(Non)probability Sampling in Survey Reseach

Inauguraldissertation zur Erlangung des akademischen Grades einer Doktorin der Sozialwissenschaften der Universität Mannheim

Vorgelegt von Hannah Bucher

Hauptamtlicher Dekan der Fakultät für Sozialwissenschaften:

Prof. Dr. Michael Diehl

Erstbetreuer:

Prof. Dr. Florian Keusch

Zweitbetreuer:

PD Dr. Tobias Gummer

Erstgutachter:

Prof. Dr. Florian Keusch

Zweitgutachter:

PD Dr. Tobias Gummer

Drittgutachterin:

Prof. Dr. Bella Struminskaya

Tag der Disputation:

13.09.2023

Acknowledgements

I would like to express my heartfelt gratitude to the numerous people who supported me throughout the journey of writing this dissertation.

Firstly, I extend my sincere appreciation to my supervisors, Florian Keusch and Tobias Gummer, for their invaluable feedback, encouragement, and unwavering support. Their dedication and guidance over the past five years played an instrumental role in the success of this dissertation. Additionally, their enthusiasm for survey research was truly inspirational.

I am also deeply grateful to the GLES team for providing a stimulating work environment and the necessary resources to pursue my research. Joss, thank you for the insightful discussions on methodological questions and for co-authoring one of the papers. Jan, thank you for your assistance with coding issues and for the great time we spent together in the office. Marie and Max, thank you for making the start of my PhD journey memorable. Thanks to the fabulous OSC team, it was great working with you.

My gratitude extends to my GESIS PhD colleagues, particularly to Anne Stroppe for the rewarding two years we spent as PhD representatives together and to Caroline Hahn for the countless hours we spent writing together. I would also like to thank Alex Wenz for his valuable feedback.

I am grateful to Fred Conrad for inviting me to the University of Michigan and to the PhD students from MPSDS for making my stay there enjoyable and productive.

Lastly, I want to express my heartfelt appreciation to my family and friends for their unconditional love and unwavering support. Special thanks to my parents, Kilian (the best son I could ask for), and Alex for always being there for me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction and Summary

1.1 Survey sampling: Origins, basic concepts and new challenges

In 1934, Neyman published an article that laid the foundation of survey research as commonly practiced to this day. His article "On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection" (Neyman, 1934) established probability sampling. In the article, Neyman combined the concepts of survey design and statistical inference (Smith, 1976) by introducing a new statistical inference theory, inference based on confidence intervals (Neyman, 1934). Confidence intervals are intervals in which the values of the estimated population parameters, such as the mean or proportion, are likely to fall. Further, they provide a measure of the uncertainty associated with the estimate (Neyman, 1934). In his article, Neyman (1934) elaborates that confidence intervals would produce reliable inference when applied to repeated samples obtained through the random sampling method, that is selecting elements from the target population at random with known probabilities of selection for each element of the target population (Neyman, 1934; Smith, 1976). This idea paves the path of probability-based sampling. Brick (2011) refers to the article as "paradigm-changing" (Brick, 2011, p. 874). Smith (1976) describes his work as "the Neyman revolution" (Smith, 1976, p.184). It provided a framework for estimating population values based on survey samples. In other words, the idea Neyman published in 1934 ultimately allowed us to estimate the prevalence characteristics in populations of millions of people by surveying only a few hundred or thousands – as long as the sample is randomly drawn. With this, random sampling became the generally accepted framework in survey sampling (Smith, 1976; Groves et al., 2009; Brick, 2011; Bethlehem, 2016).

Although widely considered the standard framework in survey sampling (Groves et al., 2009; Brick, 2011), certain developments nowadays confront probability sampling with new challenges, namely decreasing response rates, increasing survey costs, and the prevalence of the Internet that make the application of probability sampling difficult (Couper, 2000; Brick, 2011; Callegaro et al., 2014; Bethlehem, 2016; Couper, 2017). Next to these challenges, we can observe an increasing usage of sampling approaches not based on randomization in survey practice (Baker et al., 2010). Altogether, these approaches are summarized under "nonprobability sampling". The classification of these alternative sampling methods as nonprobability sampling implies that they are not based on the probability selection of respondents. This terminology does not represent a new conceptualization of survey sampling but rather indicates a deviation from probability sampling. It can be interpreted as an indication of the paradigm of probability sampling in survey research, as it clarifies that nonprobability sampling approaches are judged by probability sampling.

One particular type of nonprobability sampling, self-selection samples recruited from online panels, is most prevalent in survey research nowadays (Baker et al., 2010; Cornesse et al., 2020). Here, instead of being randomly selected, respondents volunteer to participate in surveys by registering on an Internet platform (Marsden and Wright, 2010). These samples allow survey data collection to be fast, easy, and cheap. However, as they are not based on a random selection, they do not fit the framework of probability sampling, and a widely accepted statistical theory that allows for drawing inference is missing (Cornesse et al., 2020).

These developments - the thread to probability sampling on the one and the increasing usage of nonprobability sampling approaches on the other hand – led to a coexistence of probability and nonprobability survey sampling in survey practice nowadays and led to one fundamental question regarding the future of survey sampling: Are we witnessing a framework shift? (Brick, 2011). Given the coexistence of probability and nonprobability sampling in recent survey research and survey sampling´s vague future, this dissertation places emphasis on exploring possibilities that address the limitations of both probability and nonprobability sampling approaches. Its aim is to provide insights and contributions that can inform the development of both, probability and nonprobability survey sampling methods.

In this chapter, I will will briefly introduce the concepts of probability and nonprobability sampling. Further, I will outline current developments in survey sampling and discuss the advantages and disadvantages of probability and nonprobability approaches.

1.1.1 The basic concept of probability sampling

Probability samples are based on randomization theory. The basic assumption of randomization theory is that units are randomly selected with known probabilities from finite populations (Kish, 1995). For this purpose, a sampling frame is a sufficient ingredient to identify and allow access to the target population's elements (Wright and Tsao, 1983; Couper, 2000), e.g., all residential registered addresses. Combined with the concept of confidence intervals introduced by Neyman (Neyman, 1934), this random selection allows for drawing inferences on populations without any additional assumptions on the distribution of variables in the population (Neyman, 1934; Smith, 1976). Therefore, probability sampling is also called a "design-based approach" (Dumelle et al., 2022). If - and only if - selection probabilities for all units in the target population are positive and known, we can accurately estimate parameters in the target population from surveys based on probability samples (Bethlehem, 2016). Further, randomization theory not only allows for precisely estimating parameters in the target population, but it also allows to conclude on possible biases in data collected based on probability samples. As researchers are aware of the whole selection process of individuals from the target population in the sample, they can model the different steps of the dropout of respondents from the sample. With this, it is possible to distinguish between possible sources of error introduced by the sampling process and offer the possibility to describe biases in the sample accurately[1].

1.1.2 The basic concept of nonprobability sampling

For nonprobability samples, we cannot apply randomization theory. Here, researchers are not aware of the sampling process. Instead, individuals self-select into samples rather than being randomly selected. If, as is a widely used approach,

[1]For a detailed description of how, e.g., nonresponse bias can be calculated based on a probability sample, see Bethlehem (2016).

respondents are recruited from opt-in online panels, three main factors influence the self-selection of respondents into samples:

1. Respondents must possess the necessary resources to respond to the survey. As self-selected samples are mainly based on opt-in online panels and surveys based on such samples are conducted online, in practice, respondents need to have access to and actively use the Internet.
2. Possible respondents need to be aware of the survey, which implements they have seen an advertisement for participating in a survey, e.g., on an Internet platform.
3. Respondents must decide to participate in the survey (Bethlehem, 2016).

With these three factors influencing self-selection into samples, we cannot differentiate whether units of the target population are missing in our sample because they do not belong to the Internet population (coverage error), have not received our invitation (sampling error), or do not want to participate in our survey (nonresponse error) (Brick, 2011; Ansolabehere and Rivers, 2013; Bethlehem, 2016; Kohler et al., 2019; Cornesse et al., 2020). Therefore, we can hardly estimate any participation probabilities for respondents and researchers need to rely on distributional assumptions of variables in the target population to apply statistical inference to data derived from nonprobability samples. Therefore, sampling methods not based on randomization theory are also known as "model-based approaches" (Dumelle et al., 2022). Although several research approaches investigate how models that allow to apply inferential statistics to nonprobability samples and derive precise estimates for parameters in the target population based in such samples (Loosveldt and Sonck, 2008; Valliant and Dever, 2011; Ghitza and Gelman, 2013b; Wang et al., 2015; Trangucci et al., 2018; Kennedy and Gelman, 2021), there is no generally accepted statistical theory justifying the usage of nonprobability samples to apply inferences on general populations (Mercer et al., 2017; Cornesse et al., 2020) [2].

To summarize, from a theoretical perspective, probability sampling comes with the benefit of a generally accepted theoretical foundation: randomization theory. This theory allows for (1) applying statistical inference without any additional distributional assumptions about variables in the target population and (2) precisely estimating possible biases in the realized sample. Nonprobability samples, on the contrary, lack this generally accepted theoretical foundation. Due to its basis in self-selection, additional assumptions on the distribution of variables in the target population are needed to enable the application of inferential statistics and to describe biases in surveys based on nonprobability samples precisely.

1.1.3 Current developments in survey research: Challenges and chances for survey sampling

One might expect social scientists to value probability sampling over nonprobability sampling due to its superior theoretical foundation. However, in survey practice, nonprobability sampling is gaining ground. More and more scholars rely on nonprobability samples when conducting surveys (Callegaro et al., 2014). The popularity of nonprobability sampling can be attributed to rapid societal developments in the

[2] For example, Elliott and Valliant (2017) introduce the concepts of quasi-randomization and superpopulation modeling to allow the application of inferential statistics to nonprobability samples. While in quasi-randomization, the inclusion probabilities for every unit of the target population in the sample are modeled (Elliott and Valliant, 2017), in superpopulation modeling, the variable of interest collected with survey data based on nonprobability samples is modeled (Elliott and Valliant, 2017).

last decades affecting survey research, and also pose new challenges to probability sampling.

Table 1.1 gives an overview of the benefits and disbenefits of probability and nonprobability sampling in survey research.

Table 1.1: Comparison of probability and nonprobability sampling

	Probability sampling	Nonprobability sampling
Selection	Random sample	Self-selection
Statistical theory	Random theory	No widely accepted theory
Costs	High	Low
Applicability to Web surveys	Difficult	Easy

Especially the significant shift in communication methods among people, brought about by the prevalence of the Internet and other technological advancements such as smartphones, has led to a surge in survey sampling research. Accompanying this development is an increasing usage of Web surveys in survey research (Couper, 2000; Brick, 2011). By now, collecting survey data on the Web has, in terms of the total number of conducted surveys, outperformed other survey modes (Daikeler et al., 2020). The COVID-19 pandemic has accelerated this trend, as conducting personal interviews was impossible during the crisis phase. Consequently, Web surveys have partially replaced traditional data collection methods, such as face-to-face or telephone surveys (Biffignandi and Bethlehem, 2021). This change in commonly applied data collection methods poses a big challenge for probability sampling and questions its position as a commonly applied framework for sample selection in survey research. The main difficulty in applying probability sampling to Web surveys is the absence of an adequate sampling frame to directly sample individuals on the Web for general population surveys. For Web surveys, a possible frame could, for example, be a list of e-mail addresses for all members of the target population (Couper, 2000). While lists of e-mail addresses are available for particular sub-populations, such as students of a specific University, there is no such frame available for general population surveys (Couper, 2000; Brick, 2011; Bethlehem, 2016). Instead, researchers need to rely on other modes to invite people to Web surveys, such as sending postal invitation letters with an included link to access the Web survey or phone calls to request e-mail addresses for sending invitations. With this, applying probability sampling methods for Web surveys, an increasingly utilized mode of data collection, becomes difficult.

Decreasing response rates pose further challenges to using probability sampling (Baker et al., 2010; Couper, 2017). Fewer and fewer people participate in surveys (Leeper, 2019). Some researchers also worry that low response rates could amplify nonresponse bias (Callegaro et al., 2014; Couper, 2017). Empirical evidence needs to be clarified about the consequences of decreasing response rates for nonresponse bias (Groves, 2006; Groves and Peytcheva, 2008; Wagner, 2012). Instead, it matters more whether nonresponse is systematic or random. In other words, nonresponse introduces bias when survey participation is correlated with target variables. It does not introduce bias when participation is uncorrelated with target variables (Groves, 2006; Groves and Peytcheva, 2008; Wagner, 2012). Nonetheless, some researchers criticize the applicability of inferential statistical methods in probability sampling due to decreasing response rates, which led to a challenge to the dominant position of probability sampling as the main framework of survey sampling (Brick, 2011).

However, even though lower response rates undermine the quality of the survey responses, they still increase the overall survey costs. The decrease in response rates has made fieldwork more challenging and labor-intensive. When only a few individuals are willing to participate, direct interventions during fieldwork are needed to ensure accuracy and completeness (Stoop, 2005; Wolf et al., 2021) and these efforts raise costs (Wolf et al., 2021). As researchers are usually facing cost constraints in the survey design (Groves, 2004), the need for cost-effective alternatives in survey sampling is growing stronger with increasing survey costs.

Altogether, technological innovations, decreasing response rates, and increasing survey costs challenge the status of probability sampling as the standard framework in survey sampling. These challenges provide a fertile ground for the spread of alternative approaches. Nonprobability sampling may address these challenges: Nonprobability sampling allows for exploiting new technologies. This sampling method can be easily applied to Web surveys. Recruited chiefly from online opt-in panels, respondents can be directly invited to participate in a survey on the Web (Baker et al., 2010; Couper, 2017). Further, the main advantage of nonprobability samples is their cost-effectiveness. Compared to probability sample surveys, conducting surveys based on such samples is cheap and easy to implement (Baker et al., 2010; Couper, 2017).

However, the primary drawback of nonprobability sampling, and more specifically sampling respondents from opt-in online panels, is the absence of a solid theoretical foundation to support its use, as no widely accepted statistical theory allows for applying inferential statistics to nonprobability samples (Brick, 2011; Kohler et al., 2019; Cornesse et al., 2020). This lack of theoretical foundation may be why the usage of nonprobability sampling to draw reliable conclusions on populations is, in survey research, perceived to be critical and not generally recommended (Baker et al., 2010; Cornesse et al., 2020).

