilized almost everywhere in public. Additionally, distinction according the general definition of the so-called “digital divide”: “a division between people who have access and use of digital media and those who do not” (Van Dijk 2019), is undergoing change. Therefore, we expect that results related to selection effects in online surveys from 2010 cannot be transferred to more recent years.

This article aims to formulate more adequate models on current internet use. This includes individual usage styles, like posting on social media, and motivational factors (Bethlehem 2010). As a framework we use the ESS data from 2018³. For

¹ *Missing not at random.*

² Note: The data originate from the Income and Consumption Survey (EVS), a quota sample with quotas based on the 2016 Microcensus. Therefore, a standard error cannot be reported. However, it can be assumed that the error margins of the quota sample approximately correspond to those of a stratified random sample. According to Destatis, the unmarked results presented here are fully publishable without restrictions (*Qualitätsbericht EVS 2018*).

³ This was the latest wave at the start of our analyses.

bias correction we use the QR-approach and calibration to population totals which is common among survey practitioners. We use the votes for political parties in the previous general election of the German Bundestag 2017 as outcome variable and analyze how well the bias-corrected party percentages meet the simulated values of the ESS-universe. In order to scale these differences, we compare them to the differences of original ESS-estimates and the official election results. We will also discuss selection models where we include variables which are closer to the outcome variable than the standard demographic variables.

2 Selection models for self-selected online surveys

In the early days of the internet survey, statisticians were concerned about the proportion of the households with access to the internet. For example, in Germany this percentage was as low as 8.1% in 1998 (Statistisches Bundesamt 2022). An example provided by Valliant (2020) pertains to the Michigan Behavioral Risk Factor Surveillance Survey (BRFSS) conducted in 2003, which served as a basis for simulating selection via internet access. According to Valliant, the internet coverage rate for this survey was 60%. However, the internet coverage has dramatically risen in the last two decades. Within 10 years the internet coverage in Germany has risen from 8.1% to 64.4% (2008) and in 2018 to 92.7% of the German households.

While internet access has become apparently less important for participation, the duration of daily internet use can be a good indicator to measure the probability of coming across invitations to participate in an internet survey. The longer the duration, the higher the probability of coming across an interesting widget.

Early on, the internet was merely a tool for emails and reading pre-made HTML documents. However, over the years, the purposes of internet usage have become more diverse, see van Eimeren and Gerhard (2000), with activities such as “mainly reads online articles”, “tends to write emails”, or “plays online games”. Hargittai (2002) referred to disparities based on online skills as the “second-level digital divide”, which influences how people use the internet.

Nowadays, aspects of interactivity seem to be important for online survey participation. Multiple studies (Chang and Krosnick 2009; Dever et al. 2008; Malhotra and Krosnick 2007) have reported that active internet users tend to be over-represented in online panels. A good indicator of active use is the online posting behavior.

Motivational aspects may also be important for the self-selection (Bethlehem 2010). Chang and Krosnick (2009) found that political participation in the United States is approximately 10% higher in online samples compared to telephone or probability-based internet samples. Therefore, if the outcome variable is voting behaviour, a recent posting about politics could be a relevant motivation to follow an online survey about political party preference.

A different aspect of online surveys are incentives to participate. A traditional incentive is a payment. Often such payments are linked with the recruitment to an access panel, which cumulates the information of its respondents from the separate surveys (Amarov and Rendtel 2013). A different reward which is specific for online surveys is the immediate return of the voting of the other participants on the topic of

interest. For example, the German online survey agency Civey poses its widgets on daily emerging topics of relevance in online newspaper platforms, readers interesting in these topics are probably keen on learning about the views of others on the matter, see Richter et al. (2023) for details. Thus, it makes sense to use an indicator of interest, for example, interest in politics, in a participation model for online surveys. However, one should not ignore the risk of topical self-selection (Lehdonvirta et al. 2020), which is inherent to this approach. This suggests to use interest in the topic and curiosity about the opinion of others participants as a means to contact potential respondents. But one should refrain from using these responses as valid answers which are representative for the population. Therefore Civey does not use these first contact responses which are framed by an actual event in a newspaper, see Richter et al. (2023). Nevertheless, interest in the topic seems to be a reasonable indicator for participation in online surveys.

3 Weighting procedures for online samples

The use of survey weights is the standard procedure for the design-based estimation of population totals and proportions, see, for example, the textbook of Särndal et al. (1992) or Lohr (2023). In this section we shortly discuss two weighting approaches which can be also used for non-probability surveys. Elliot and Valliant (2017) designed the quasi-randomisation (QR)-approach especially for the estimation of population parameters from non-probability samples. Calibration has a long history in survey sampling, with fundamental contributions from Deming and Stephan (1940), Deville and Särndal (1992), Deville et al. (1993). It serves as a standard tool to correct nonresponse bias. A second important issue is the reduction of the variance of population estimates, see (Lundström and Särndal 2001) or the overview of Särndal (2007). In the context of non-probability samples, Yang et al. (2018) demonstrate an application where units come from an opt-in online sample.

Both procedures implicitly model selection probabilities such that the resulting survey weights apply to all outcome variables. This is an advantage over model-based approaches which predict the value of the outcome variable for the non-observed units, see Little and Rubin (2020) for a model-based treatment of missing values.

3.1 Quasi-randomisation approach

The Quasi-randomisation (QR)-approach compares for a set of variables denoted by \mathbf{X} the frequencies in a probability and a non-probability sample. The value of \mathbf{X} must be known for both samples. Let S be the sampling indicator for the probability (or reference) sample and S^* for the non-probability sample. The value of the outcome variable is Y .

The QR-approach delivers an estimate of the unknown inclusion probability of the non-probability sample under the assumption that it does not depend on the outcome variable:

$$P(S_i^* | Y_i, \mathbf{X}_i) = P(S_i^* | \mathbf{X}_i) \tag{1}$$

As we do not observe unselected or nonresponding units in the non-probability sample, Eq. 1 describes a typical missing at random **MAR** pattern in the sense of Rubin (1976).

Applying Bayes' rule, we have:

$$\begin{aligned} P(S_i^* = 1 | X = x_i) &= \frac{P(X = x_i | S_i^* = 1)P(S_i^* = 1)}{P(X = x_i)} \\ &= \frac{P(X = x_i | S_i^* = 1)P(S_i^* = 1)P(S_i = 1 | X = x_i)}{P(S_i = 1)P(X = x_i | S_i = 1)} \\ &\propto \frac{P(X = x_i | S_i^* = 1)P(S_i = 1 | X = x_i)}{P(X = x_i | S_i = 1)} \end{aligned} \tag{2}$$

Elliot and Valliant (2017) pool the reference and non-probability samples into a single dataset and use an indicator Z_i for the membership to the reference sample ($Z_i = 0$) or the non-probability sample ($Z_i = 1$). Implicitly they assume that the probability for a joint membership in both surveys is so small, that it may be ignored. Thus we obtain:

$$\begin{aligned} \frac{P(X = x_i | S_i^* = 1)}{P(X = x_i | S_i = 1)} &\approx \frac{P(X = x_i | Z_i = 1)}{P(X = x_i | Z_i = 0)} \\ &= \frac{P(Z_i = 1 | X = x_i)P(X = x_i)/P(Z_i = 1)}{P(Z_i = 0 | X = x_i)P(X = x_i)/P(Z_i = 0)} \\ &\propto \frac{P(Z_i = 1 | X = x_i)}{P(Z_i = 0 | X = x_i)} \end{aligned} \tag{3}$$

Therefore we can rewrite (2) as:

$$P(S_i^* = 1 | X = x_i) \propto P(S_i = 1 | X = x_i) \frac{P(Z_i = 1 | X = x_i)}{P(Z_i = 0 | X = x_i)} \tag{4}$$

The survey weights are the reciprocals of the two components of Eq. 4. If the selection probabilities of the reference sample do not directly depend on the covariates \mathbf{X} , Elliot and Valliant suggest a so-called beta regression, see Ferrari and Cribari-Neto (2004). In our simulation, the reference sample is generated from an artificial population. For simplicity, we assign equal weights to all units, although it would be possible to define them differently if required. In the combined sample, the probability $\hat{P}(Z_i = z | X = x_i)$ can be estimated using means like logis-

tic regression, least absolute shrinkage and regression operator (LASSO)(Tibshirani 1996) or Bayesian additive regression trees (BART) (Chipman et al. 2010). In our simulations below we use logistic regression.

3.2 Calibration

The basic concept of Calibration is to construct survey weights such that the weighted population estimates match known population totals, which are typically obtained from an external source such as a census or a microcensus.