So, the current probability sampling faces a variety of challenges: The lacking of a frame to directly sample individuals on the Web, decreasing response rates, increasing survey costs, and increased usage of nonprobability samples in survey practices. All of this raises the question of what the future of survey sampling may look like. Will probability sampling remain the standard of survey sampling against which other sampling methods are judged? The past of survey sampling may give us a glimpse into its future. A recent review of the last fifty years of survey sampling (Brick, 2011) identified two main factors that facilitated or hampered change to survey sampling: cost and statistical theory. They summarize:

> "… [C]ost has emerged as the primary agent for changes in sampling methods, even if these have largely been incremental changes. Statistical theory has seldom, if ever, been the leading agent of change. But statistical theory has been essential to supporting new developments. When statistical theory for a sampling method does not garner widespread acceptance for that methodology, then the sampling method is not likely to be accepted across disciplines and applications" (Brick, 2011, p. 879).

In other words, this quote suggests that cost is often the main factor driving changes in sampling methods. However, statistical theory plays a crucial role in supporting and legitimizing these changes. Without a widely accepted statistical theory that supports the inference drawn from survey sampling, change in survey sampling may be prevented.

Transferring these two factors to the question of the future of survey sampling, we see the clear advantage of nonprobability sampling regarding survey costs. However, as nonprobability sampling lacks a widely accepted theoretical theory supporting its application, this does not indicate a fundamental change in survey sampling. Brick (2011) conclude that, although we can be sure that change in survey sampling will happen, we have much less certainty regarding the direction the future will take. Concerning the standard of probability sampling as the general accepted framework in survey sampling and the role nonprobability sampling may play, he concludes:

> "Data collection cost is going to continue to force samplers to examine how they can take advantage of cheaper methods of data collection. A statistical theory that supports collecting observations from the Web from a nonprobability sample is an indispensible ingredient if we are to achieve this much-sought goal. An unresolved question is whether this goal can be accomplished within design-based probability sampling theory. If it is possible, then it is likely that sampling will be invigorated with many new applications and extensions. If not, two outcomes seem realistic: (1) a new paradigm could be introduced that accommodates Web surveys and this theory becomes generally accepted, replacing or supplementing design-based probability sampling; (2) collection of data from volunteers on the Web will be restricted to specific disciplines or applications because of the weak theoretical basis" (Brick, 2011, p. 811).

With these current developments in survey sampling and the assessment of Brick (2011) regarding a possible framework shift from probability to nonprobability sampling, the future of survey sampling mainly depends on finding a statistical theory that allows drawing inferences for data obtained from the Web. Whether this theory fits in the framework of probability sampling or represents it, a new framework remains, by now, open. By now, we can observe a coexistence of probability and nonprobability sampling in survey practice, with both approaches having specific advantages and disadvantages. With this coexistence, we need to (1) explore new ways in how we can adopt probability sampling approaches for Web surveys and reduce their survey costs and survey errors and (2) find ways to make nonprobability sampling approaches more suitable for drawing inferences about populations. This is the aim of this dissertation.

1.2 Why this dissertation?

In the first part of this introductory chapter, I introduced the basic concept of probability and nonprobability sampling. Subsequently, I outlined current developments in survey research that led to a co-existence of these survey sampling approaches in survey practice, featuring some benefits and disadvantages. This dualism illustrates the need for methodological research to address the unresolved issues surrounding the usability of both probability and nonprobability sampling. This dissertation aims to solve the challenges faced by both sampling techniques. In the following, I will position my dissertation within the dual framework of probability and nonprobability sampling and explain its scientific merit. Table 1.2 classifies the issues addressed in this dissertation in the dual framework of probability and nonprobability sampling.

Chapter 2 compares four survey benchmark statistics to answer whether differences associated with different sampling approaches can also be found in empirical analysis

Table 1.2: Research areas this dissertation contributes to in the dual framework of probability and nonprobability sampling

	Probability sampling	Nonprobability sampling
Selection	Chapter 2	Chapter 2 and 4
Statistical theory		Chapter 4
Costs	Chapter 3	
Applicability to Web surveys	Chapter 3	

based on data derived from different survey samples. The surveys I compare differ in mode and sampling: a probability face-to-face survey, a probability self-administered mixed-mode survey, a telephone-recruited nonprobability online survey, and a Web-recruited nonprobability survey. With this approach, this study allows drawing attention to the mode compatibility of probability and nonprobability samples. While probability samples are suitable for different modes of data collection, such as face-to-face, telephone, and mixed-mode surveys but lack applicability to directly sample respondents from the Web, surveys based on such samples are traditionally conducted in different modes. Nonprobability samples, however, are primarily used for Web surveys. With this, the decision for or against a sampling approach also has particular implications for the data collection mode. By comparing four surveys that differ in survey sampling and mode, the current study allows us to collectively analyze different sources of error and investigate to what extent the results of analyses differ between surveys conducted in different modes and based on different samples. Moreover, while comparing various survey types, I contrast surveys that vary in their survey mode and sampling method and their associated costs. As discussed in Chapter 1.1., survey costs represent a crucial aspect to consider when establishing a survey design, in addition to potential sources of error. Given the budgetary limitations researchers often face while conducting a scientific survey (Groves, 2004), survey costs significantly determine the data collection approach (Groves, 2004; Brick, 2011). I compare the four surveys in terms of the estimation of benchmark statistics that refer to voting behavior. Further, I also compare distributions in over 80 variables covering measures of political attitudes and behavior. Lastly, I look at differences in the results of multivariate analyses using a multimodel approach with turnout as the dependent variable. With this approach, I can quantify the comparability of the four surveys and include different theoretical explanations of voting behavior in this comparison.

Chapters 3 and 4 address the deficiencies of probability and nonprobability sampling introduced in Chapter 1.1. In these two chapters, the focus lies on a methodological contribution aimed at resolving issues related to either probability or nonprobability sampling.

Chapter 3 focuses on probability sampling [3]. With this chapter, I aim to contribute to resolving two issues that probability sampling faces today. First, saving survey costs and second, presenting a method to recruit respondents for Web surveys quickly. Here, I explore the feasibility of a probability sampling approach for surveying people directly via smartphones. For this aim, I conducted a case study in Germany. I recruited respondents from a mobile random digit dialing sample via text messages with a Web survey link. This approach shows a new possibility to combine invitation and data collection on one device - the smartphone.

Chapter 4 focuses on nonprobability sampling [4]. It aims to tackle two issues that are associated with nonprobability sampling. First, it addresses bias introduced by the

[3] The study was conducted with Matthias Sand and has already been published in JSSAM (see Bucher and Sand (2022)).

[4] The study was conducted with Joss Roßmann.

self-selection of respondents. Second, it also contributes to elaborating approaches to draw inferences based on nonprobability sampling within the framework of model-based approaches. In this chapter, I focus on enhancing model-based adjustments of nonprobability surveys and propose a selection strategy of adjustment variables that accounts for high correlations of adjustment variables with survey participation and survey questions. Further, I outline the need for empirical examinations on whether the underlying assumptions hold before running post-survey adjustment models. With this, this study contributes to the ongoing discussion on the usability of nonprobability sampling for survey research.

To summarize, the goal of this dissertation s is twofold: (1) To investigate whether disparities related to probability and nonprobability sampling methods can also be identified in empirical analyses that utilize data from different survey samples; (2) To contribute to the methodological enhancement of both, probability and nonprobability sampling.

1.3 Extended summary of chapters

The subsequent sections provide a detailed overview of each primary chapter, along with an extended summary.

1.3.1 Would electoral research show different findings if we replaced probability face-to-face surveys with cheaper alternatives of data collection?

Chapter two compares four surveys that vary in mode and sampling. The analysis focuses on estimating benchmark statistics, distribution patterns across over 80 variables related to political attitudes and behavior, and variations in the results of multivariate analyses. The analyses replicate around 30 studies with individual-level voter turnout as the dependent variable. This approach allows for quantifying the differences in the data results from the four surveys. With this, the present study contributes knowledge on the impact of using survey data collected in different modes and based on different sampling approaches for analyzing individual-level voter turnout.

The data utilized comes from the German Longitudinal Election Study (GLES). I combine data from a probability face-to-face survey (GLES, 2019a), data from a nonprobability online survey recruited via telephone panel (GLES, 2019b), data from a probability self-administered mixed-mode survey (GLES, 2022a), and data from a nonprobability online survey recruited via Web advertisement (GLES, 2022b). Whereas data for the first two surveys were collected before the 2017 German Federal Election, data for the latter were collected before the 2021 German Federal Election. This design allows me to compare surveys based on different samples in terms of their accuracy of point estimates. I will compare only the two surveys with overlapping field periods for uni- and multivariate analysis. However, I also will compare the relative differences between the two surveys across all surveys.

I employ three distinct analytical methods. First, I compare the four surveys in terms of the accuracy of the estimates of external population benchmarks. As a benchmark, I use the official voting statistics for the 2017/2021 German Federal Election. I use survey data weighted with the adjustment weights provided by GLES to calculate point estimates and 95 percent confidence intervals. I calculate the absolute relative bias of each survey that tells about the average difference between the estimates in the surveys and the actual outcome of the election. Second, I investigate differences between variables for which no external benchmarks are available. I compare a set of 89 variables for the 2017 data and 83 for the 2021 data consisting of sociodemographic characteristics, political attitudes, and political behavior. Third, I examine the differences between the surveys in multivariate analyses through a multimodel comparison. For this purpose, I replicate models with individual-level voter turnout as the dependent variable. Models for replication were selected based on a meta-analysis of individual-level voter turnout conducted by Smets and van Ham (2013). To investigate whether the results of the multivariate analyses differ between the surveys, I first compare whether the calculated models differ in their accuracy by comparing whether the predicted values of models coincide with the observed outcome. For this purpose, I use receiver operator statistics (ROC) as a statistical measure. Further, I calculate average marginal effects (AMEs) for all regression models to compare results from the surveys in terms of the associations of variables in multivariate models. I investigate whether the AMEs of the calculated models differ substantially between the surveys.

The findings reveal several vital observations: 1. My analysis consistently demonstrates that the probability face-to-face survey is the most accurate regarding point estimates, indicating that survey sampling and survey mode affect the accuracy of point estimates. 2. Significant differences exist between the probability and non-probability surveys concerning the distributions of various variables. 3. The results suggest that collecting data via surveys with varying modes and sampling leads to different outcomes across multiple regression models and variable associations.

However, the study only compares two surveys in each instance, thus not enabling definitive conclusions regarding their relative performance. Nonetheless, the findings consistently demonstrate that both sampling and mode are influential in this regard. As a result, this chapter provides insights into how survey sampling and mode decisions affect results derived from different surveys. It further examines that decisions on the survey design concerning survey mode and survey sampling have far-reaching implications for concluding the drivers of individual-level voter turnout and voting behavior.

1.3.2 Exploring the feasibility of recruiting respondents and collecting Web data via smartphone: A case study of text-to-Web recruitment for a general population survey in Germany[5].

The third chapter of this dissertation sets up an innovative approach to survey people directly on their smartphones. The widespread use of smartphones has revolutionized how we communicate and access the Internet, making it an ideal platform for survey research. With most of the population now reachable via smartphones, combining

[5]The study was conducted with Matthias Sand and has already been published in JSSAM (see Bucher and Sand (2022)).

text messages for recruitment and direct surveying via smartphone presents new possibilities for survey research. This approach was investigated in a case study conducted in Germany in November 2018. The text-to-Web approach we used in this study can be described in three steps: First, randomly sampling German cellphones via Random Digit Dialing (RDD) and second, sending invitation text messages to the generated numbers with the link to the Web survey and third, data collection.

A central finding of our study is that recruiting respondents for Web surveys via text messaging to smartphones is feasible. However, this approach is hampered by numerous issues relating to its implementation and the resulting data. Although RDD mobile sampling is easy and quick to implement, it may introduce biases due to the exclusion of certain network providers and limitations in generating a random sample of only smartphone numbers. Text message invitations are fast and relatively cheap. Nevertheless, the high rate of undelivered messages and low willingness to participate may impact data quality. The article suggests that further research is needed to address these challenges and experiment with different invitation designs and content to increase participation. Still, our study is the first to explore and demonstrate the different stages of conducting a text-to-Web survey combined with mobile RDD sampling. We have provided a basis for future research in this area by highlighting the various steps and potential challenges.

1.3.3 Enhancing model-based adjustments of nonprobability surveys: Selecting auxiliary variables based on theoretical assumptions about their association with survey participation and variables of interest[6].

Chapter four proposes a selection strategy of adjustment variables that accounts for high correlations of adjustment variables with survey participation and survey questions. This chapter further shows how it can be empirically checked whether the calculated post-survey adjustment with the selected variables can reduce selection bias in nonprobability surveys. This endeavor results in a six-step approach that survey researchers can quickly implement.

To empirically demonstrate our approach, we conducted two case studies: In both studies, we conducted surveys on political attitudes and behavior in Germany based on a nonprobability sample from a German opt-in online panel. We applied the six-step approach in both studies to select influential adjustment variables. To investigate whether the adjustment based on the theoretically selected variables helps in reducing bias in our surveys, we compared estimates for the outcome of two elections in Germany from (1) unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights.

Although promising in theory, we saw that the approach did not perform as desired concerning reducing bias in estimates obtained from the nonprobability survey data in both cases. Our results consistently showed that adjusting the data from

[6]The study was conducted with Joss Roßmann.

nonprobability online surveys with enhanced, theory-based weights did not substantively reduce selection biases in estimates. For both studies, we assume a reasonable explanation for the poor performance of the post-survey adjustment based on the theoretically informed selection of covariates in the weak correlations of these variables with (1) survey participation and (2) political attitudes and behavior.

Moreover, adjusting the data with standard socio-demographic weights amplifies the absolute relative bias in estimates of voting behavior. We take this finding as a warning that misspecification in adjustment models can exacerbate biases in estimates and lead to erroneous conclusions.

References

Ansolabehere, Stephen, and Douglas Rivers. 2013. "Cooperative Survey Research." *Annual Review of Political Science* 16 (1): 307–29. https://doi.org/10.1146/annurev-polisci-022811-160625.

Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. Dillman, et al. 2010. "Research Synthesis: AAPOR Report on Online Panels." *Public Opinion Quarterly* 74 (4): 711–81. https://doi.org/10.1093/poq/nfq048.

Bethlehem, Jelke. 2016. "Solving the Nonresponse Problem With Sample Matching?" *Social Science Computer Review* 34 (1): 59–77. https://doi.org/10.1177/0894439315573926.

Biffignandi, Silvia, and Jelke G. Bethlehem. 2021. *Handbook of Web Surveys*. Second edition. Hoboken, NJ: Wiley.

Brick, J. M. 2011. "The Future of Survey Sampling." *Public Opinion Quarterly* 75 (5): 872–88. https://doi.org/10.1093/poq/nfr045.

Bucher, Hannah, and Matthias Sand. 2022. "Exploring the Feasibility of Recruiting Respondents and Collecting Web Data via Smartphone: A Case Study of Text-To-Web Recruitment for a General Population Survey in Germany." *Journal of Survey Statistics and Methodology* 10 (4): 886–97. https://doi.org/10.1093/jssam/smab006.

Callegaro, Mario, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas, eds. 2014. *Online Panel Research*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118763520.

Cornesse, Carina, Annelies G Blom, David Dutwin, Jon A Krosnick, Edith D De Leeuw, Stéphane Legleye, Josh Pasek, et al. 2020. "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research." *Journal of Survey Statistics and Methodology* 8 (1): 4–36. https://doi.org/10.1093/jssam/smz041.

Couper, Mick P. 2000. "Web Surveys." *Public Opinion Quarterly* 64 (4): 464–94. https://doi.org/10.1086/318641.

———. 2017. "New Developments in Survey Data Collection." *Annual Review of Sociology* 43 (1): 121–45. https://doi.org/10.1146/annurev-soc-060116-053613.

Daikeler, Jessica, Michael Bošnjak, and Katja Lozar Manfreda. 2020. "Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates." *Journal of Survey Statistics and Methodology* 8 (3): 513–39. https://doi.org/10.1093/jssam/smz008.

Dumelle, Michael, Matt Higham, Jay M. Ver Hoef, Anthony R. Olsen, and Lisa Madsen. 2022. "A Comparison of Design-Based and Model-Based Approaches for Finite Population Spatial Sampling and Inference." *Methods in Ecology and Evolution*, June, 2041–210X.13919. https://doi.org/10.1111/2041-210X.13919.

Elliott, Michael R., and Richard Valliant. 2017. "Inference for Nonprobability Samples." *Statistical Science* 32 (2): 249–64. https://doi.org/10.1214/16-STS598.

Ghitza, Yair, and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups." *American Journal of Political Science* 57 (3): 762–76. https://doi.org/10.1111/ajps.12004.

GLES. 2019a. "Langfrist-Online-Tracking T37 (GLES)." https://doi.org/10.4232/1.13295.

———. 2019b. "Vorwahl-Querschnitt (GLES 2017)." https://doi.org/10.4232/1.13234.

GLES. 2022a. "GLES Cross-Section 2021, Pre-Election GLES Querschnitt 2021, Vorwahl." GESIS Data Archive. https://doi.org/10.4232/1.13860.

———. 2022b. "GLES Tracking September 2021, T50 GLES Tracking September 2021, T50." GESIS. https://doi.org/10.4232/1.14000.

Groves, Robert M. 2004. *Survey Errors and Survey Costs.* Repr. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley-Interscience.

———. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70 (5): 646–75. https://doi.org/10.1093/poq/nfl033.

Groves, Robert M., Floyd J. Fowler, Mick Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau, eds. 2009. *Survey Methodology.* 2. ed. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley.

Groves, Robert M., and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72 (2): 167–89. https://doi.org/10.1093/poq/nfn011.

Kennedy, Lauren, and Andrew Gelman. 2021. "Know Your Population and Know Your Model: Using Model-Based Regression and Poststratification to Generalize Findings Beyond the Observed Sample." *Psychological Methods* 26 (5): 547–58. https://doi.org/10.1037/met0000362.

Kish, Leslie. 1995. *Survey Sampling.* A Wiley Interscience Publication. New York: Wiley.

Kohler, Ulrich, Frauke Kreuter, and Elizabeth A. Stuart. 2019. "Nonprobability Sampling and Causal Analysis." *Annual Review of Statistics and Its Application* 6 (1): 149–72. https://doi.org/10.1146/annurev-statistics-030718-104951.

Leeper, Thomas J. 2019. "Where Have the Respondents Gone? Perhaps We Ate Them All." *Public Opinion Quarterly* 83 (S1): 280–88. https://doi.org/10.1093/poq/nfz010.

Loosveldt, Geert, and Nathalie Sonck. 2008. "An Evaluation of the Weighting Procedures for an Online Access Panel Survey." *Survey Research Methods* 2 (2): 93–105. https://doi.org/10.18148/srm/2008.v2i2.82.

Marsden, Peter V., and James D. Wright. 2010. *Handbook of Survey Research.* 2nd ed. Bingley, UK: Emerald.

Mercer, Andrew W., Frauke Kreuter, Scott Keeter, and Elizabeth A. Stuart. 2017. "Theory and Practice in Nonprobability Surveys." *Public Opinion Quarterly* 81 (S1): 250–71. https://doi.org/10.1093/poq/nfw060.

Neyman, Jerzy. 1934. "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection." *Journal of the Royal Statistical Society* 97 (4): 558. https://doi.org/10.2307/2342192.

Smets, Kaat, and Carolien van Ham. 2013. "The Embarrassment of Riches? A Meta-Analysis of Individual-Level Research on Voter Turnout." *Electoral Studies* 32 (2): 344–59. https://doi.org/10.1016/j.electstud.2012.12.006.

Smith, T. M. F. 1976. "The Foundations of Survey Sampling: A Review." *Journal of the Royal Statistical Society. Series A (General)* 139 (2): 183. https://doi.org/10.2307/2345174.

Stoop, Ineke A. L. 2005. *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. SCP-report 2005,8. The Hague: Social and Cultural Planning Office of the Netherlands.

Trangucci, Rob, Imad Ali, Andrew Gelman, and Doug Rivers. 2018. "Voting Patterns in 2016: Exploration Using Multilevel Regression and Poststratification (MRP) on Pre-Election Polls." arXiv. https://doi.org/10.48550/ARXIV.1802.00842.

Valliant, Richard, and Jill A. Dever. 2011. "Estimating Propensity Adjustments for Volunteer Web Surveys." *Sociological Methods & Research* 40 (1): 105–37. https://doi.org/10.1177/0049124110392533.

Wagner, J. 2012. "A Comparison of Alternative Indicators for the Risk of Nonresponse Bias." *Public Opinion Quarterly* 76 (3): 555–75. https://doi.org/10.1093/poq/nfs032.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31 (3): 980–91. https://doi.org/10.1016/j.ijforecast.2014.06.001.

Wolf, Christof, Pablo Christmann, Tobias Gummer, Christian Schnaudt, and Sascha Verhoeven. 2021. "Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys." *Public Opinion Quarterly* 85 (2): 623–48. https://doi.org/10.1093/poq/nfab039.

Wright, Tommy, and How J. Tsao. 1983. "A FRAME ON FRAMES: AN ANNOTATED BIBLIOGRAPHY." In *Statistical Methods and the Improvement of Data Quality*, 25–72. Elsevier. https://doi.org/10.1016/B978-0-12-765480-5.50008-4.

2 Would Electoral Resarch Show Different Findings if we Replaced Probability Face-to-Face Surveys with Cheaper Alternatives of Data Collection?

Abstract

In this paper, I compare four surveys that differ in mode and sampling by the German Longitudinal Election Study (GLES), namely a probability face-to-face survey, a probability self-administered mixed mode survey, a nonprobability online survey recruited via telephone interviews and a nonprobability online survey recruited via Web advertisement in terms of estimation of benchmark statistics; distributions in over 80 variables covering measures of political attitudes and behavior; and differences in results of multivariate analyses replicating ~30 studies with individual-level

voter turnout as the dependent variable. The probability face-to-face survey performs best in estimating characteristics with external benchmarks. Further, I found substantial differences in uni-and multivariate analysis between the surveys. Thus, switching from a probability face-to-face survey to another mode and/or sample for data collection affects empirical findings on individual-level voter turnout and the conclusions drawn therefrom.

2.1 Introduction

National election studies are characterized by methodological diversity. They are conducted in a telephone-, face-to-face-, or, more recently, online mode. However, face-to-face surveys based on a random selection of individuals in a population - also known as probability-based personal surveys - are still considered the gold standard in survey research (Baker et al., 2010; Callegaro et al., 2014; Cornesse et al., 2020). They are conducted in national election studies across a wide range of countries, for example, in the American National Election Studies (ANES, 2017), the British Election Study (Fieldhouse et al., 2022), and the German Longitudinal Election Study (GLES, 2019a).

However, due to decreasing response rates and increasing survey costs, collecting survey data via personal interviews based on probability samples becomes increasingly challenging. The COVID-19 pandemic compounded this problem, as conducting personal interviews was impossible during the crisis phase. Therefore, many national election studies have recently changed their data collection strategies by first, switching from a face-to-face mode to self-administered modes of data collection such as online and mail surveys (ANES, 2021; GLES, 2022a) and second, started complement their data collection strategies with nonprobability surveys (Sanders et al., 2007; Dassonneville et al., 2018; Wagner et al., 2018; GLES, 2019b).

The main benefit of these approaches is that they reduce costs of data collection. However, the discussion is ongoing as to whether changes in survey design, such as shifting from face-to-face to self-administered modes (de Leeuw, 2010) or from probability to nonprobability sampling, comes at the price of non-ignorable biases in estimating variables of interest (Baker et al., 2010; Callegaro et al., 2014; Cornesse et al., 2020). The current study aims to investigate whether electoral research would show different findings if we replaced probability face-to-face surveys with cheaper data collection approaches.

In this study, I focus on voter turnout. For this aim, I compare four different surveys from the German Longitudinal Election Study (GLES), differing in survey mode and sampling, as well as ranging from expensive to cheap in total data collection costs, regarding differences in accuracy of point estimates, as well as variables that are commonly applied to study voting turnout. I use data from a probability face-to-face survey, a probability self-administered mixed mode survey, and two different nonprobability online surveys - one that recruited participants via telephone, and one that recruited participants via advertisement on the Internet. With this approach, the current study allows to collectively analyze different sources of error, namely survey sampling and survey mode and investigate to what extent results of analyses of individual-level voter turnout differ between surveys conducted in different modes and based on different samples.

Further, I use three different analytical approaches with different statistical measures to compare a large set of variables commonly used to study voting turnout that reflect several broad theoretical models of individual-level voter turnout (Smets and van Ham, 2013). This allows for quantifying the differences in the results of data from the four surveys. With this approach, the present study makes an important contribution to knowledge on the impact of using survey data collected in different modes and based on different sampling approaches for the analysis of individual-level voter turnout.

The results show, first, that the probability face-to-face survey performs best in terms of the accuracy of estimates for population parameters. Second, there are substantial differences in the distributions of most variables for which no benchmark data are available between the probability and the nonprobability surveys. Third, many of the associations between variables differ in their significance and their direction. However, I find no conclusive evidence that the type of survey affects the goodness of fit of multivariate models.

2.2 Background

As different potential sources of error are mixed when comparing data collected using probability face-to-face surveys, probability self-administered mixed mode surveys, and nonprobability online surveys, this may lead to differing estimates. These potential sources of error are survey sampling and survey mode.

2.2.1 Survey sampling

2.2.1.1 Survey sampling in probability surveys

Probability surveys have in common that their samples are drawn by randomly selecting units from a finite population (Kish, 1995; Bethlehem, 2009; Groves et al., 2009; Ansolabehere and Rivers, 2013). Thus, the probability of receiving an invitation to participate in the survey can be determined for every member of the target population. Therefore, it is possible to apply inferential statistics and draw conclusions regarding possible biases in the estimates. From a theoretical point of view, probability surveys are thus suitable for making inferences about general populations. As they are based on randomization theory, they are also referred to as "design-based approaches" (Dumelle et al., 2022).

However, probability surveys have potential sources of bias[1], the most significant of which is that the likelihood of the selected units responding to the survey cannot be controlled (Kish, 1995; Bethlehem, 2009, 2016). Nonresponse is therefore the main source of potential bias in probability surveys (Groves, 2006; Groves and Peytcheva, 2008; Bethlehem, 2009; Groves et al., 2009; Bethlehem, 2016). As nonresponse in probability surveys has been steadily increasing in recent years (Callegaro et al., 2014; Couper et al., 2017), the risk of nonresponse bias is also increasing.

However, studies investigating the relationship between decreasing response rates and nonresponse bias indicate that increasing nonresponse contributes only to a limited extent to an increase in nonresponse bias (Groves, 2006; Groves and Peytcheva, 2008; Wagner, 2012). What matters is not so much the presence of nonresponse, but rather

[1]For a useful summary of potential sources of bias, see Biemer (2010).

whether it is systematic - that is, correlated with the target variables - or random - that is, independent of these variables (Groves, 2006; Groves and Peytcheva, 2008; Wagner, 2012). Further, current findings suggest that in most probability surveys, bias caused by nonresponse is not sufficiently large to render inferential statistical methods inapplicable (Groves, 2006; Groves and Peytcheva, 2008; Wagner, 2012). Thus, probability surveys are still considered the gold standard in survey research (Baker et al., 2010; Callegaro et al., 2014; Cornesse et al., 2020).

2.2.1.2 Nonprobability surveys

The main criticism of nonprobability surveys relates to the generalizability of their findings to general populations (Baker et al., 2010; Callegaro et al., 2014; Cornesse et al., 2020). This criticism is because randomization theory is not applicable to nonprobability surveys (Brick, 2011; Ansolabehere and Rivers, 2013; Bethlehem, 2016; Cornesse et al., 2020), as the fundamental assumption of randomization theory - namely, that one has a finite population from which to select a sample with known probabilities for each population unit - does not hold (Ansolabehere and Rivers, 2013; Bethlehem, 2016). Meeting this assumption constitutes the fundamental criterion for applying inferential statistics to survey data, and thus for drawing conclusions about populations on the one hand, and about possible biases in one's estimates on the other.

However, efforts have been made to develop models that allow inferences to be made about general populations based on nonprobability samples (Loosveldt and Sonck, 2008; Valliant and Dever, 2011; Ghitza and Gelman, 2013a; Wang et al., 2015; Trangucci et al., 2018; Kennedy and Gelman, 2021). Although recent research indicates that the aforementioned modeling approaches show promise in enhancing the generalizability of results of nonprobability surveys (Wang et al., 2015; Kennedy and Gelman, 2021), other studies have found that they failed to minimize biases in estimates (Loosveldt and Sonck, 2008; Schonlau et al., 2009; Valliant and Dever, 2011). These mixed results can be attributed to the fact that their ability to reduce bias depends greatly on influential covariates that are often unavailable or unknown (Park et al., 2004; Schonlau et al., 2009; Valliant and Dever, 2011).