In calibration, the information used from the external source can be the joint distribution of discrete variables, like gender specific age-groups within federal states. In this case the calibration is achieved analytically by using the ratio of expected and observed frequencies in the separate cells of the calibration scheme, which is known as post-stratification. In case, that only the marginal distribution of the calibration variables is known one has to use an iterative procedure which is known as Generalized Raking (Deville et al. 1993). In our simulation we use the Generalized Raking in R (R Core Team 2024) which is implemented in the *rake* function of the R-Package *survey* (Lumley 2024).

4 The ESS as a simulation platform

The basic approach is to use a high quality survey sample for the generation of an artificial population. We choose the German subsample of the European Social Survey (ESS) as a database for our simulation experiments. In this population we have individual information on the internet use, interesting outcome variables and variables relevant for the above weighting procedures. We formulate a plausible model of the selection process. Then we will select repeated samples from the population. On the basis of each sample we apply the above mentioned weighting procedures. In the evaluation step we compare the population values with the uncorrected sample results and the weighted sample results. Thus we see a potential bias and also its potential reduction by weighting. As we also know the value of the outcome variable we, can also check the MAR condition⁴. Thus we can check whether standard demographic variables are sufficient to explain the participation process which are formulated in variables of internet use. Variances and confidence intervals can be derived via replications of the sampling process. An early application of this design was used by Schnell et al. (2017).

A substantial advantage of this approach is that effects of a mode change are avoided. Thus, the displayed biases are solely due to the selective effect of internet use. Standard comparisons of online surveys and surveys with different interview modes, like in Cornesse et al. (2020), are always confounded with potential mode effects, see Bottoni and Fitzgerald (2021) for results from the ESS.

⁴ Note, that this is not possible in real survey situations.

4.1 The design of the ESS

The ESS is a cross-national survey that aims to examine social change and stability in Europe. Since 2001, new cross-sectional surveys have been conducted every two years. The ESS follows strict random probability sampling and adheres to rigorous translation protocols to achieve comparability across countries. It has received numerous international accolades for its high-quality survey methodology (ESS ERIC 2024). The German subsample is a stratified, two stage sample with equal design weights for each person. The ESS conducts their survey via face-to-face interviews. At the start of our project the data from the 9-th round (2018) were available⁵. The data documentation mentions a nonresponse rate of 27.6% for the German subsample. For this reason a post-stratification weight is delivered which accounts for gender, age, education, and geographic region (Kaminska 2020). Note, that we did not use these weights for our simulations where we entirely refer to the artificial ESS-population. However, we used the weights for a below comparison of the real outcomes of the German parliament election 2017 and the internet use in the population.

4.2 Variables on internet participation

The ESS includes the variable *netuse* in rounds 1–5, corresponding to the years 2002–2010. This variable specifically inquires about internet access and the amount of time spent using the internet, as shown in Table 1. Respondents who selected the categories “Refusal”, “Don’t know”, and “No answer” were excluded⁶. For the years 2012 and 2014 (rounds 6 and 7), no information regarding internet access was

Table 1 Internet access in the ESS Round 1–9

Variable	Levels	Year/Round
<i>Netuse</i>	0 (No access at home or work)	2002–2010
How often do you use the internet, the World Wide Web or e-mail—whether at home or at work—for your personal use?	1 (Never use)	ESS1–ESS5
	2 (Less than once a month)	
	...	
	6 (Several times a week)	
	7 (Every day)	
<i>Netusoft</i>	1 (Never)	2016–2018
People can use the internet on different devices such as computers, tablets and smartphones. How often do you use the internet on these or any other devices, whether for work or personal use?	2 (Only occasionally)	ESS8, ESS9
	3 (A few times a week)	
	4 (Most days)	
	5 (Every day)	

⁵ The data can be accessed via the portal under <https://www.europeansocialsurvey.org/data>.

⁶ In Round 1, the German team added a special category: “Don’t know internet, e-mail, www”. The German data for the NETUSE variable have been excluded from the international data file, but the additional answer category is available in a separate country-specific file for Germany (ESS1 Documentation Report 2018).

collected. However, in 2016 and 2018 (rounds 8 and 9), data on internet access were again collected using the variable *netusoft*, which assessed the frequency of internet use without explicitly mentioning “internet access”.

Below we present a trend analysis of the selective effect of no access to the internet. For the analysis we collapsed the categories “no access at home or work” and “never use”.

In Section 2 we argued that an increased duration in the internet may result in a higher participation rate in online surveys. Here we use a threshold value. For 2018 the variable *netustm* returns the duration of internet use on a typical day in minutes. Its weighted average is 199.4min with 21% missing data. Here we use the threshold of 180min to distinguish between low and high internet use. For the selection mechanism **SM-1** we excluded all persons with a low internet use.

The dichotomisation of internet use is a very rough tool. A refined choice would be a scheme with a selection probability proportional to the duration of internet use. Therefore, selection scheme **SM-2** uses a sample of 1000 observations via pps-sampling, where the size variable is the duration of internet use. Figure 1 compares the distribution function of the duration of internet use for the ESS universe and the selected samples under **SM-2**. The sample distribution is shifted towards longer durations and the difference between the quantiles becomes larger for longer durations.

For example, the 25% quantile is shifted from 60min in the ESS universe to 180min in the sample.

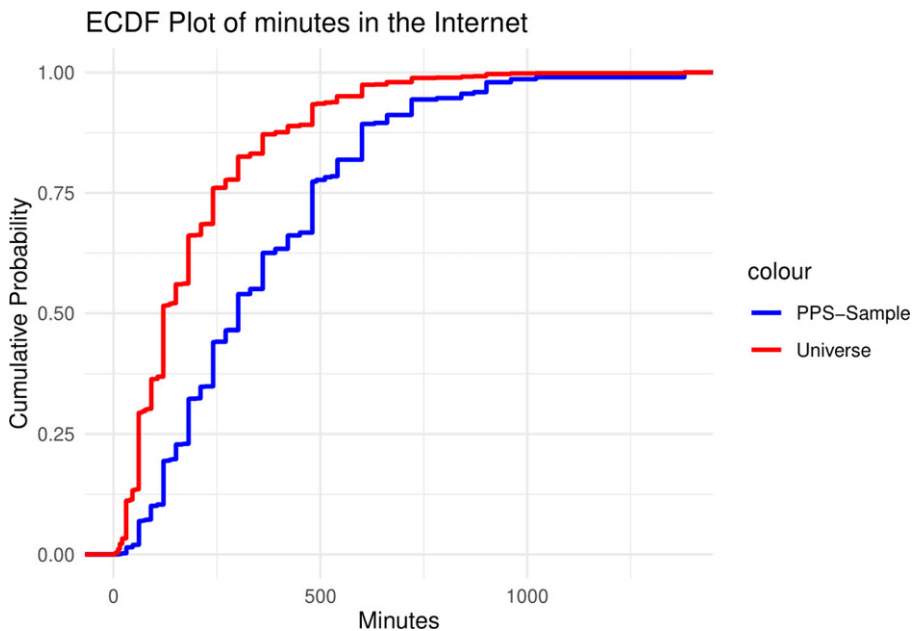


Fig. 1 Comparison of the distribution function of the duration of the internet use: ESS universe (Above, red) and the pooled pps-samples (Below, blue)

In order to include an activity indicator, we use the indicator whether one has recently shared a message. Here the ESS asked its respondents (Variable `postplon1`: “Have you posted or shared anything about politics online in the last 12 months?” (yes/no)). This selection mechanism **SM-3** aims to capture both the inclination to express opinions online and also an interest in political subjects. Finally, we combine long internet use and interest in politics which may especially attract persons to participate in online surveys about politics. This selection mechanism **SM-4** specifically addresses our outcome variable on voting behaviour.

4.3 Outcome variables

Online surveys are often evaluated by their ability to predict election outcomes. It is asked which political party the respondents would vote for if there would be general elections on the next Sunday (in German “Sonntagsfrage”). Because of the long time interval between successive rounds, the ESS questionnaire asks, about the vote in the last general national elections⁷. We consider the major parties of the Variable `prtvede2` listed in Table 3. Minor parties are collapsed in the category “Others”. Note that we did include non-voters by a separate variable `vote`, where the respondents state whether they voted or not.

We exclude respondents which were not eligible to vote, i.e. persons under the age of 18 or without German citizenship. In the trend analysis below we use the subjective health indicator (Variable `health`⁸) which has been already used by Schnell et al. (2017) for the 2010 round of the ESS. This variable is present for every round of the ESS.