In summary, the advantage of nonprobability surveys is that they are considerably less expensive than probability surveys. From a theoretical perspective, however, probability surveys are preferable, especially when it comes to making inferences about general populations.

2.2.2 Survey mode

2.2.2.1 Face-to-face surveys

Face-to-face surveys belong to the category of interviewer administered surveys. They are characterized through an Interviewer visiting the respondent to conduct a survey. With this approach, face-to-face surveys allow for conducting complex surveys that take a long time to answer since the interviewer guides the respond through the question program (Heerwegh and Loosveldt, 2008; Groves et al., 2009). Furthermore, empirical studies show that conducting surveys face-to-face results in lower break-off rates, a better sample balance, helps in reducing satisfying and generally improves the cooperativeness of survey respondents compared to other survey modes, such as telephone surveys, or more recently, online surveys (Holbrook et al., 2003; Heerwegh and Loosveldt, 2008; Neuman, 2012).

However, conducting surveys in the face-to-face mode as personal interviews results in high costs for data collection. Especially during the last years that were characterized by low response rates, the costs per interview for face-to face surveys increased (Neuman, 2012). The increase in costs is related to the efforts that must be made by survey researchers to increase the willingness to participate in surveys, such as re-contact individuals that did not respond to initial contacts or increase incentives for hard-to-reach populations (Neuman, 2012; Wolf et al., 2021). Further, as most countries were affected by a global pandemic in the last two years, conducting personal interviews was not applicable in many countries at this time.

Therefore, many surveys and with this also many national election studies shifted the data collection modes from face-to-face surveys to self-administered surveys such as mixed mode online and mail surveys (Wolf et al., 2021; GLES, 2022a), or online only surveys (ANES, 2021).

2.2.2.2 Self-administered (online and mail) surveys

In self-administered mixed mode online and mail surveys, respondents can choose whether they want to answer the questionnaire in an online, or a mail mode. With combining these two kinds of survey mode, researchers want to ensure individuals that do not or only occasionally use the Internet can participate in the survey. As current research indicates that being reachable online correlates with specific characteristics (Bandilla et al., 2009; Mohorko et al., 2013; Sterrett et al., 2017), the offer of the mail mode should prevent coverage error. Further, offering individuals multiple modes to conduct the survey may increase response rates in self-administered surveys (Millar and Dillman, 2011).

While the popularity of self-administered surveys increased in recent years, shifting from interviewer administered to self-administered data collection may influence the response to questions (Dillman and Christian, 2005; Cernat and Sakshaug, 2020; Olson et al., 2021a). The absence of an interviewer, as well as differences in the presentation of questions in an online or mail questionnaire may activate other stimuli when answering the question (Dillman and Christian, 2005; Gideon, 2012; Cernat and Sakshaug, 2020; Olson et al., 2021a). Further, self-administered survey mode may also result in selection effects, that means an influence of survey mode on sample composition (Struminskaya et al., 2016).

However, compared to face-to-face surveys, self-administered mixed mode surveys offer the potential to safe costs in data collection and with this allow to conduct a larger number of interviews within the same budget than face-to-face surveys (Wolf et al., 2021).

2.2.2.3 Online only surveys

Online only surveys belong to the category of self-administered surveys. Different to self-administered mixed mode surveys, only individuals can take part in online only surveys that use the Internet. With this, these surveys suffer from coverage error, as non-Internet users are systematically excluded from participation (Bandilla et al., 2009; Mohorko et al., 2013; Sterrett et al., 2017). The main benefit of online only surveys is that they are easy to implement and among the aforementioned modes the cheapest data collection mode (Biffignandi and Bethlehem, 2021). Further, they allow researchers a lot of flexibility in questionnaire design, as multimedia tools can be embedded in the questionnaire (Biffignandi and Bethlehem, 2021).

In this study, I collectively analyze these different sources of error, and investigate to what extent results analyzing individual-level voter turnout differ between different types of surveys, namely a probability face-to-face survey, a probability self-administered mixed mode survey, and two nonprobability online surveys.

In this study I focus merely on survey error. However, when comparing different types of surveys, I do not only compare surveys that differ in their survey mode and survey sampling, but also in their survey costs. From a practical perspective an important consideration when setting up a survey design, next to potential sources of error, are survey costs. When conducting a scientific survey, researchers are often limited by cost constraints (Groves, 2004). Therefore, survey costs are an important factor when making decisions about data collection design (Groves, 2004; Brick, 2011). The current study relates the possible sources of error to the cost of conducting a survey and with this, allows a reflection of both, survey errors and costs.

2.3 Literature review

A growing body of literature investigates differences in empirical results between different types of surveys in the field of electoral research (Berrens et al., 2003; Malhotra and Krosnick, 2007; Sanders et al., 2007; Chang and Krosnick, 2009; Stephenson and Crete, 2011; Yeager et al., 2011; Ansolabehere and Schaffner, 2014; Bytzek and Bieber, 2016; Pasek, 2016; Breton et al., 2017; Dassonneville et al., 2018). For the most part, these studies compare nonprobability online surveys with different modes of probability surveys in terms of different statistical analyses. Thus, they focus merely on sampling-related differences in estimates. Three different analysis strategies can be identified in these studies: comparison of (1) the accuracy of point estimates by comparing the estimates of the samples with external population benchmarks; (2) differences in distributions and means of attitudinal and behavioral variables; and (3) differences in associations of variables between surveys.

Regarding the accuracy of point estimates, previous studies mostly show that probability surveys perform substantially better than nonprobability surveys (Malhotra and Krosnick, 2007; Chang and Krosnick, 2009; Stephenson and Crete, 2011; Yeager et al., 2011; Ansolabehere and Schaffner, 2014; Bytzek and Bieber, 2016; Kennedy et al., 2016; Dassonneville et al., 2018). Some studies suggest that nonprobability surveys perform worse than probability surveys when estimating the results of elections - especially party vote (Ansolabehere and Schaffner, 2014; Bytzek and Bieber, 2016; Dassonneville et al., 2018). However, other studies show that probability surveys do not consistently meet the valid population parameters with higher accuracy than nonprobability surveys (Sanders et al., 2007; Pasek, 2016; Breton et al., 2017).

The evidence is mixed regarding distributions of attitudinal variables and variables measuring political behavior. Some studies suggest that the means of political knowledge and political interest are higher in nonprobability online surveys than in probability surveys (Berrens et al., 2003; Malhotra and Krosnick, 2007; Chang and Krosnick, 2009; Ansolabehere and Schaffner, 2014). Others have found differences between some variables without discovering substantial patterns (Sanders et al., 2007; Stephenson and Crete, 2011; Yeager et al., 2011; Breton et al., 2017).

Regarding associations between variables, most of the aforementioned studies suggest that the type of survey sample and mode used for analysis does not have an impact on the associations in general. However, some studies show that the strength of these associations differs between probability and nonprobability surveys (Sanders et al., 2007; Yeager et al., 2011; Pasek, 2016). For multivariate models, most of the studies considered in this review show that conclusions drawn based on results achieved with different surveys are comparable (Berrens et al., 2003; Sanders et al., 2007; Chang and Krosnick, 2009; Stephenson and Crete, 2011; Yeager et al., 2011; Ansolabehere and Schaffner, 2014; Bytzek and Bieber, 2016; Breton et al., 2017; Dassonneville et al., 2018). However, one study on the effect of survey mode and sampling on inferences about political attitudes and behavior (Malhotra and Krosnick, 2007) found significant differences between surveys in multivariate analyses in some of the calculated models. This led the authors to conclude that "researchers interested in assuring the accuracy of their findings should rely on face-to-face surveys of probability samples rather than Internet samples of volunteer respondents" (Malhotra and Krosnick, 2007, p. 286).

As this literature review shows, previous studies have yielded mixed results regarding the comparability of results obtained with nonprobability online surveys and probability-based surveys in different modes in electoral research. In sum, these studies indicate:

1. Probability surveys perform better in terms of estimating valid population parameters.
2. Distributions of individual variables for which no external benchmarks are available may differ between surveys.

3. Surveys are comparable mainly in their findings regarding associations between variables.

Using individual-level voter turnout as the dependent variable, the current study aims to extend this research by replicating a wide range of multivariate models from studies published in high-ranked political science journals (Smets and van Ham, 2013) in analyses of data from surveys conducted across different modes and based on different samples.

2.4 Data

In this study, I use four German Longitudinal Election Study (GLES) surveys: the GLES Pre-Election Cross-Sections, two probability surveys based on a register sample. While this survey was conducted in a face-to-face mode on the occasion of the German Federal Election 2017 (GLES, 2019a), the GLES shifted the data collection mode to a self-administered mixed mode design in 2021 (GLES, 2022a); and GLES Tracking, an online survey based on a quota sample from a commercial German online opt-in panel provider. While this survey was conducted by means of an online opt-in online panel provider (forsa.Omninet) that has recruited its members via a telephone in 2017 (GLES, 2019b), GLES changed the provider (Respondi.AG) and in the survey conducted in 2021 respondents were recruited with advertisement on different Internet platforms (GLES, 2022b). With this, the two online access panels may differ in their composition. However, as both surveys rely on a quota selection of members from the online access panel, they both belong to the group of nonprobability online surveys. The target population of all surveys was German citizens aged

18 years and older[2]. Data were collected in the weeks preceding the 2017 or 2021 German Federal Election.

2.4.1 Comparability of the surveys

Figure 2.1 gives an overview of the four surveys used in this study. While the probability face-to-face survey and the nonprobability online survey based on a telephone recruited online panel were conducted prior to the 2017 German Federal Election, the probability self-administered mixed mode survey and the nonprobability online survey based on Web advertisement recruitment were conducted prior to the 2021 German Federal Election. All surveys aimed to capture political attitudes and behavior during the 2017/2021 German Federal Election campaign in order to gain a deeper understanding of voting behavior and turnout.

While field periods of the two 2017 surveys partially overlapped, the field period for the GLES Pre-Election Cross-Section was longer. This can be attributed to the fact that face-to-face surveys require much more complex field management and take longer to conduct. For the 2021 surveys, the differences in field period have become smaller. However, here the survey was conducted in a mixed mode online and mail design and mail surveys require more time since they must be sent back by post.

With this design, I can compare the four surveys directly in terms of their accuracy of point estimates. For uni- and multivariate analysis, I will compare only the two

[2] In the probability surveys, people aged 16 years and over were interviewed. However, persons who were not yet 18 years old at the time of data collection were excluded from the present analyses to ensure comparability with the nonprobability online surveys.

surveys with overlapping field periods. However, I also will compare the relative differences between the two surveys each across all surveys.

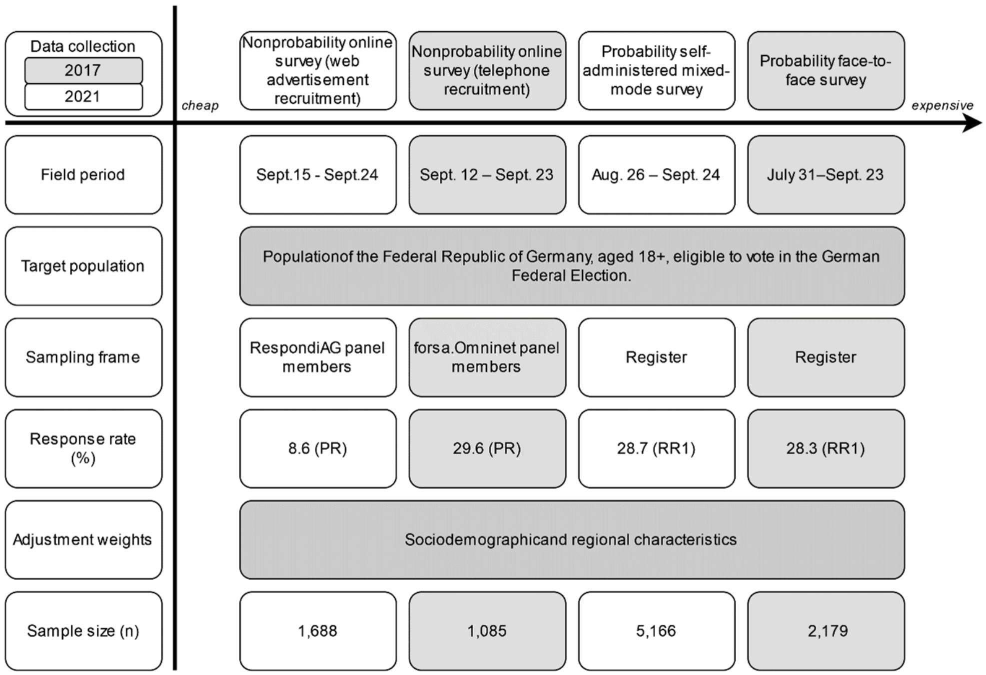

Figure 2.1: Description of the datasets

2.5 Methods

2.5.1 Analyzing the accuracy of point estimates in the surveys

In this first step of the data analysis, I compare the probability face-to-face survey, the probability self-administered mixed mode survey and the two nonprobability online surveys in terms of the accuracy of the estimates of external population benchmarks. As a benchmark, I use the official voting statistics for the 2017/2021 German Federal

Election provided by the Federal Election Commission of Germany (Kobold and Schmiedel, 2018; Bundeswahlleiter, 2021). These data contain information on voter turnout and party vote, differentiated by gender, age group, and region for adults eligible to vote in Germany.

I use survey data weighted with the adjustment weights provided by GLES to calculate point estimates and 95 percent confidence intervals. I calculate the absolute relative bias of each survey that tells about the average difference between the estimates in the surveys and the actual outcome of the election. Mathematically, the absolute relative bias is defined as:

$$AbsoluteRelativeBias(Y) = \sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)}{n}$$

, where

$$n$$

denotes the data points on all variables considered,

$$y_i$$

denotes the values of these variables observed in the target population, and

$$\hat{y}_i$$

denotes the values of these variables estimated by the surveys. To calculate the absolute relative bias, I dichotomized the variables.

2.5.2 Comparing variables and their distribution

In a second step, I investigate differences between variables for which no external benchmarks are available. Since at this step of data analysis I compare the surveys directly with each other, I will compare each two of the surveys with overlapping field periods. In this comparison, I use all individual variables in the multivariate models as independent variables. In total, I compare a set of 89 variables for the 2017 data and 83 variables for the 2021 data consisting of sociodemographic characteristics, political attitudes, and political behavior.

This large set of variables includes variables with different scales. To achieve comparability between these variables, I rescaled from 0 to 1 all variables that are at least ordinal-scaled, and I dichotomized nominally scaled variables. I compare the variables in terms of their means and distributions with different measures. To examine whether the distributions of the at least ordinal-scaled variables differ between the two surveys, I compute the Kolmogorov-Smirnov test for differences in distributions. For the dichotomized variables, I compute the test for equality of proportions. These tests provide information on whether the distributions/proportions differ significantly between the two surveys, but not on the extent of this difference. Therefore, I further calculate the effect size using Cohen's d. I use the Stata command "ksmirnov" to test for differences in distributions and the Stata command "prtest" to test for equality of proportions. Cohen's d is computed with the Stata command "esize."

2.5.3 Comparing associations between variables in multivariate models and the models' goodness of fit

2.5.3.1 Model selection process

In a third step, I examine the differences between the surveys in multivariate analyses through a multimodel comparison. For this purpose, I replicate models with individual-level voter turnout as the dependent variable (Green and Shachar, 2000; Heath, 2000; Lyons and Alexander, 2000; Highton and Wolfinger, 2001; Holbrook et al., 2001; Clarke et al., 2002; Goldstein and Freedman, 2002; Mughan and Lacy, 2002; Mutz, 2002; Perea, 2002; Anduiza–Perea, 2005; Jackson, 2003; Blais et al., 2004; Rubenson et al., 2004; Chong and Rogers, 2005; Adams et al., 2006; Leighley and Nagler, 2007; Malhotra and Krosnick, 2007; Sanders et al., 2007; Wass, 2007; Killian et al., 2008; Stevens et al., 2008; Pattie and Johnston, 2009; Yoo, 2010). Models for replication were selected based on a meta-analysis of individual-level voter turnout conducted by Smets and van Ham (2013). This meta-analysis reviews 90 empirical studies of individual-level voter turnout in national elections published in high-ranked political science journals.

I examined whether I could replicate the models calculated in these 90 studies with GLES data. I included those models in my analysis, for which at least the predictors in the original studies could be calculated with GLES data. In doing so, I followed the operationalization of variables in the original studies as closely as possible. However, the implementation of some variables and concepts differs significantly between data used in the original studies and GLES data. Therefore, it was possible to conceptually replicate only some of the studies. As the aim of the present analysis is not to explain individual-level voter turnout, but rather to investigate whether results of multivariate analyses differ across a wide range of possible explanatory factors of voter turnout between surveys that differ in mode and sampling, I consider conceptual replication to be sufficient for my purposes. I could calculate the predictors for 29 models with the 2017 data sets and for 27 models with the 2021 data sets of the 90 studies considered by Smets and van Ham (2013) in their meta-analysis[3]. Figure 2.2 gives an overview of the process of selection of models for inclusion in the analysis.

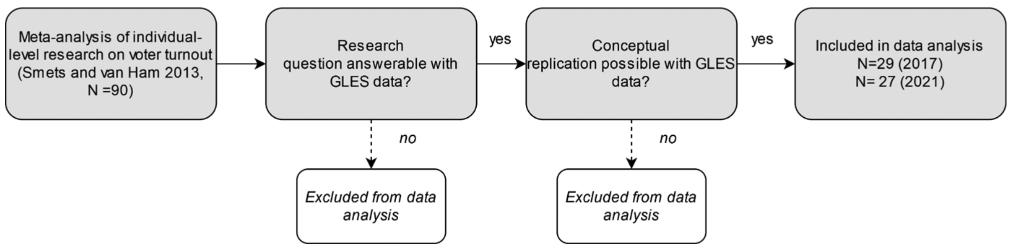

Figure 2.2: Model selection process

In their meta-analysis, Smets and van Ham (2013) divided the reviewed studies into "broad theoretical models" of individual-level voter turnout that they felt "reflect the main theoretical approaches in the literature" (Smets and van Ham, 2013, p. 346). In my study, I adopted this categorization for the replicated models. Table 2.1 gives an overview of the broad theoretical models and indicates how many of the models replicated in the present study belong to the respective broad theoretical models.

2.5.3.2 Operationalization and regression models

The dependent variable in all the regression models calculated is individual-level voter turnout, operationalized as turnout intention. I dichotomized this variable to

[3]For a detailed overview of the replicated models, see table 6.1 in the appendix (A.2).

Table 2.1: Broad theoretical models of individual-level voter turnout (Smets and van Ham 2013)

Broad Theoretical Model	Description	n (2017)	n (2021)
Resource model	Turnout driven by resources (e.g., time, money, skills)	14	13
Rational choice model	Turnout driven by personal cost-benefit calculation	5	5
Mobilization model	Turnout driven by mobilization (e.g., by parties, candidates)	6	5
Psychological model	Turnout driven by cognitive characteristics (e.g., political interest)	3	3
Political institutional model	Decisions to vote driven by the political and institutional context	1	1
Total			
		29	27

distinguish between potential voters and nonvoters. To this end, I divided respondents into two groups: (1) those who reported that they would certainly or likely vote in the 2017/2021 German Federal Election; and (2) those who indicated that they might vote, were not likely to vote, or were sure that they would not vote. However, as I could classify only a small proportion of respondents as nonvoters, the problem of poor variance of the dependent variable occurred. I used Firth's logistic regression to correct this bias in all calculated models. This type of logistic regression uses penalty maximum likelihood estimation to reduce bias in unbalanced samples. It is commonly applied in data analysis with rare event data (King and Zeng, 2001; Heinze and Schemper, 2002; Wang, 2014; Puhr et al., 2017). I used the R package "logistf" to calculate the regression models (Heinze et al., 2022).

2.5.3.3 Investigating differences in the goodness of fit of the models between the two surveys

To investigate whether results of the multivariate analyses differ between the surveys, I first compare whether the calculated models differ in their accuracy between the surveys. With this approach, I investigate whether substantial differences occur in the overall performance of the models, depending on whether they are calculated with data from the probability face-to-face survey or the nonprobability online survey with telephone recruitment for the 2017 data, and between the probability self-administered mixed mode survey and the nonprobability online survey with Web recruitment for the 2021 data. To explore whether the models are suitable for all types of surveys considered in this comparison, I use receiver operating characteristic (ROC) analysis (Fan et al., 2006; Carter et al., 2016), a frequently employed statistical tool for checking whether the predicted values of models coincide with the observed outcome. In binary regression models, it indicates the ability of models to distinguish the two classes of outcomes (Yin, 2017).

To summarize the accuracy of the regression models in a single number, I report the area under the curve (AUC). This commonly used measure takes values between 0.5 to 1, with 0.5 indicating that a model has no discrimination ability and 1 indicating that a model has perfect discrimination ability. To compare whether substantial differences in model accuracy occur between the surveys, I compute significance tests for differences in AUCs. To calculate ROC curves, I use the R package pROC (Robin et al., 2021).

2.5.3.4 Assessing differences in associations of variables between the surveys

I calculate average marginal effects (AMEs) for all regression models to compare results from the surveys in terms of the associations of variables in multivariate models. I investigate whether the AMEs of the calculated models differ substantially between the surveys. By doing so, I aim to explore whether conclusions on determinants of individual-level voter turnout differ when the analysis is conducted with data from a probability face-to-face survey or from a nonprobability online survey with telephone recruitment, or with data from a probability self-administered mixed mode survey or a nonprobability online survey with Web recruitment. To analyze differences in associations of variables between the surveys, I investigate whether

1. the AMEs differ significantly between the two surveys;
2. the AMEs are significant ($p <= 0.05$) in only one of the 2017/2021 surveys;
3. the AMEs differ in direction between the two surveys.

By using three different statistical measures, this analysis allows me to draw a comprehensive conclusion on differences in associations of variables between the surveys.

Figure 2.3 provides an overview of the applied analysis strategy for the AMEs.

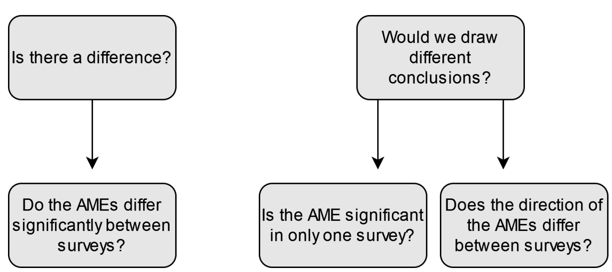

Figure 2.3: Multivariate analysis strategy

Figure 2.4 gives an overview on the different analytical approaches used in this study to compare the surveys.

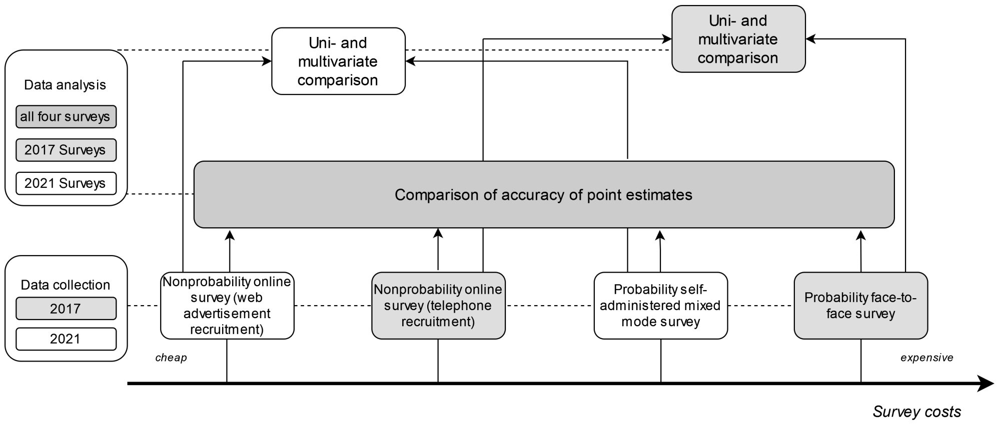

Figure 2.4: Analytical approaches to compare the surveys

2.6 Results

2.6.1 Analyzing the accuracy of point estimates in the surveys

Figure 2.5 shows the absolute bias (value of the variable observed in the target population minus value of the variables estimated by the surveys) with 95 percent confidence intervals for the four surveys, compared with the population parameter from the validated benchmark data (for gender, region, age, party vote and voter turnout). As can be seen from this figure, there are some crucial differences between the surveys and the accuracy of their point estimates for some variables, while similarities for others.

First, looking at gender and region, we can see hardly any differences between the surveys. Looking at age, we can see the bias is largest for the nonprobability online survey recruited via Web interviews (0.079), indicating that adults aged 50 years and older are underrepresented in this surveys. Also the nonprobability online survey recruited via phone interviews (0.042), as well as the probability self-administered mixed mode survey (0.023) fail to give an accurate estimation of adults aged 50 years and older. Only the probability face-to-face survey meets the benchmark value accurate.

Second, we see some remarkable differences in party vote. For the right-wing party (AfD) the bias of the nonprobability online survey recruited via Phone is largest and negative (-0.032), which means that the voters for the right-wing party are overrepresented in this survey. For the conservative party (CDU/CSU) the picture is somewhat different. Here, the bias is largest and positive for the nonprobability online survey recruited via Web advertisement (0.084), and negative (-0.048) for the probability face-to-face survey. With respect to the liberal party (FDP), the differences are minor. Regarding the estimate of green party voters, the bias is largest for the probability self-administered survey (-0.046) and negative, which means that green party voters are overrepresented in this survey. Social democratic party (SPD) voters are slightly overrepresented in the two probability based surveys. Left-wing party (Die LINKE) voters are overrepresented in the nonprobability online surveys recruited via Web advertisment (-0.038) and underrepresented in the three other surveys.

And finally, third, looking at voter turnout, we can see that all four surveys underestimated the proportion of nonvoters. However, the bias of turnout is largest in the nonprobability online survey recruited via phone (−0.21), and smallest for the probability face-to-face survey (−0.136).

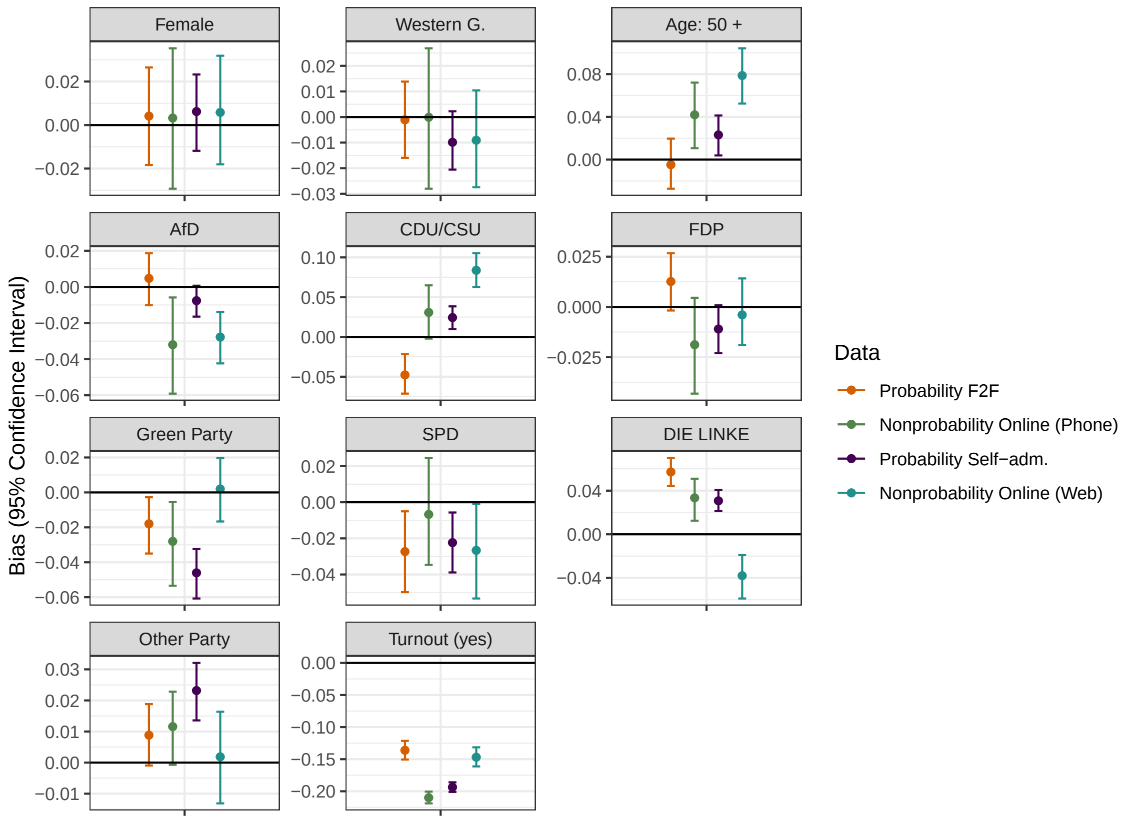

Figure 2.5: Bias in point estimates in the four surveys

Table 2.2 quantifies these differences by reporting the absolute and (absolute) relative bias for all surveys. Overall, we can see that the probability face-to-face survey performs best in terms of the accuracy of point estimates, as the total average bias in the probability face-to-face survey is the smallest among all four surveys in this comparison (0.029). The point estimates of the probability face-to-face survey are also most accurate for the sociodemographic variables, as well as the variables for turnout and party vote.