4.4 Selection of variables for the construction of survey weights

For the application of the two weighting procedures we have to select appropriate variables. On the one hand, these variables should be predictive for the selection process, see, for example, Lundström and Särndal (2001). Their choice is a difficult task, as we do not observe the nonrespondents. On the other hand, if the selected variables are not linked to the likelihood of participation, they have only minor potential to correct a selection bias (Lee and Valliant 2009). A major advantage of using standard demographics is that they are usually found in both, the online survey and the reference survey. We aim to investigate the value of standard demographics for the bias reduction in online surveys.

For demographic variables we use dummy variables for *gender*, *age*, *marital status*, *educational attainment*, and *type of residence*. The variable *agea*, representing age in years, is divided into age categories as described in Fig. 2. The categories “Legally married” and “In a registered civil union” are referred to the category *married*. “Legally separated”, “Divorced/Civil union dissolved” and “Widowed/Civil partner died” are combined into the category *formerly married*, while “None of

⁷ Here we use the so-called Second Vote (“Zweitstimme”) which is representative for the proportions of the parties in the parliament.

⁸ With measurements on a Lickert scale ranging from “very good” to “very bad”.

Demographic variables		Additional variables	
Variable	Expression	Variable	Expression
Gender	Female Male	Satisfied with life	dissatisfied medium satisfied very satisfied
Age categories	18-30 31-40 41-60 60-70 over 70 years old	Trust in people	little trust medium trust strong trust
Marital status	married formerly married never married	Close to political Party?	-Christian Union -Social Democratic Party (SPD) -The Left -Alliance'90/The Greens (Greens) -Free Democratic Party (FDP) -Alternative for Germany (AfD) -Others -None
School graduation	category 1 category 2 category 3		
Residential area	city town village		

Fig. 2 Demographic variables and additional Variables

these” is placed in the category *never married*. Regarding education, the categories “Abitur” and “Fachhochschule” refer to *category 3 “tertiary level”*, “Mittlere Reife” into *category 2 “Upper secondary level”*, and “Förderschule” or “Hauptschule” into *category 1 “lower secondary”*. This aligns with the standard classification in the German school system. Additionally, individuals without any graduation are included in the lowest *category 1* due to low case numbers. To describe the residential area, the categories “A big city” and “Suburbs or outskirts of a big city” are grouped as *City*, “Town or small city” as *Town*, and “Farm or home in the countryside” and “Country village” as *Village*.

Nevertheless, the use of specific variables beyond demographics can still be valuable. Large online institutes use so-called “webographic” variables as a supplement to the demographic variables (Schonlau et al. 2007). These variables capture differences in the frequency distributions of telephone and online surveys. They do not directly relate to the selection process, rather, they are more linked to lifestyle variables, for example, whether the respondent has experienced a violation of privacy or has read a book in the last month (Schonlau et al. 2007; Table 1). However, such variables are not present in the ESS.

As mentioned in Section 2, motivational variables may play an essential role in the selection process. These variables can explain participation in an online survey beyond demographic variables. For instance, life satisfaction may trigger participa-

Table 2 Definitions of alternatives selection mechanisms

Selection criterion	Selection mechanism
At least 180 min of internet use	SM-1
Pps-sampling with duration of internet use as size variable	SM-2
Posted or shared anything about politics online last 12 months (yes)	SM-3
Combination of SM-1 and people who are at least quite interested in politics	SM-4

tion in surveys on political issues. Similarly, partisanship for a political party may be a strong predictor of one’s actual voting behavior. Furthermore, strong trust in people may motivate participation in surveys, either because these individuals believe in privacy regulations or because they hope that the survey results will be used in a positive manner.

Figure 2 presents, in the right column, the additional auxiliary variables we use for our analysis. The question “How satisfied are you with life as a whole?” utilizes a scale ranging from 0 (Extremely dissatisfied) to 10 (Extremely satisfied). We categorized as follows: 0–5 as “dissatisfied”, 6 or 7 as “medium”, 8 as “satisfied” and 9 or higher as “very satisfied”. The ESS also includes a question regarding trust in others, specifically “Do you think most people can be trusted, or do you believe that you can’t be too careful?” Participants responded on a scale from 0 (You can’t be too careful) to 10 (Most people can be trusted). We recode the answers as follows: below 4 as “little trust”, 4 and 5 as “medium trust”, and above 5 as “strong trust”. Another potentially influential predictor is party affiliation, which was determined by the question “Which party do you feel closer to?” The same parties as those used for the variable of interest were included, with the exception that the Non-voters

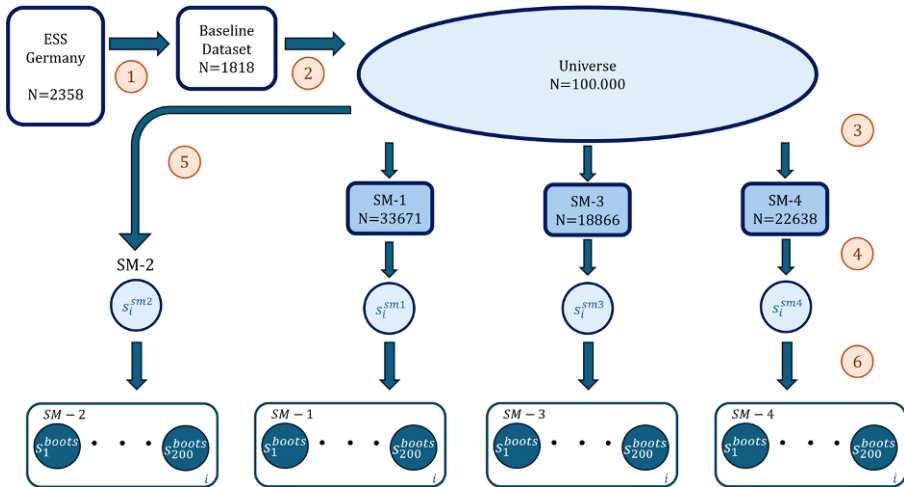


Fig. 3 Simulation Setup: (1) Use only people eligible to vote and delete missing information (2) Create universe with simple random sampling with replacement (3) Create subsets based on the selection variables (4) Draw $i = 1000$ simple random samples of size $n = 1000$ without replacement (5) same procedure as Step 4 with replacement and selection probability proportional to minutes in internet (6) Draw in each SM, for each i , 200 bootstrap samples

were again replaced, with a separate Variable, with respondents which are not close to any political party.

4.5 Design of the simulation

We created an artificial population of 100,000 persons from the 9-th round of the ESS by simple random sampling with replacement. In order to avoid later problems with missing data, we skipped all persons with missing demographic information and missing outcome variables from the population frame. This artificial universe serves as the reference for estimates obtained from the samples. Subsequently, three subsets were created based on the selection mechanisms **SM-1**, **SM-3**, and **SM-4** as described in Table 2. Samples with a size of 1000 are drawn from these subsets using simple random sampling without replacement. In case of selection mechanism **SM-2** the sample is directly drawn from the universe using proportional to size sampling based on the daily minutes on the internet. This process is repeated $i = 1000$ times for each selection mechanism, resulting in a total of $S = 1000$ samples for each of the four selection mechanisms. These stages are displayed in Fig. 3 as steps 1–6.

To estimate the selection bias, we compare the estimated values from the samples under different selection scenarios with the population values.

The bias across all samples S is calculated as follows:

$$bias(\hat{\theta}) = S^{-1} \sum_{s=1}^S (\hat{\theta}_s - \theta) \tag{5}$$

where $\hat{\theta}_s$ is the estimate from sample s , θ is the full population value.

We do not scale the bias to the relative bias. Instead, we measure biases in percentage points of votes, which is a convenient scale to interpret voting results.

Regarding the quasi-randomization approach, each sample is then merged with the universe, and the Z indicator is set to 1 for units in the non-probability sample

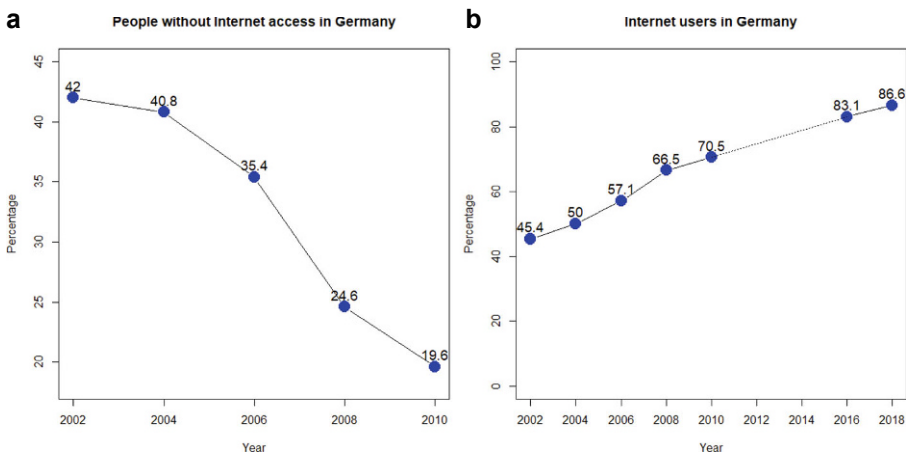


Fig. 4 Internet users over time according to ESS1–ESS9

and 0 for units in the reference sample, as described in Eq. 3. The reciprocal term of Eq. 4, that is $\hat{P}(Z_i = 0 | X = x_i) / \hat{P}(Z_i = 1 | X = x_i)$ is then calculated using a logistic regression with auxiliary variables from the combined datasets. Since the reference sample is the artificial population there is no need to use design weights in the computation of the QR-weights. The calibration information is also derived from the artificial population.

In Step 6 of Fig. 3 we take the variance of $\hat{\theta}$ into account. We are interested in the parametric 95% confidence interval coverage for the QR- and the calibration estimates. For this purpose we compute the t-statistic $t(\hat{\theta}_s) = (\hat{\theta}_s - \theta) / \sqrt{v(\hat{\theta}_s)}$, where $\hat{\theta}_s$ represents the estimate from sample s , θ is the population value, and $v(\hat{\theta}_s)$ is the variance estimator in sample s . Here we use no analytical expression for $v(\hat{\theta}_s)$, instead we use for each sample $R=200$ replications of the estimate⁹ and use their empirical variance. The mean of the t-values over the $S=1000$ samples estimates the coverage rate $CI(\hat{\theta})$:

$$CI(\hat{\theta}) = S^{-1} \sum_{s=1}^S I(-1.96 \leq t(\hat{\theta}_s) \leq 1.96) \quad (6)$$

where I is the indicator for the event in the parenthesis. $CI(\hat{\theta})$ measures how often the 95-percent confidence interval of the sample estimates covers the population value. Ideally, it should be equal to the nominal value 0.95. However, the lack to correct the bias completely may shift the coverage rate below the nominal level.

5 Empirical results

We first display the development of access to the internet according to the ESS data. We will then present a trend analysis on the bias of two outcome variables which is due to a simple restriction to internet users, i.e. all ESS persons who reported an access to the internet remain in the analysis while all persons with no access or no reported use of the internet are discarded from the analysis. The first outcome variable is the subjective health index which was used by Schnell et al. (2017). The second outcome variable is the reported vote in the election of the German parliament. In the main part of our analyses, we first check how well the demographic variables fit within the framework of the selection mechanisms **SM-1** to **SM-4**. Then we investigate the reduction of bias by the QR-weights and the calibration procedure. Finally we scale the bias relative to the difference of the ESS estimates and the real election result for the 2017 Bundestag.

5.1 Temporal trend in internet access and internet use

Figure 4 reveals that the proportion of people without internet access has more than halved between 2002 and 2010. Note that for the results presented in Fig. 3 as well

⁹ Thus results in 200×1000 replications for each selection mechanism.

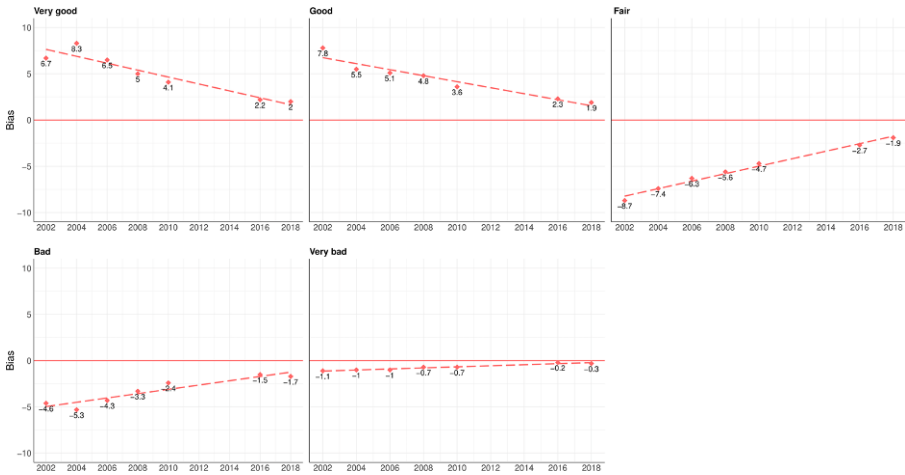


Fig. 5 Bias of Subjective health status for internet users in Germany estimated by the European Social Survey. Positive (negative) figures indicate an over-estimation (under-estimation) by the subsample of persons with internet use compared to the entire sample

as in Figs. 4 and 5, we use the post-stratification weights from the ESS introduced in Section 4.1. Although the ESS only provides information about internet access up until 2010, Fig. 4 suggests that the number of people without internet access in the ESS (2018) should be in the range of low single digits.

Looking at these figures one might expect that the potential of a bias due to a mere non-access or non-use is diminishing. This could imply that bias statements based on a low proportion of internet users in the population are no longer valid.

Therefore, we replicate the result of Schnell et al. (2017) on the over-representation of healthy persons among internet users. Figure 5 demonstrates that this over-

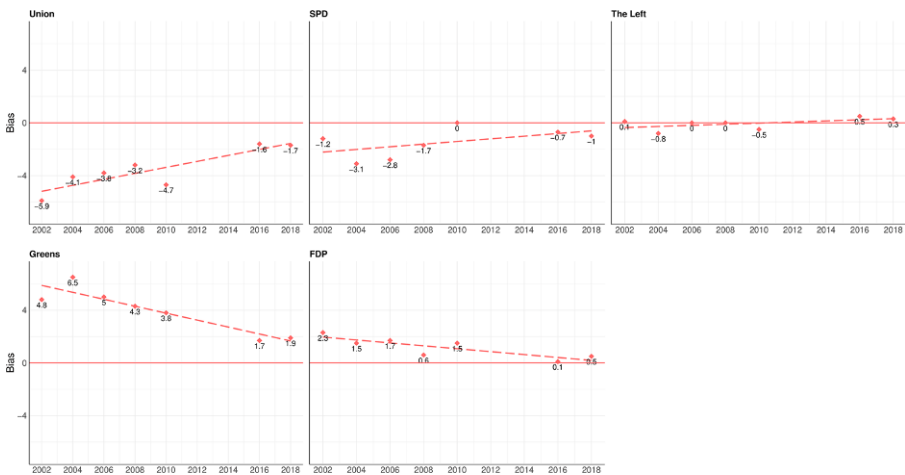


Fig. 6 Bias of political vote for internet users in Germany

representation in 2002 has melted down until 2018. Such a finding also holds for the voting behaviour in the 2017 election of the German Bundestag. Figure 6 displays the bias in the results of the respective parties over time, assuming that only eligible voters who use the internet are sampled.

The bias is illustrated in terms of the deviation from the zero line for each party over time. If a point is above the zero line, it indicates an over-estimation of the voting result in the selected subset of internet users. For instance, in 2002, the Greens would have 4.8% points more votes than in the underlying complete ESS. Conversely, if a point is below the zero line, it signifies an under-estimation of the election result. For example in 2002, the Union would have 5.9% points fewer votes compared to the entire sample. The analysis reveals that the Conservative Union tends to be under-estimated in online samples, while the Green party, which focuses on environmental awareness and post-materialistic values, tends to be over-estimated.

A similar pattern can be observed for SPD and FDP. It is noteworthy that the points marked by the dashed line¹⁰ approach the zero line over time.

In these examples no bias correction was employed and the distinction between internet users and non-users alone does not seem to be relevant anymore. However, the argument may fail if the selection process is more complicated, like in the selection models **SM-1** to **SM-4** below.

5.2 The impact of demographic and additional control variables

We first study the impact of selection models **SM-1** to **SM-4** on demographic and other control variables using the 2018 ESS dataset for Germany. In the second part of this section we will check whether or not the selection according to these models can be controlled by demographic variables.

Figure 7 compares the proportions in the artificial universe (first column) and the samples according to selection model **SM-1** to **SM-4** (column 2 to 5). In terms of standard demographics, there is a male surplus in the samples. The proportion of younger individuals up to 40 years of age is higher in all subsets, while the number of older people over 70 is lower in all scenarios. Furthermore, it is evident that the participants in the different selection scenarios, tend to be never married, highly educated (as indicated by the substantive increase in school category 3), and more likely to reside in urban areas. In summary, these selection models result in under- and over-coverage in terms of standard demographics.