An interesting note on the accuracy of point estimates is the unexpectedly high bias in the probability self-administered mixed-mode survey. This method exhibits a bias comparable to that of the nonprobability surveys, making it one of the less accurate among the four types assessed in this study, as indicated by the data (0.036). This result contradicts previous findings which generally indicated that probability surveys outperform nonprobability surveys in achieving accurate point estimates against validated benchmarks (Malhotra and Krosnick, 2007; Chang and Krosnick, 2009; Stephenson and Crete, 2011; Yeager et al., 2011). However, my study finds this higher accuracy is specific to the probability face-to-face survey, not extending to the probability self-administered mixed-mode survey, whose bias closely matches that of nonprobability surveys.

Table 2.2: Absolute relative bias across the four surveys

Variable	Response	Nonprobability online survey (web recruitment)	Nonprobability online survey (phone recruitment)	Probability self-adm. mixed mode survey	Probability face-to-face survey
Sociodemographics					
Gender	Female	0.006	0.003	0.006	0.004
Age	50 y +	0.079	0.042	0.023	-0.005
Region	Western G.	-0.009	0.000	-0.010	-0.001
Absolute relative bias		0.031	0.015	0.013	0.003
Vote					
Turnout	Yes	-0.147	-0.210	-0.194	-0.136
Vote Forecast	CDU/CSU	0.084	0.031	0.024	-0.048
	SPD	-0.027	-0.007	-0.022	-0.027
	FDP	-0.004	-0.019	-0.011	0.013
	GRUENE	0.002	-0.028	-0.046	-0.018
	LINKE	-0.028	-0.032	-0.008	0.005
	AfD	-0.038	0.033	0.031	0.057
	Other	0.079	0.042	0.023	0.009
Absolute relative bias		0.041	0.046	0.045	0.039
Total					
Absolute relative bias		0.039	0.038	0.036	0.029

Note: 50 y+ = aged 50 years and over; Western G. = western Germany

2.6.2 Comparing variables and their distributions

Figure 2.6 shows for each of the two surveys with overlapping field periods the differences between means of the variables for which no external benchmark data are available. The differences between the means of the probability surveys and the nonprobability surveys are plotted on the x-axis. Positive values indicate the mean in the probability survey being larger than in the nonprobability survey, negative values indicate the mean in the nonprobability survey being larger than the mean in the probability survey. The variables used in this comparison are plotted on the y-axis. Points closer to zero indicate smaller differences in the means between the surveys; points further away indicate larger differences. The variables labeled are often used to explain individual voting behavior in electoral research (Smets and van Ham, 2013).

We can see from this figure that there is no systematic pattern regarding differences in means of variables between the two surveys for the 2017 data. While the magnitude of differences varies greatly by variable, no specific type of variables (sociodemographic variables, variables that capture political behavior or attitudes) differs particularly strongly in its means between the surveys.

The picture looks different for the 2021 data. For variables that capture political behavior we see here that the mean is higher in the probability self-administered mixed mode survey on almost all variables, which means that respondents from the probability self-administered mixed mode survey seem to be more engaged in politics than respondents from the nonprobability online survey recruited via Web advertisement.

Looking at the variables often used to explain individual voting behavior in electoral research, we can see that respondents from the nonprobability surveys identify less strongly with political parties, and - in the 2017 data - are less interested in politics, which is contrary to findings from previous studies that reported political interest to be higher in nonprobability online surveys (Berrens et al., 2003; Malhotra and Krosnick, 2007; Chang and Krosnick, 2009; Ansolabehere and Schaffner, 2014). Looking at the sociodemographic variables we can see the proportion of married respondents being higher in the probability surveys. Further, looking at the mean income and educational status of the respondents we can see that the mean for these variables was higher in the nonprobability survey 2017, but higher in the probability survey in 2021.

Table 2.3: Univariate analysis: significance and effect sizes of differences in means/proportion between the surveys

	2017 Data					2021 Data				
		Effect size (%)					Effect size(%)			
Variable	significant	small	medium	large	n	significant	small	medium	large	n
Sociodemographic Variable	60.5	65.8	31.6	2.6	38	82.9	40.0	45.7	14.3	35
Political Attitudes	71.1	65.8	28.9	5.3	38	92.5	65.0	35.0	0.0	40
Political Behavior	75.0	66.7	8.3	25.0	12	100.0	16.7	33.3	50.0	6
total	67.0	65.9	27.3	6.8	88	88.9	50.6	39.5	9.9	81

Table 2.3 summarizes the results of the tests conducted to analyze whether the differences between the distributions or proportions in each of the two surveys are significant. The effect sizes of the differences are also reported in this table. For this table, I classified Cohen's d <= 0.2 as small effects, Cohen's d >.02 and <= 0.5 as medium effects, and Cohen's d >= 0.5 as large effects (Cohen, 2013).

In total, 67 percent of the variables in the 2017 data differ significantly in their distribution/proportion between the two surveys, in the 2021 data mostly all (88.9

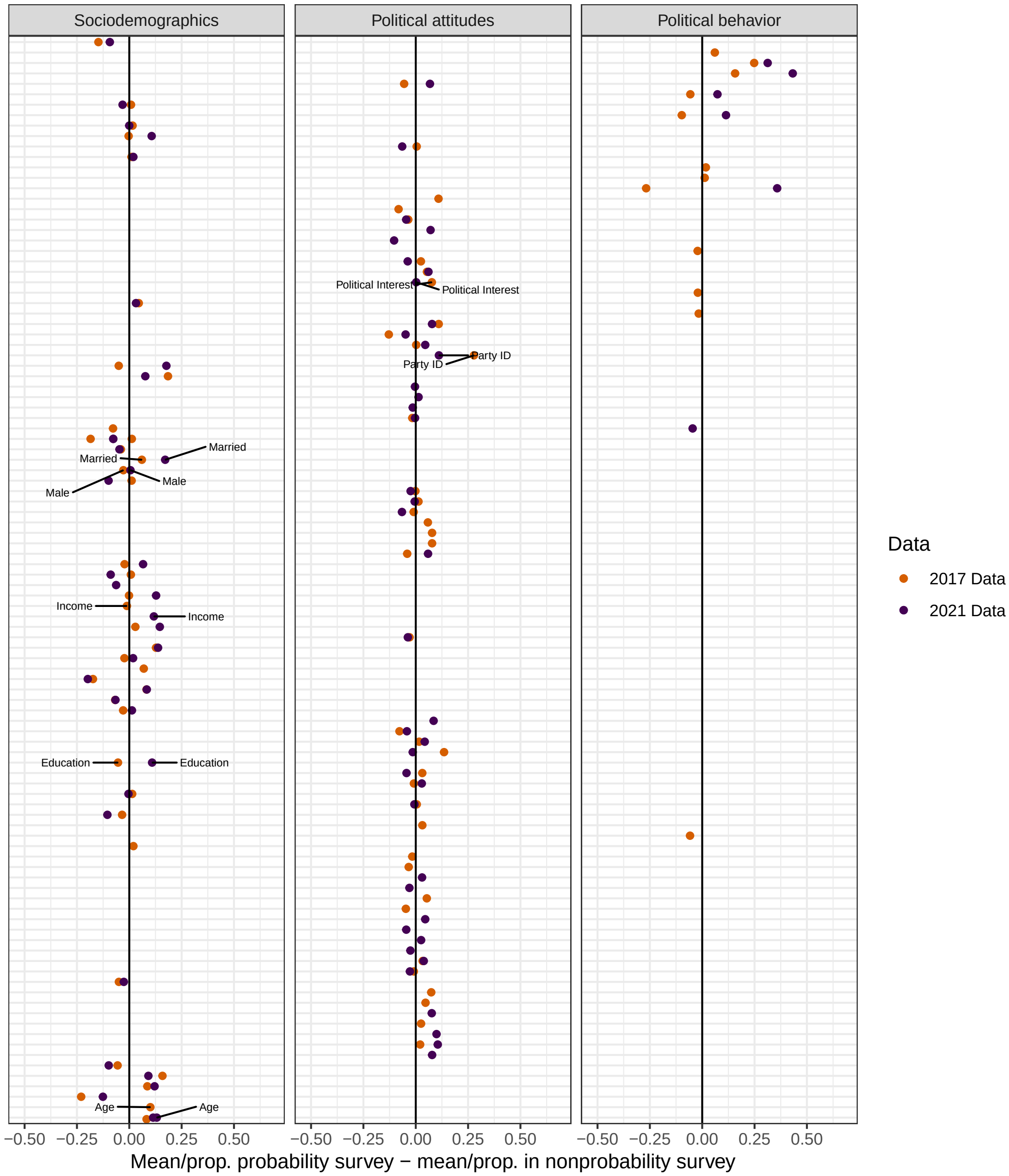

Figure 2.6: Differences in means/proportion between the probability survey and the nonprobability survey

percent) of the variables differ significantly. Looking at each subgroup of variables (sociodemographic variables, political attitudes and behavior), the percent of differences between surveys that are significant are also larger in the 2021 comparison. While for the sociodemographic variables (n = 38), 60.5 percent of the differences between the surveys are significant in the 2017 data, 82.9 percent of the differences between the surveys are significant in the 2021 data (n = 35). The significant differences for variables covering political attitudes and behavior are considerably larger (about 70 percent in the 2017 data and about 90 percent in the 2021 data).

Looking at the effect sizes of the differences, we see that most of the differences in the distributions/proportions between the surveys are small or of a medium effect size (93 percent in the 2017 data, 90 in the 2021 data). These small/medium effects indicate an overlap of values for at least 80.3 percent (see Magnussion (2022)). With this high overlap of values, I assume minor/medium differences are relatively unproblematic for using these variables in statistical analysis. However, in two-thirds of the variables under investigation covering political attitudes in the 2021 data, the effect sizes are large. For these variables, the chance that an individual picked at random from the probability survey will have a higher score on the variable under investigation than an individual picked at random from the nonprobability survey is 71.4 percent (see Magnussion (2022)). With this, I consider large differences are problematic for using these variables in statistical analysis since the difference between the probability and the nonprobability survey is sufficiently large.

2.6.3 Comparing models' goodness of fit and associations in multivariate models

2.6.3.1 Investigating differences in models' goodness of fit between the surveys

Figure 2.7 shows for each of the two surveys with overlapping field periods the differences between the AUCs for all calculated models. The differences between the AUCs of the probability and the nonprobability surveys are plotted on the x-axis. The models used in this comparison are plotted on the y-axis. Positive values indicate the AUC in the probability survey being larger than in the nonprobability survey, negative values indicate the AUC in the nonprobability survey being larger than the AUC in the probability survey. Points closer to zero indicate smaller differences in the AUCs between the surveys; points further away indicate larger differences. I also indicate the broad theoretical model of individual-level voter turnout to which each regression model belongs (Smets and van Ham, 2013).

This figure shows that most of the differences in accuracy between regression models are minor fo each, the 2017 and the 2021 data. However, in the regression model that belong to the broad theoretical model "psychological model", we see in two of the AUCs larger differences. In both cases, the AUC in the nonprobability online survey with Web recruitment is larger than in the probability self-administered mixed mode survey, which means that the discrimination ability is higher in the nonprobability online survey.

In Table 2.4, we see the results from the significance test of the difference between the AUCs of the calculated regression models in the probability survey and the

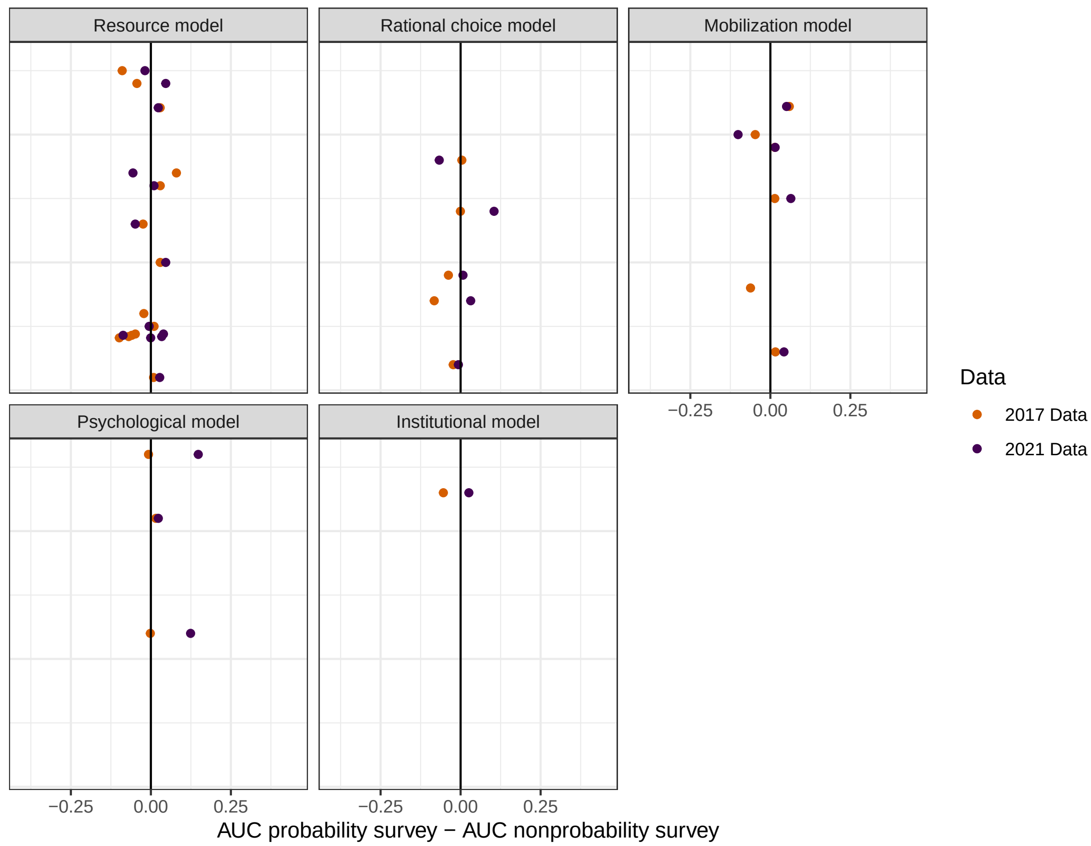

Figure 2.7: Differences in areas under the curve (AUCs) for all calculated regression models

nonprobability survey for the 2017 and 2021 data.