Regarding the additional variables, Fig. 7 indicates that potential participants are slightly more likely to have trust in people compared to the universe, especially under selection model **SM-4**. In terms of political partisanship, the share of Union and SPD preferences appears to be under-represented, while the Greens are over-represented.

The AFD also shows a particularly strong over-representation among the subset of individuals who publicly posted their political view (**SM-3**). Also the percentage of people without any political partisanship is exceptionally lower in **SM-3** and

¹⁰ The dashed line is the OLS regression line of the separate bias figures.

Demographic variables					Additional variables					
Universe	SM-1	SM-2	SM-3	SM-4	Universe	SM-1	SM-2	SM-3	SM-4	
Male	52.4	56.2	57.4	53.2	59.3	Satisfaction				
Age 18-30	15.7	34.9	29.5	27.5	29.1	Dissatisfied	13.8	13.7	14.5	11.4
Age 30-40	12.5	18.8	19.2	20.6	17.3	Medium sat.	20.1	21.6	20.3	20.5
Age 40-60	34.8	34.0	36.0	33.1	37.6	Satisfied	31.1	30.0	32.0	31.7
Age 60-70	19.8	8.5	11.0	14.1	11.1	Very sat.	36.4	35.2	34.7	33.3
Age over 70	17.2	3.7	4.3	4.7	4.9	Trust				
Married	58.2	44.5	47.8	47.6	48.3	Little	19.2	19.2	21.0	19.1
Form.married	15.9	9.3	9.5	9.6	10.4	Medium	29.2	25.8	25.0	25.5
Nev.married	25.9	46.2	42.7	42.9	41.4	Strong	51.6	55.0	54.0	55.3
School cat.1	24.3	11.7	13.9	13.1	8.3	Partisanship				
School cat.2	33.5	31.4	32.0	30.7	27.2	Union	21.5	18.6	18.3	17.8
School cat.3	42.2	56.8	54.1	56.2	64.5	SPD	11.8	9.6	9.0	9.9
City	30.2	37.0	38.2	34.8	37.9	The Left	5.8	5.7	5.9	7.5
Town	35.8	35.6	34.0	34.6	35.8	The Greens	12.6	15.4	14.7	18.2
Village	34.0	27.3	27.8	30.7	26.4	FDP	3.4	3.7	4.1	4.1
						AfD	3.4	3.4	3.4	6.3
						Others	1.4	1.4	2.0	2.8
						None	40.0	42.2	42.6	33.4
										33.3

Fig. 7 Descriptive distribution of demographic and related variables under the Selection mechanisms SM-1 to SM-4

SM-4. Additionally the selection schemes show only minor differences with respect to life satisfaction.

5.2.1 Testing the MAR assumption

One advantage of the simulation approach is that we can test the appropriateness of the MAR assumption as we know the value of the Y -variable also for the non-selected units. In the subsequent analysis we estimate different models for the selection indicator, which is dichotomous in the case of **SM-1**, **SM-3** and **SM-4**. In the case of **SM-2**, where we modeled the mechanism proportionally, we obtained a metric indicator. In this case, we chose a linear model for the sake of simplicity.

Table 3 Variable of interest in the ESS

Variable	Expression
Party Vote	–
	<i>Christian (Democratic/Social) Union (CDU/CSU)</i>
	<i>Social Democratic Party (SPD)</i>
	<i>The Left</i>
	<i>Alliance '90/The Greens (Greens)</i>
	<i>Free Democratic Party (FDP)</i>
	<i>Alternative for Germany (AfD)</i>
	<i>Others</i>
	<i>Non-voters</i>

The first model, the so-called “restricted model”, uses the chosen X-variables. The second model, the so-called “unrestricted model”, adds Y as an additional variable to the explaining variables. Here we use the party vote. To be specific, we used 8 indicators for the votes listed in Table 3. We then test the joint significance of these indicators by a likelihood ratio test. The degrees of freedom of the resulting χ^2 -statistic is $8 - 1 = 7$.

Table 4 presents the results of likelihood ratio tests for the four Selection Models **SM-1** to **SM-4** when using the demographic variables as covariates. With the exception for **SM-1**, all selection schemes show a separate impact of voting for a specific party. This means the MAR assumption is rejected here. However, for the **SM-1**, the MAR assumption holds. Based on the *Chisq*-value we assume that the correction by demographic variables will perform best under **SM-1**.

Table 5 demonstrates that including additional variables fulfill the MAR assumption for **SM-4**. This is not surprising as **SM-4** is a combination of **SM-1** and interest in politics. Here the additional variable is political partisanship which is a strong indicator for interest in politics. Note, however, that this variable does not necessarily coincide with the vote in the previous elections. For **SM-3** (Posted anything about politics) the *Chisq* value of the Y-variable has decreased by one third, from 29 to 20. Thus we also expect a substantial bias reduction by using the additional variables.

In the next section we investigate the bias reduction under the different selection schemes. The rate of bias reduction turns-out to be independent from the fact whether the MAR-condition holds or not. Thus, the claim of Schnell et al. (2017) that under the violation of the MAR-condition “no weighting strategy will be able to eliminate health bias in web surveys” is somewhat restrictive as it ignores the possibility of a substantial bias reduction even under the violation of the MAR-condition. This will be demonstrated for our political voting example below.

5.3 Bias correction

Table 6 and 7 present averaged estimates for the party votes based on the simulated 1000 samples. Table 6 shows the correction of the bias by the QR- and the calibration-weights using demographics only. Table 7 incorporates additional information on life satisfaction, trust in people, and political orientation. The second row provides the percentages for the ESS universe. They are the target values which have

Table 4 MAR with standard demographics

Selection	Model	DF	LogLik	Chisq	$Pr(> Chisq)$
SM-1	Restricted	12	-927.06	-	
	Unrestricted	19	-922.68	8.7613	0.2703
SM-2	Restricted	13	-11709	-	
	Unrestricted	20	-11696	24.736	0.000845***
SM-3	Restricted	12	-819.68	-	
	Unrestricted	19	-805.09	29.186	0.0001338***
SM-4	Restricted	12	-856.37	-	
	Unrestricted	19	-847.83	17.082	0.01687*

Table 5 MAR with additional variables

Selection	Model	DF	LogLik	Chisq	$Pr(> Chisq)$
SM-1	Restricted	24	-921.54	-	
	Unrestricted	31	-918.46	6.1481	0.5226
SM-2	Restricted	25	-11702	-	
	Unrestricted	32	-11690	24.057	0.001113**
SM-3	Restricted	24	-802.01	-	
	Unrestricted	31	-791.97	20.064	0.005432**
SM-4	Restricted	24	-844.03	-	
	Unrestricted	31	-837.08	13.898	0.05302

to be approximated by the application of the QR- and the calibration weights. The subsequent rows display the average bias across the samples for the uncorrected sample, the QR-weights, and the calibration weights for each selection mechanism. The last column, Bias, presents the absolute difference between the ESS universe and the respective correction result, summed over all political parties.

The tables reveal that Union and SPD are throughout under-represented by all uncorrected selection mechanisms, while the Greens tend to be over-represented. Considering the overall over- or under-estimation of election results by samples, the uncorrected absolute bias ranges from 15.3 percentage points in **SM-1** to 22.2 percentage points in **SM-3**.

The correction in Table 6 appear to work well for the major parties, Union, SPD, and Greens. For instance, the uncorrected result of 25.6% for Union in **SM-1** approaches the universe value of 30.2%, reaching 30.0% with QR-weights and 29.7% with calibration weights. The corrections using only demographics effectively reduce the substantial bias for bigger parties across different selection mechanisms. However, the results for smaller parties are mixed. For example, the uncorrected under-estimation of The Left under **SM-2** of 5.6% compares to 6.6% in the universe. However, this under-representation becomes more severe with 4.4 and 4.3% by the weighting routines. Nonetheless, the overall absolute bias across all political parties is reduced under all selection mechanisms except for **SM-3**. The overall reduction is considerable. For example, in **SM-4**, the weighting approaches reduce the overall bias from 21.1% to 11.6 and 11.3%. An exception is the performance under **SM-3**. The model with demographic variables do clearly not fulfill the MAR-condition.

Table 6 Estimated mean Bias using only demographics

Selection	Population	Union	SPD	The Left	Greens	FDP	AFD	Others	Nonvoters	Bias
–	Universe	30.2	19.2	6.6	16.1	7.9	5.9	2.3	11.7	(–)
SM-1	Uncorrected	25.6	16.5	6.3	19.6	9.3	6.1	3.0	13.6	15.3
	QR	30.0	19.5	4.0	15.7	10.7	7.0	1.9	11.3	8.2
SM-2	Calibrated	29.7	19.7	3.9	15.5	10.8	7.0	1.9	11.5	8.9
	Uncorrected	25.3	16.4	5.6	19.1	9.8	6.3	2.7	14.8	17.5
SM-3	QR	29.5	18.7	4.4	16.0	10.4	7.1	2.0	11.9	7.7
	Calibrated	29.4	18.8	4.3	15.6	10.4	7.1	2.0	12.3	8.6
SM-4	Uncorrected	24.8	16.7	7.2	20.4	7.4	10.8	3.6	9.0	22.2
	QR	25.0	19.9	4.8	18.1	9.4	12.6	2.7	7.5	22.5
SM-4	Calibrated	25.2	19.4	4.3	18.1	9.7	13.2	2.7	7.6	23.1
	Uncorrected	25.9	18.1	7.9	22.7	9.7	5.9	3.2	6.6	21.1
SM-4	QR	30.8	20.1	5.1	17.0	10.5	6.6	2.5	7.5	11.6
	Calibrated	30.2	20.1	5.3	17.6	10.4	6.4	2.6	7.4	11.3

Table 7 Estimated mean Bias using demographics and additional variables

Selection	Population	Union	SPD	The Left	Greens	FDP	AFD	Others	Novoters	Bias
SM-1	Universe	30.2	19.2	6.6	16.1	7.9	5.9	2.3	11.7	(-)
	Uncorrected	25.6	16.5	6.3	19.6	9.3	6.1	3.0	13.6	15.3
	QR	29.7	19.3	4.6	15.6	10.9	6.4	2.1	11.4	7.1
SM-2	Calibrated	30.4	19.0	5.5	16.2	9.9	5.9	2.4	10.8	4.6
	Uncorrected	25.3	16.4	5.6	19.1	9.8	6.3	2.7	14.8	17.5
	QR	29.4	19.5	4.8	16.2	10.0	6.5	2.0	11.8	6.1
SM-3	Calibrated	29.7	19.8	5.2	16.0	9.8	6.2	2.2	11.2	5.4
	Uncorrected	24.8	16.7	7.2	20.4	7.4	10.8	3.6	9.0	22.2
	QR	26.8	21.9	5.0	16.5	9.1	10.0	2.4	8.3	16.9
SM-4	Calibrated	28.3	22.1	4.7	16.7	7.3	9.0	2.5	9.5	13.4
	Uncorrected	25.9	18.1	7.9	22.7	9.7	5.9	3.2	6.6	21.1
	QR	31.4	19.8	4.9	15.5	10.2	7.2	3.2	7.8	12.5
	Calibrated	31.5	19.3	6.3	17.4	8.8	6.4	3.6	6.8	10.6

This can be also seen from the absolute bias values in Table 6. The uncorrected bias amounts to 22.2%. Despite the QR correction, the bias increases to 22.5%, and with the calibration correction, it further rises to 23.1%. Note, however, that the demographic correction variables work quite well under **SM-2** where the MAR-condition does not hold.

The inclusion of additional variables in the correction schemes in Table 7, reduces the overall bias, with the exception of the QR-approach under **SM-4**, which performs slightly worse. It is noticeable that in **SM-3**, both weighting approaches now significantly reduce the overall bias, from 22.2 to 16.9 and 13.4, respectively. Thus the additional variables, while not sufficient to meet the MAR conditions, improve the quality of the weighting procedure.

If we compare the two weighting approaches we find only minor differences in their performance. If we use demographic variables, the QR-approach works slightly better. However, in the presence of influential calibration variables, like political partisanship, the calibration approach becomes more powerful.

Figure 8 and display the distribution of the QR estimates from the 1000 samples, incorporating additional variables. The rhombus's distance from the dashed line represents the uncorrected bias, while the boxplots illustrate the single estimates, which, when averaged, produce the corrected estimators presented in Table 6 and 7. The results for the calibrated estimator are similar and can be found in the appendix Fig. 11.

Notably, the corrections demonstrate the effectiveness of the weighting schemes, particularly for the major parties Union, SPD, and Greens, as evidenced by the boxplots' proximity to the dashed line, indicating only small biases. However, for smaller parties, the corrections consistently under- or over-estimate the population value, see, for example, the results for The Left under **SM-1** and **SM-2**. One may also wonder why the correction sometimes overshoots the ESS target value, despite the MAR condition is fulfilled in **SM-1**. The reason is our simple selection model:

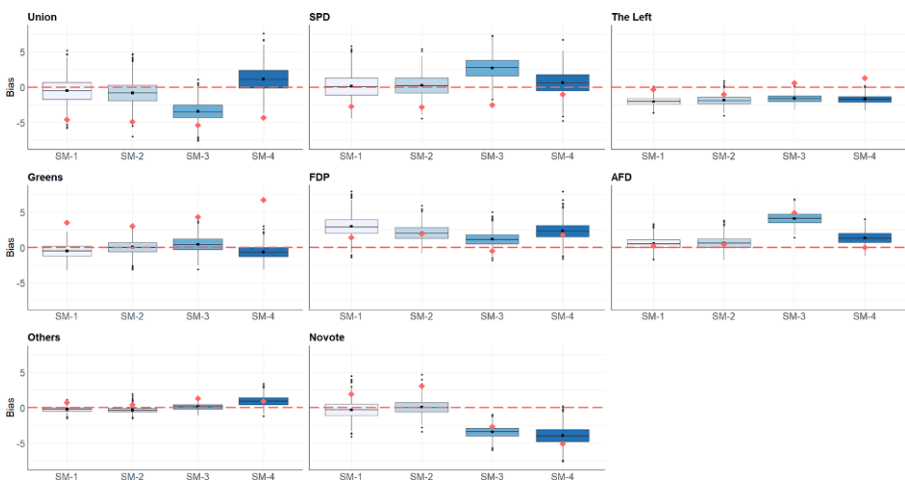


Fig. 8 QR estimates using demographics and additional variables

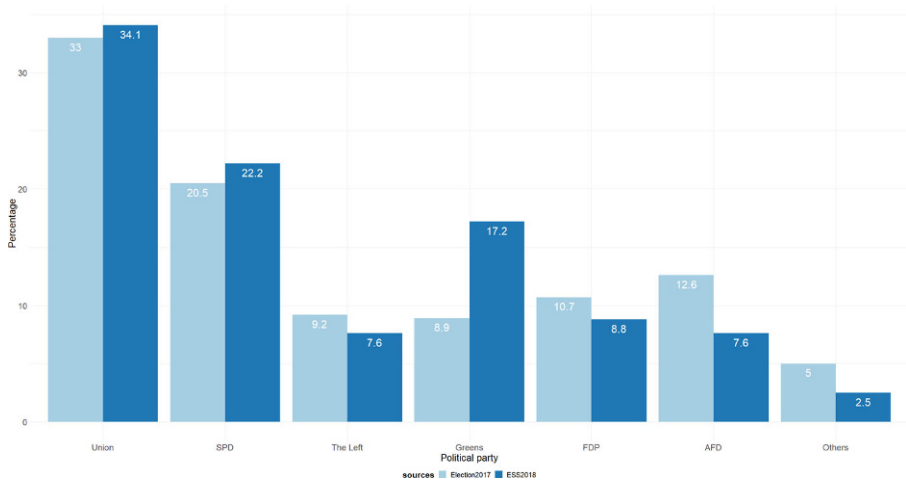


Fig. 9 Comparison of ESS with the Bundestagswahl 2017

for the QR-approach we used a main effects logit model and for the calibration we used only the marginal population distributions of the demographic and the additional variables. In both cases we omitted interactions of the covariates which may be relevant for the selection process.

5.4 Coverage rates

When considering coverage rates, both correction procedures do not precisely align with the population values at the nominal level of 95%, except for the QR-estimate under **SM-2** and the calibration estimate under **SM-1**, both for the Green party. Of course, one cannot expect that the nominal value of 950/1000 is met perfectly, as the bias of the weighted estimates is only partially removed. For example, under

Table 8 Coverage rate of 1000 QR-estimates using demographic and additional variables

Selection	Union	SPD	The Left	Greens	FDP	AfD	Others	Novote
SM-1	942	936	97	929	415	928	850	905
SM-2	899	953	281	948	574	921	792	943
SM-3	311	609	219	955	833	0	956	40
SM-4	921	928	163	871	586	738	841	182

Table 9 Coverage rates of 1000 calibration estimates using demographic and additional variables

Selection	Union	SPD	The Left	Greens	FDP	AfD	Others	Novote
SM-1	932	926	600	957	531	943	937	817
SM-2	932	934	512	942	596	944	888	893
SM-3	731	580	88	918	822	19	959	462
SM-4	875	948	907	805	856	910	613	19

SM-1 the population value of the Union is approached quite well by the calibration estimate with additional variables and consequently the empirical coverage rate of 932/1000 fits the nominal value of 0.95 quite well. However, the QR-estimate of the Union still exhibits a bias of approximately 2 percentage points under **SM-3**. Because of the low variances of the estimate the corresponding coverage rate drops to 311/1000 in Table 8. Thus, the coverage rates are highly sensitive with respect to a potential bias.