The amount of significant differences is smaller in the 2021 data, with 11 percent of significant differences in the AUCs compared to 21 percent of significant differences in the AUCs of the 2017 data. Since, to the best of my knowledge, this is the first study to compare probability surveys with nonprobability surveys in terms of the accuracy of regression models, I had no expectations regarding the differences between the surveys in this regard. However, this analysis shows that the surveys are comparable regarding the discrimination ability - and thus the accuracy of the regression models.

Table 2.4: Comparison of AUCs for all calculated regression models categorized into broad theoretical models of individual-level voter turnout

Broad theoretical model	2017 Data		2021 Data	
	significant difference (in %)	n	significant difference (in %)	n
Resource	21.4	14	0.0	13
Rational choice	20.0	5	20.0	5
Mobilization	16.7	6	0.0	5
Psychological	0.0	3	66.7	3
Institutional	100.0	1	0.0	1
total	20.7	29	11.1	27

2.6.3.2 Assessing differences in associations of variables between the surveys

The picture is somewhat different when investigating differences in the associations of variables in multivariate models between the surveys. Figure 2.8 shows for each of the two surveys with overlapping field periods the differences between the AMEs for all calculated models. The differences between the AMEs of the probability and the nonprobability surveys are plotted on the x-axis. The models used in this comparison are plotted on the y-axis. Positive values indicate the AME in the probability survey being larger than in the nonprobability survey, negative values indicate the AME in the nonprobability survey being larger than the AME in the probability survey. Points closer to zero indicate smaller differences in the AMEs between the surveys; points further away indicate larger differences. I also indicate the broad theoretical model of individual-level voter turnout to which each regression model belongs (Smets and van Ham, 2013).

This figure shows that most of the AMEs are plotted close to each other, indicating that they do not differ greatly between the two surveys. However, for some of the AMEs, the differences between the nonprobability survey and the probability survey appear to be more extensive, especially for AMEs in models that belong to the broad theoretical model categories “Resource”, “Rational choice” and “Psychological”.

However, what is hard to derive from this figure is whether conclusions about the determinants of voting behavior would differ if the data were collected with a probability face-to-face, a nonprobability online survey recruited via phone or a probability self-administered mixed mode survey or a nonprobability online survey recruited via phone interviews. To answer this question, I conducted several follow-up tests on the AMEs (Table 2.5). I first ran a significance in difference test to quantify whether the AME in the probability survey differs significantly from the AME in the nonprobability survey. To further explore whether conclusions on determinants of individual-level voter turnout differ when the analysis is conducted with data from the different surveys, I investigate whether they become significant in both surveys. Third, I look deeper at the direction of the significant AMEs and investigate whether the effects are positive in one survey and negative in the other.

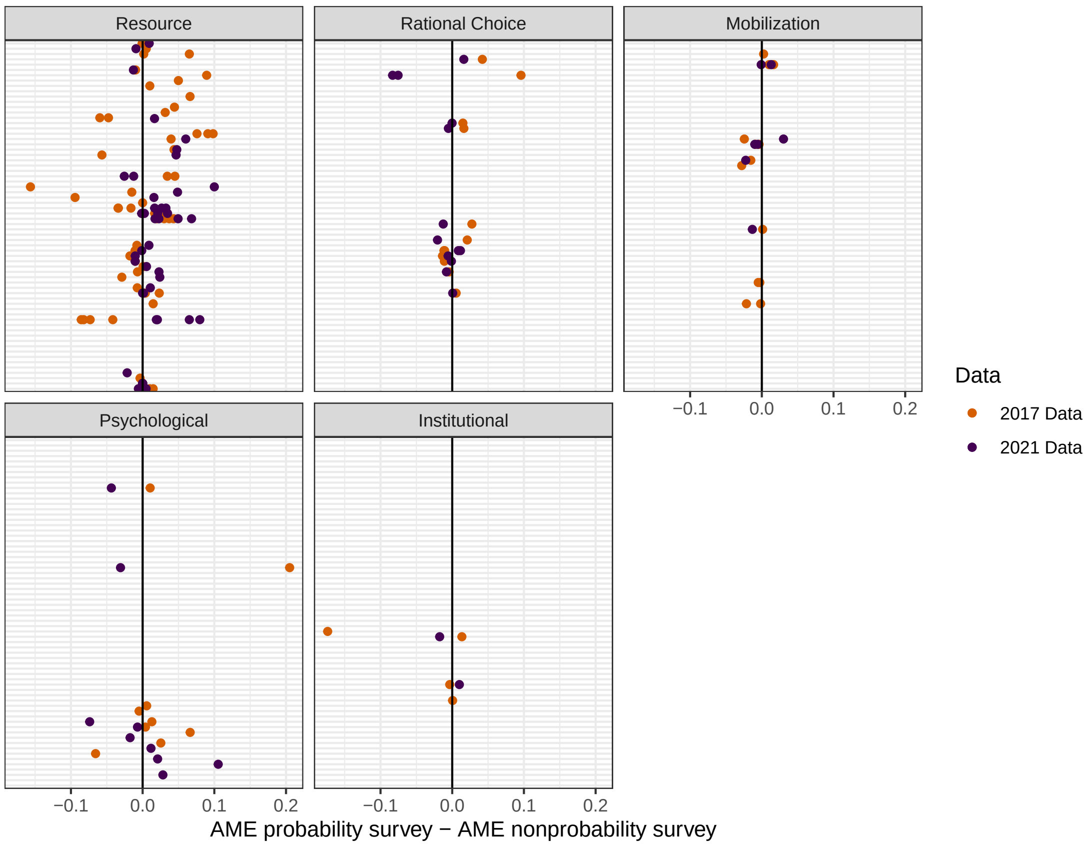

Figure 2.8: Differences in average marginal effects for all calculated regression models

Table 2.5: Differences in the significance and direction of AMEs between the surveys

	Is there a difference?		Would we draw different conclusions?					
Broad theoretical model	significant difference (in %)		significant in only one survey (in %)		Difference in direction		n	
	2017 Data	2021 Data	2017 Data	2021 Data	2017 Data	2021 Data	2017 Data	2021 Data
Resource	53.0	62.0	50.0	26.0	28.8	52.0	66	50
Rational choice	61.5	15.4	53.8	23.1	15.4	23.1	13	13
Mobilization	42.9	28.6	21.4	14.3	35.7	57.1	14	7
Psychological	66.7	55.6	22.2	33.3	55.6	55.6	9	9
Institutional	75.0	0.0	25.0	50.0	0.0	0.0	4	2
total	54.7	49.4	43.4	25.9	29.2	46.9	106	81

Investigating differences between the AMEs we can see that 54.7 percent of the AMEs in the 2017 data differ significantly between the two surveys. The proportion of significant differences of AMEs is considerably larger in the 2021 data, with 49.4 percent of the investigated AMEs differ significantly between the two surveys. Here, in the regression models that belong to the broad theoretical models "resource" and "psychological", more than 50 percent of the AMEs differ significantly between the probability self-administered mixed mode survey and the nonprobability online survey, recruited via Web advertisement.

Looking at the significance of the AMEs across the surveys, we see that the significance differs for 43.4 percent of the AMEs in the 2017 data and 25.9 percent of the 2021 data. Here, the effect is significant either in the probability survey or in the nonprobability survey. Moreover, the direction of 29.2 percent of the AMEs differs between the two surveys in the 2017 data and 46.9 percent in the 2021 data, indicating that we would draw different conclusions on how the independent variables influence individual-level voter turnout depending on which survey we used.

The latter finding is somewhat surprising, as previous studies mostly show no differences in multivariate associations between variables depending on whether they are collected with a probability (face-to-face) survey or a nonprobability online survey (Berrens et al., 2003; Chang and Krosnick, 2009; Stephenson and Crete, 2011; Yeager et al., 2011; Ansolabehere and Schaffner, 2014; Breton et al., 2017; Bytzek and Bieber, 2016; Dassonneville et al., 2018).

2.7 Discussion

This study set out to extend previous research on differences in empirical analysis between surveys that differ in mode and sampling in electoral research. From a sampling theory perspective, I argued that probability surveys can be expected to perform better, as sampling theory is not applicable to nonprobability surveys. Further, I argued that differences could be expected between interviewer-administered and self-administered surveys. However, the overview of empirical studies provided in Section 3 of this study shows that findings are mixed as to whether results of empirical analyses are affected by the type of survey with which the data are collected. Based on these results, I extended the analysis strategy in the current comparison, hoping to gain new insights into outcome differences between surveys that differ in mode and sampling.

By comparing a large set of variables commonly used in electoral research, by using three different analytical approaches and multiple statistical measures, and by considering different broad theoretical models of individual-level voter turnout (Smets and van Ham, 2013), I could quantify the comparability of the probability face-to-face survey, the probability self-administered mixed mode survey and the nonprobability online surveys and include different theoretical explanations of voting behavior in this comparison.

The results of my analysis consistently show that the probability face-to-face survey performs best in terms of to the accuracy of point estimates. Therefore, I would encourage electoral researchers to rely on probability face-to-face surveys for estimating valid population values.

Regarding the use of different kinds of surveys in multivariate analysis of individual-level voter turnout, the results suggest that for many regression models across different broad theoretical models and for many associations between variables in these regression models we would obtain different results if we collect data with surveys that differ in mode and sampling. However, in this step of data analysis I was only able to compare two surveys each - the probability face-to-face survey with the nonprobability online survey recruited via phone interviews and the probability self-administered mixed mode survey with the nonprobability online survey recruited via Web advertisement. As we do not have "true" values for these estimates, I cannot conclude that one of the surveys performs worse or better than the other one. However, we can see from these results that sampling and mode do make a difference, since the findings are consistent across both comparisons.

Although probability self-administered mixed mode and nonprobability online surveys are cost-effective and allow for bigger sample sizes than probability-based face-to-face surveys, we should bear in mind that results from the surveys are comparable only to a limited extent. The key finding of the present study is that we may achieve different results in the analysis of individual-level voter turnout and draw different conclusions on the drivers of individual-level voter turnout and voting behavior if we replace probability face-to-face surveys with surveys that differ in mode and/or sampling.

The limitations of this study should be acknowledged. I compared a probability face-to-face survey and a self-administered mixed mode survey with two nonprobability online surveys. With this approach, I could show what differences in estimates could be expected if we replaced probability face-to-face surveys with surveys that differ in mode and sampling. However, I cannot say with certainty that the differences observed between the surveys are due to the sources of error considered, or whether unobserved/non-considered factors in which the surveys differed are responsible for the observed differences. Previous research and theoretical arguments indicate that it is very likely that the differences are due to the the sampling approach and data collection mode. Despite these limitations, this study contributes to investigating differences in estimates based on different surveys in the field of electoral research.

Lastly, I argued that survey costs are the main driver in changing data collection from probability to nonprobability and from interviewer-administered to self-administered surveys. To further assess how to reduce costs in data collection while achieving the best possible results within given budgets in statistical analyses, future studies should investigate the accuracy of surveys per unit of currency spent. For example, one could estimate a valid population parameter with external benchmark data available across different survey samples and modes considering the costs for each interview for the different surveys. As very few studies to date have compared the survey costs of different survey modes (Olson et al., 2021b), and a systematic comparison is lacking, such a study could help researchers to choose the best possible data collection method within their given budget and to minimize biases in estimates within cost constraints.

References

Adams, James, Jay Dow, and Samuel Merrill. 2006. "The Political Consequences of Alienation-Based and Indifference-Based Voter Abstention: Applications to Presidential Elections." *Political Behavior* 28 (1): 65–86. https://doi.org/10.1007/s11109-005-9002-1.

AnduizaPerea, Eva. 2005. "Campaign Effects in the Spanish Election of 2000." *Journal of Elections, Public Opinion and Parties* 15 (2): 215–36. https://doi.org/10.1080/13689880500178815.

ANES. 2017. "ANES 2016 Time Series Study: Version 2." ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR36824.V2.

———. 2021. "ANES 2020 Time Series Study Full Release."

Ansolabehere, Stephen, and Douglas Rivers. 2013. "Cooperative Survey Research." *Annual Review of Political Science* 16 (1): 307–29. https://doi.org/10.1146/annurev-polisci-022811-160625.

Ansolabehere, Stephen, and Brian F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* 22 (3): 285–303. https://doi.org/10.1093/pan/mpt025.

Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. Dillman, et al. 2010. "Research Synthesis: AAPOR Report on Online Panels." *Public Opinion Quarterly* 74 (4): 711–81. https://doi.org/10.1093/poq/nfq048.

Bandilla, Wolfgang, Lars Kaczmirek, Michael Blohm, and Wolfgang Neubarth. 2009. "Coverage- Und Nonresponse-Effekte Bei Online-Bevölkerungsumfragen." In *Sozialforschung Im Internet*, edited by Nikolaus Jackob, Harald Schoen, and Thomas Zerback, 129–43. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91791-7_8.

Berrens, Robert P., Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer. 2003. "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples." *Political Analysis* 11 (1): 1–22. https://doi.org/10.1093/pan/11.1.1.

Bethlehem, Jelke. 2009. *Applied Survey Methods.* Hoboken, NJ, USA: John Wiley

& Sons, Inc. https://doi.org/10.1002/9780470494998.

———. 2016. "Solving the Nonresponse Problem With Sample Matching?" *Social Science Computer Review* 34 (1): 59–77. https://doi.org/10.1177/0894439315573926.

Biemer, P. P. 2010. "Total Survey Error: Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74 (5): 817–48. https://doi.org/10.1093/poq/nfq058.

Biffignandi, Silvia, and Jelke G. Bethlehem. 2021. *Handbook of Web Surveys*. Second edition. Hoboken, NJ: Wiley.

Blais, Andre, Elisabeth Gidengil, and Neil Nevitte. 2004. "Where Does Turnout Decline Come From?" *European Journal of Political Research* 43 (2): 221–36. https://doi.org/10.1111/j.1475-6765.2004.00152.x.