There is no clear ranking of the coverage rates between the QR-estimates and the calibration estimates (Table 9). Similar results are observed when using only demographics (see Table 10 and 11 in the appendix).

5.5 Comparison with real voting results

In our simulation setting, we generated subsets from an artificial universe to observe a selection bias and its correction. However, as the actual values in the population are typically unknown, a direct comparison with population values is not feasible. Nevertheless, in this example, we have access to the election results of the German Bundestag in 2017. Therefore, we can scale the biases in our simulation setting to the differences of the ESS population estimates and the results of the Bundestag elections. This provides a basis for evaluating the effectiveness of the correction in our simulated setting. To compare the ESS with the election results, we apply the provided ESS weights, which reflect the design weights and a demographic calibration to correct for the nonresponse in the ESS. Figure 9 compares the weighted ESS results from 2018 with the Bundestag election results of 2017¹¹. The ESS over-

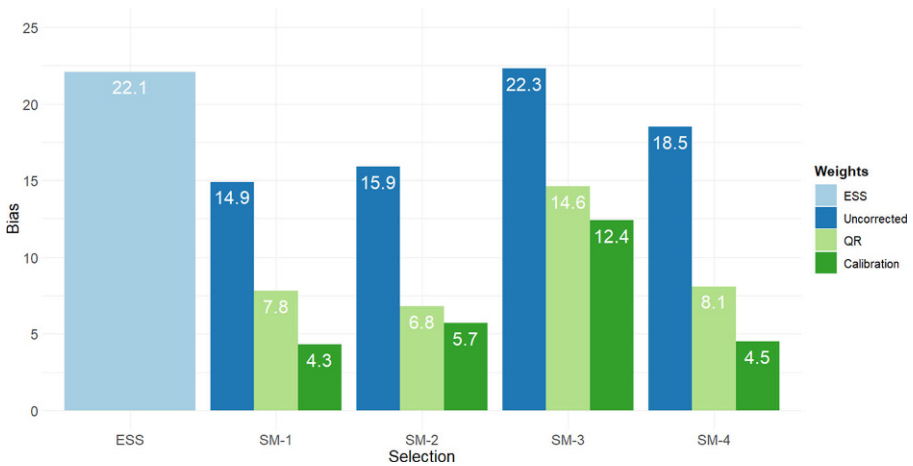


Fig. 10 Comparison of bias and corrections using demographics and additional variables. The first bar (ESS) represents the sum of the weighted absolute biases relative to the actual election results, based on the post-stratification weights from the ESS, while the SM-1 to SM-4 bars display the corresponding sums of deviations with respect to the artificial population under different correction methods

¹¹ Note, that the ESS questionnaire asked for the recent voting in the 2017 Bundestag election. It did not ask how one would vote at the moment of the interview in 2018.

estimates the vote-share of the Greens by nearly double while under-estimating the AFD share. This highlights the challenge of accurately estimating voting results even for a well-executed probability survey.

Figure 10 displays the sum of the absolute biases over all political parties before and after the correction procedure. The first bar shows the absolute bias for the comparison of the ESS population estimate and the Bundestag results. In the subsequent bars we replicate the results from Table 7 for the selection models **SM-1** to **SM-4** while excluding non-voters. Figure 10 indicates that the deviations of a real probabilistic survey, like the ESS, from the population values have approximately the same magnitude as the uncorrected online surveys estimates under various selection models. The discrepancy between the ESS and the actual election results is approximately 22.1 percentage points, primarily due to overestimation of the Greens. The dark blue bars represent biases of the artificial universe in our simulation setting, while the two green bars represent the biases after applying the correction methods. It becomes evident that with the application of these correction methods the corrected non-probability estimates have smaller deviations from their population values than the final ESS-estimates from the true population values.

6 Discussion

We have demonstrated the use of the ESS as a simulation basis for the evaluation of weighting strategies for data from online surveys. The selection process of the data is assumed to be not observed, like in River Sampling from the internet. Therefore we formulated four potential selection models which are motivated by a-priori knowledge on internet use: the duration of daily internet use, the activity in social channels via posting, and the interest in politics.

We used two weighting strategies which do not rely on design-based sampling probabilities. First the Quasi Randomization (QR) approach which compares the non-probability sample distribution with a probability reference distribution, and second, the calibration approach, which restricts the weighted sample results to known population values. These weighting procedures use variables which have to be recorded in the non-probability sample and also in the reference sample or population. For this reason, the set of variables is often restricted to demographic variables. However, if available, additional variables which are potentially linked to the selection process, can be used too. Therefore we used life satisfaction, trust in people, and political partisanship. The latter variable seems to be especially useful for the prediction of survey participation in an online survey on politics. As an outcome variable we used the votes for the previous parliament election in Germany.

In this simulation setting, we could test the MAR-condition. In only one out of four for selection models the demographic variables were sufficient to guarantee the MAR-condition. This couldn't be improved by the use of additional non-demographic variables, like political partisanship. However, despite whether the MAR-condition holds or not, we achieved substantive bias reductions in both cases.

We compared the ESS population estimates for the 2017 Bundestag election with the real election result. We used the difference of this high quality probability survey

to the population count as a means to judge the differences in our simulation setting. It turned out that the total absolute differences of the original ESS estimates and the population results were about the same size as the uncorrected biases of our four selection schemes. The weighting procedures did reduce these differences in some cases by more than a half of their initial size. Thus, we obtained a better fit of the population values than the ESS in the case of the Bundestag results.

We have also shown that previous selection models which are based on the mere access to the internet, either defined by no internet connection at home or by non-use, have meanwhile lost their potential to create a substantial bias. For this result we have used the biannual replication of the ESS from 2002 to 2018. Therefore, previous results on internet online survey which are based on mere access are probably no longer valid.

We choose our selection models on general plausibility arguments of participation in surveys. In these models a selection bias can be controlled to a size which is not bigger (or even smaller) than in a high quality survey. The weakness of our approach is the lacking empirical knowledge on the selection process. Therefore, it is important to further investigate potential determinants associated for online participation. Survey agencies conducting online surveys might run experiments with selected subgroups of their access panel participants and their information on internet use.

A different useful dimension of the ESS which has not been used here is the comparability over its European subsamples. For the analysis presented here we have only used the German subsample. However, it would be interesting to investigate our sampling models also for the other European participants of the ESS. In this case, one should use a variable with an overall general meaning, like the satisfaction with health.

7 Appendix

Table 10 Coverage rate of 1000 QR-estimates using demographic variables

Selection	Union	SPD	The Left	Greens	FDP	AfD	Others	Novote
SM-1	944	940	13	932	520	859	743	906
SM-2	919	929	203	943	504	821	811	953
SM-3	115	938	145	713	815	0	936	0
SM-4	936	920	288	894	643	913	939	92

Table 11 Coverage rate of 1000 calibration estimates using demographic variables

Selection	Union	SPD	The Left	Greens	FDP	AfD	Others	Novote
SM-1	934	935	10	918	463	852	746	922
SM-2	911	937	185	915	522	818	809	929
SM-3	139	953	8	740	733	0	940	1
SM-4	952	923	472	753	562	931	937	59

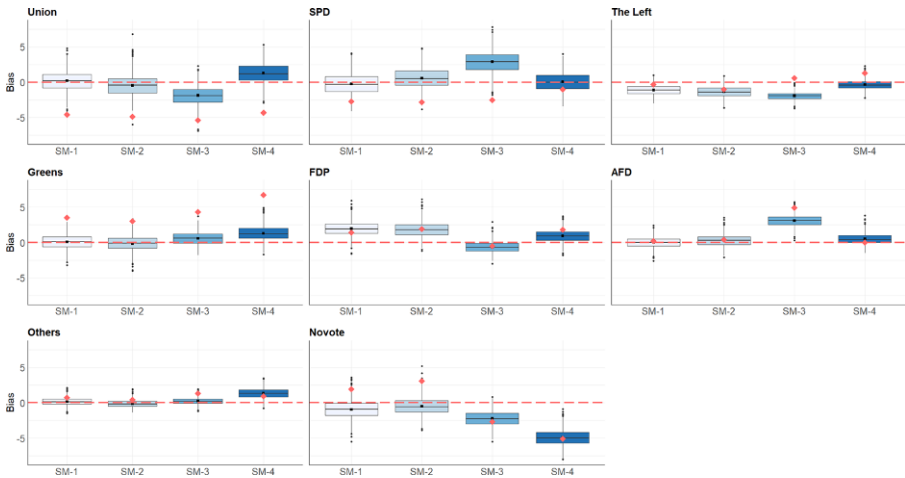


Fig. 11 Calibrated estimates using demographics and additional variables

Acknowledgements We would like to thank Tobias Wolfram, Johann Krümmel and Jakob Richter from Civey GmbH for numerous hints on River sampling and careful comments on a draft version of this paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Conflict of interest F. Prücklmair declare that he has no competing interests. U. Rendtel is a member of an advisory board of Civey. This activity is unpaid.

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

References

- Amarov B, Rendtel U (2013) The recruitment of the access panel of German official statistics from a large survey in 2006: empirical results and methodological aspects. *Surv Res Methods* 7:103–114
- American Association for Public Opinion Research (AAPOR) (2010) Report on Online Panels. <https://aapor.org/wp-content/uploads/2022/11/nfq048.pdf>. Accessed 28 July 2023
- Bethlehem J (2010) Selection bias in web surveys. *Int Stat Rev* 78:161–188
- Blom AG, Gathmann C, Krieger U (2015) Setting up an online panel representative of the general population: the German Internet panel. *Field Methods* 27(4):391–408
- Bottoni G, Fitzgerald R (2021) Establishing a baseline: bringing innovation to the evaluation of cross-national probability-based online panels. *Surv Res Methods* 15:115–133. <https://doi.org/10.18148/srm/2021.v15i2.7457>
- Callegaro M, Lozar Manfreda K, Vehovar V (2015) *Web survey methodology*. SAGE, London
- Chang L, Krosnick JA (2009) National surveys via RDD telephone interviewing versus the Internet: comparing sample representativeness and response quality. *Public Opin Q* 73:641–678

- Chipman HA, George EI, McCulloch RE (2010) BART: Bayesian additive regression trees. *Ann Appl Stat* 4(1):266–298
- Cornesse C, Blom AG, Dutwin D, Krosnick JA, de Leeuw ED, Legleye S, Pasek J, Pennay D, Phillips B, Sakshaug JW, Struminskaya B, Wenz A (2020) A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *J Surv Stat Methodol* 8:4–36
- Deming W, Stephan F (1940) On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann Math Stat* 11(4):427–444
- Dever JA, Rafferty A, Valliant R (2008) Internet surveys: can statistical adjustments eliminate coverage bias? *Surv Res Methods* 2:47–62
- Deville JC, Särndal CE (1992) Calibration estimators in survey sampling. *J Am Stat Assoc* 87(418):376–382
- Deville JC, Särndal CE, Soutory O (1993) Generalized raking procedures in survey sampling. *J Am Stat Assoc* 88(423):1013–1020
- van Eimeren B, Gerhard H (2000) Entwicklung der Onlinemedien in Deutschland, ARD/ZDF-Online-Studie 2000: Gebrauchswert entscheidet über Internetnutzung. *Media Perspekt* 8:338–349 (https://www.ard-zdf-onlinestudie.de/file/s/2000/Online00_Nutzung.pdf (accessed 3.7.2024))
- Elliot MR, Valliant R (2017) Inference for nonprobability samples. *Stat Sci* 32(2):249–264
- Enderle T, Münnich R, Bruch C (2013) On the impact of response patterns on survey estimates from access panels. *Surv Res Methods* 7:91–101
- ESS1 edition 6.6 (2018) ESS1 - 2002 Documentation Report, The ESS Data Archive. https://stessrelpubprodwe.blob.core.windows.net/data/round1/survey/ESS1_data_documentation_report_e06_7.pdf. Accessed 17 Mar 2022
- ESS ERIC (2024) About the European Social Survey European Research Infrastructure (ESS ERIC). <https://www.europeansocialsurvey.org/about-ess>. Accessed 3 July 2024
- ESS Round 9: European Social Survey Round 9 Data (2018) Data file edition 3.1. Sikt - Norwegian Agency for Shared Services in Education and Research, Norway – Data Archive and distributor of ESS data for ESS ERIC
- Ferrari S, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. *J Appl Stat* 31(7):799–815
- Hargittai E (2002) Second level digital divide: mapping differences in people’s online skills. *First Monday* 7:4
- Kaminska O (2020) Guide to using weights and sample design indicators with ESS data. https://www.europeansocialsurvey.org/sites/default/files/2023-06/ESS_weighting_data_1_1.pdf. Accessed 3 July 2024
- Lee S, Valliant R (2009) Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociol Methods Res* 37(3):319–343
- Lehdonvirta V, Oksanen A, Räsänen P, Blank G (2020) Social media, web, and panel surveys: using non-probability samples in social and policy research. *Policy Internet* 13:134–155. <https://doi.org/10.1002/poi3.238>
- Little RJA, Rubin D (2020) *Statistical analysis with missing data*, 3rd edn. Wiley, Hoboken
- Lohr S (2023) *Sampling. Design and analysis*, 3rd edn. Chapman and Hall/CRC, Boca Raton
- Lumley T (2024) *survey* (version 4.4-2), *Analysis of Complex Survey Samples*. <https://www.rdocumentation.org/packages/survey/versions/4.4-2>. Accessed 3 July 2024
- Lundström S, Särndal CE (2001) Estimation in the presence of nonresponse and frame imperfections. *Statistiska centralbyrån, Statistics Sweden*
- Malhotra N, Krosnick JA (2007) The effect of survey mode and sampling on inferences about political attitudes and behavior: comparing the 2000 and 2004 ANES to Internet surveys with Nonprobability samples. *Polit Anal* 15:286–323
- R Core Team (2024) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna (<https://www.R-project.org/> (accessed 3.7.2024))
- Rendtel U, Amarov B (2014) The access panel of German official statistics as a selection frame. In: Engel JL, Scherpenzeel, Strugis (eds) *Improving survey methods*. Taylor & Francis, Routledge, pp 236–249 (Chapter 20)
- Richter G, Wolfram T, Weber C (2023) Die Statistische Methodik von Givvy. Eine Einordnung im Kontext gegenwärtiger Debatten über das Für und Wider internetbasierter nicht-probabilistischer Stichprobenziehung. https://assets.ctfassets.net/ublc0iceiwck/2ui4wF5JBzs0ZaLJdqMR9/bff2a3db6b690d056ea807b61ce748e9/whitepaper_04_05_23.pdf. Accessed 27 July 2023
- Rubin D (1976) Inference and missing data. *Biometrika* 63:581–592
- Särndal CE (2007) The calibration approach in survey theory and practice. *Surv Methodol* 33(2):99–119

- Särndal CE, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer
- Schnell R, Noack M, Torregroza S (2017) Differences in general health of Internet users and non-users and implications for the use of web surveys. *Surv Res Methods* 11(2):105–123
- Schonlau M, van Soest A, Kapteyn A (2007) Are “Webographic” or attitudinal questions useful for adjusting estimates from Web surveys using propensity scoring? *Surv Res Methods* 1(3):155–163
- Statistisches Bundesamt (Destatis) (2021) Qualitätsbericht, Einkommens- und Verbrauchsstichprobe EVS 2018. <https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Einkommen-Konsum-Lebensbedingungen/einkommens-verbrauchsstichprobe-2018.html>
- Statistisches Bundesamt (Destatis) (2022) Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik - Deutschland. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Ausstattung-Gebrauchsgueter/Tabellen/liste-infotechnik-d.html>. Accessed 3 July 2024
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Royal Stat Soc Ser B* 58(1):267–288
- Valliant R (2020) Comparing alternatives for estimation from nonprobability samples. *J Surv Stat Methodol* 8(2):231–263. <https://doi.org/10.1093/jssam/smz003>
- Valliant R, Dever J (2011) Estimating propensity adjustments for volunteer web surveys. *Social Methods Res* 40(1):105–137
- Van Dijk J (2019) The digital divide. Polity Press
- Vehovar V, Toepoel V, Steinmetz S (2016) Non-probability sampling. In: Wolf C, Joye D, Smith TW, Fu Y (eds) *The SAGE handbook of survey methodology*, pp 329–345
- Yang M, Ganesh N, Mulrow E, Pineau V (2018) Estimation methods for nonprobability samples with a companion probability sample. Joint Statistical Meetings (JSM) Proceedings, Survey Research Methods Section. American Statistical Association, Alexandria

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.