Breton, Charles, Fred Cutler, Sarah Lachance, and Alex Mierke-Zatwarnicki. 2017. "Telephone Versus Online Survey Modes for Election Studies: Comparing Canadian Public Opinion and Vote Choice in the 2015 Federal Election." *Canadian Journal of Political Science* 50 (4): 1005–36. https://doi.org/10.1017/S0008423917000610.

Brick, J. M. 2011. "The Future of Survey Sampling." *Public Opinion Quarterly* 75 (5): 872–88. https://doi.org/10.1093/poq/nfr045.

Bundeswahlleiter, Der. 2021. "Repräsentative Wahlstatistik Zur Bundestagswahl 2021." Statistisches Bundesamt, Wiesbaden.

Bytzek, Evelyn, and Ina E. Bieber. 2016. "Does Survey Mode Matter for Studying Electoral Behaviour? Evidence from the 2009 German Longitudinal Election Study." *Electoral Studies* 43: 41–51. https://doi.org/10.1016/j.electstud.2016.04.007.

Callegaro, Mario, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas, eds. 2014. *Online Panel Research*. Chichester, UK: John

Wiley & Sons, Ltd. https://doi.org/10.1002/9781118763520.

Carter, Jane V., Jianmin Pan, Shesh N. Rai, and Susan Galandiuk. 2016. "ROC-ing Along: Evaluation and Interpretation of Receiver Operating Characteristic Curves." *Surgery* 159 (6): 1638–45. https://doi.org/10.1016/j.surg.2015.12.029.

Cernat, Alexandru, and Joseph Sakshaug. 2020. "The Impact of Mixed Modes on Multiple Types of Measurement Error." *Survey Research Methods*, April, 79–91 Pages. https://doi.org/10.18148/SRM/2020.V14I1.7450.

Chang, Linchiat, and Jon A. Krosnick. 2009. "National Surveys Via Rdd Telephone Interviewing Versus the Internet." *Public Opinion Quarterly* 73 (4): 641–78. https://doi.org/10.1093/poq/nfp075.

Chong, Dennis, and Reuel Rogers. 2005. "Racial Solidarity and Political Participation." *Political Behavior* 27 (4): 347–74. https://doi.org/10.1007/s11109-005-5880-5.

Clarke, Harold D., David Sanders, Marianne C. Stewart, and Paul F. Whiteley. 2002. "Downs, Stokes and Modified Rational Choice: Modelling Turnout in 2001." *British Elections & Parties Review* 12 (1): 28–47. https://doi.org/10.1080/13689880208413068.

Cohen, Jacob. 2013. *Statistical Power Analysis for the Behavioral Sciences*. Zeroth. Routledge. https://doi.org/10.4324/9780203771587.

Cornesse, Carina, Annelies G Blom, David Dutwin, Jon A Krosnick, Edith D De Leeuw, Stéphane Legleye, Josh Pasek, et al. 2020. "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research." *Journal of Survey Statistics and Methodology* 8 (1): 4–36. https://doi.org/10.1093/jssam/smz041.

Couper, Mick P., Christopher Antoun, and Aigul Mavletova. 2017. "Mobile Web

Surveys." In *Total Survey Error in Practice*, edited by Paul P. Biemer, Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West, 28:133–54. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119041702.ch7.

Dassonneville, Ruth, André Blais, Marc Hooghe, and Kris Deschouwer. 2018. "The Effects of Survey Mode and Sampling in Belgian Election Studies: A Comparison of a National Probability Face-to-Face Survey and a Nonprobability Internet Survey." *Acta Politica* 11 (1): 23. https://doi.org/10.1057/s41269-018-0110-4.

de Leeuw, Edith D. 2010. "Mixed-Mode Surveys and the Internet." *Survey Practice* 3 (6): 1–5. https://doi.org/10.29115/SP-2010-0030.

Dillman, Don A., and Leah Melani Christian. 2005. "Survey Mode as a Source of Instability in Responses Across Surveys." *Field Methods* 17 (1): 30–52. https://doi.org/10.1177/1525822X04269550.

Dumelle, Michael, Matt Higham, Jay M. Ver Hoef, Anthony R. Olsen, and Lisa Madsen. 2022. "A Comparison of Design-Based and Model-Based Approaches for Finite Population Spatial Sampling and Inference." *Methods in Ecology and Evolution*, June, 2041–210X.13919. https://doi.org/10.1111/2041-210X.13919.

Fan, Jerome, Suneel Upadhye, and Andrew Worster. 2006. "Understanding Receiver Operating Characteristic (ROC) Curves." *CJEM* 8 (01): 19–20. https://doi.org/10.1017/S1481803500013336.

Fieldhouse, E., J. Green, G. Evans, C. Prosser, R. De Geus, J. Bailey, H. Schmitt, C. Van Der Eijk, and J. Mellon. 2022. "BES British Election Studies, 1969-British Election Study, 2019: Post-Election Random Probability Survey." UK Data Service. https://doi.org/10.5255/UKDA-SN-8875-1.

Ghitza, Yair, and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups." *American Jour-*

nal of Political Science 57 (3): 762–76. https://doi.org/10.1111/ajps.12004.

Gideon, Lior, ed. 2012. *Handbook of Survey Methodology for the Social Sciences*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3876-2.

GLES. 2019a. "Langfrist-Online-Tracking T37 (GLES)." https://doi.org/10.4232/1.13295.

———. 2019b. "Vorwahl-Querschnitt (GLES 2017)." https://doi.org/10.4232/1.13234.

GLES. 2022a. "GLES Cross-Section 2021, Pre-Election GLES Querschnitt 2021, Vorwahl." GESIS Data Archive. https://doi.org/10.4232/1.13860.

———. 2022b. "GLES Tracking September 2021, T50 GLES Tracking September 2021, T50." GESIS. https://doi.org/10.4232/1.14000.

Goldstein, Ken, and Paul Freedman. 2002. "Campaign Advertising and Voter Turnout: New Evidence for a Stimulation Effect." *The Journal of Politics* 64 (3): 721–40. https://doi.org/10.1111/0022-3816.00143.

Green, Donald P., and Ron Shachar. 2000. "Habit Formation and Political Behaviour: Evidence of Consuetude in Voter Turnou." *British Journal of Political Science* 30: 561–73.

Groves, Robert M. 2004. *Survey Errors and Survey Costs*. Repr. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley-Interscience.

———. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70 (5): 646–75. https://doi.org/10.1093/poq/nfl033.

Groves, Robert M., Floyd J. Fowler, Mick Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau, eds. 2009. *Survey Methodology*. 2. ed. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley.

Groves, Robert M., and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72 (2): 167–89.

https://doi.org/10.1093/poq/nfn011.

Heath, Anthony. 2000. "Were Traditional Labour Voters Disillusioned with New Labour? Abstention at the 1997 General Election." *British Elections & Parties Review* 10 (1): 32–46. https://doi.org/10.1080/13689880008413035.

Heerwegh, D., and G. Loosveldt. 2008. "Face-to-Face Versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality." *Public Opinion Quarterly* 72 (5): 836–46. https://doi.org/10.1093/poq/nfn045.

Heinze, Georg, Meinhard Ploner, Daniela Dunkler, Harry Southworth, and Lena Jiricka. 2022. "Package 'Logistf': Firth's Bias-Reduced Logistic Regression." CRAN.

Heinze, Georg, and Michael Schemper. 2002. "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21 (16): 2409–19. https://doi.org/10.1002/sim.1047.

Highton, Benjamin, and Raymond E. Wolfinger. 2001. "The First Seven Years of the Political Life Cycle." *American Journal of Political Science* 45 (1): 202. https://doi.org/10.2307/2669367.

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires." *Public Opinion Quarterly* 67 (1): 79–125. https://doi.org/10.1086/346010.

Holbrook, Allyson L., Jon A. Krosnick, Penny S. Visser, Wendi L. Gardner, and John T. Cacioppo. 2001. "Attitudes Toward Presidential Candidates and Political Parties: Initial Optimism, Inertial First Impressions, and a Focus on Flaws." *American Journal of Political Science* 45 (4): 930. https://doi.org/10.2307/2669333.

Jackson, Robert A. 2003. "Differential Influences on Latino Electoral Participation." *Political Behavior* 25 (4): 339–66. https://doi.org/10.1023/B:POBE.0000004062.

12215.63.

Kennedy, Courtney, Andrew W. Mercer, Keeter Scott, Nick Hatley, Kyley McGeeney, and Alejandra Gimenez. 2016. "Evaluating Online Nonprobability Surveys Vendor Choice Matters; Widespread Errors Found for Estimates Based on Blacks and Hispanics." PEW Research Center.

Kennedy, Lauren, and Andrew Gelman. 2021. "Know Your Population and Know Your Model: Using Model-Based Regression and Poststratification to Generalize Findings Beyond the Observed Sample." *Psychological Methods* 26 (5): 547–58. https://doi.org/10.1037/met0000362.

Killian, Mitchell, Ryan Schoen, and Aaron Dusso. 2008. "Keeping Up with the Joneses: The Interplay of Personal and Collective Evaluations in Voter Turnout." *Political Behavior* 30 (3): 323–40. https://doi.org/10.1007/s11109-007-9051-8.

King, Gary, and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9 (2): 137–63. https://doi.org/10.1093/oxfordjournals.pan.a004868.

Kish, Leslie. 1995. *Survey Sampling.* A Wiley Interscience Publication. New York: Wiley.

Kobold, Kevin, and Sven Schmiedel. 2018. "Wahlverhalten Bei Der Bundestagswahl 2017 Nach Geschlecht Und Alter: Ergebnisse Der Repräsentativen Wahlstatistik." *WISTA* 3: 142–56.

Leighley, Jan E., and Jonathan Nagler. 2007. "Unions, Voter Turnout, and Class Bias in the U.S. Electorate, 1964–2004." *The Journal of Politics* 69 (2): 430–41. https://doi.org/10.1111/j.1468-2508.2007.00541.x.

Loosveldt, Geert, and Nathalie Sonck. 2008. "An Evaluation of the Weighting Procedures for an Online Access Panel Survey." *Survey Research Methods* 2 (2): 93–105. https://doi.org/10.18148/srm/2008.v2i2.82.

Lyons, William, and Robert Alexander. 2000. "A Tale of Two Electorates: Generational Replacement and the Decline of Voting in Presidential Elections." *The Journal of Politics* 62 (4): 1014–34. https://doi.org/10.1111/0022-3816.00044.

Magnussion, Kristoffer. 2022. "Interpreting Cohen's d Effect Size: An Interactive Visualization (Version 2.5.2) [Web App]. R Psychologist."

Malhotra, Neil, and Jon A. Krosnick. 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." *Political Analysis* 15 (3): 286–323. https://doi.org/10.1093/pan/mpm003.

Millar, M. M., and D. A. Dillman. 2011. "Improving Response to Web and Mixed-Mode Surveys." *Public Opinion Quarterly* 75 (2): 249–69. https://doi.org/10.1093/poq/nfr003.

Mohorko, Anja, Edith de Leeuw, and Joop Hox. 2013. "Internet Coverage and Coverage Bias in Europe: Developments Across Countries and Over Time." *Journal of Official Statistics* 29 (4): 609–22. https://doi.org/10.2478/jos-2013-0042.

Mughan, Anthony, and Dean Lacy. 2002. "Economic Performance, Job Insecurity and Electoral Choice." *British Journal of Political Science* 32 (03): 513–33. https://doi.org/10.1017/S0007123402000212.

Mutz, Diana C. 2002. "The Consequences of Cross-Cutting Networks for Political Participation." *American Journal of Political Science* 46 (4): 838. https://doi.org/10.2307/3088437.

Neuman, W. Lawrence. 2012. "Designing the Face-to-Face Survey." In *Handbook of Survey Methodology for the Social Sciences*, edited by Lior Gideon, 227–48. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3876-2_14.

Olson, Kristen, Jolene D Smyth, Rachel Horwitz, Scott Keeter, Virginia Lesser, Stephanie Marken, Nancy A Mathiowetz, et al. 2021. "Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys: AAPOR Task Force Report." *Journal of Survey Statistics and Methodology* 9 (3): 381–411. https://doi.org/10.1093/jssam/smz062.

Olson, Kristen, James Wagner, and Raeda Anderson. 2021. "Survey Costs: Where Are We and What Is the Way Forward?" *Journal of Survey Statistics and Methodology* 9 (5): 921–42. https://doi.org/10.1093/jssam/smaa014.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12 (4): 375–85. https://doi.org/10.1093/pan/mph024.

Pasek, Josh. 2016. "When Will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence." *International Journal of Public Opinion Research* 28 (2): 269–91. https://doi.org/10.1093/ijpor/edv016.

Pattie, C. J., and R. J. Johnston. 2009. "Conversation, Disagreement and Political Participation." *Political Behavior* 31 (2): 261–85. https://doi.org/10.1007/s11109-008-9071-z.

Perea, Eva Anduiza. 2002. "Individual Characteristics, Institutional Incentives and Electoral Abstention in Western Europe." *European Journal of Political Research* 41 (5): 643–73. https://doi.org/10.1111/1475-6765.00025.

Puhr, Rainer, Georg Heinze, Mariana Nold, Lara Lusa, and Angelika Geroldinger. 2017. "Firth's Logistic Regression with Rare Events: Accurate Effect Estimates and Predictions?" *Statistics in Medicine*. https://doi.org/10.1002/sim.7273.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, Markus Müller, Stefan Siegert, Matthias Doering, and Zane Billings. 2021. "Package 'pROC': Display and Analyze ROC Curves."

CRAN.

Rubenson, Daniel, André Blais, Patrick Fournier, Elisabeth Gidengil, and Neil Nevitte. 2004. "Accounting for the Age Gap in Turnout." *Acta Politica* 39 (4): 407–21. https://doi.org/10.1057/palgrave.ap.5500079.

Sanders, David, Harold D. Clarke, Marianne C. Stewart, and Paul Whiteley. 2007. "Does Mode Matter For Modeling Political Choice? Evidence From the 2005 British Election Study." *Political Analysis* 15 (3): 257–85. https://doi.org/10.1093/pan/mpl010.

Schonlau, Matthias, Arthur van Soest, Arie Kapteyn, and Mick Couper. 2009. "Selection Bias in Web Surveys and the Use of Propensity Scores." *Sociological Methods & Research* 37 (3): 291–318. https://doi.org/10.1177/0049124108327128.

Smets, Kaat, and Carolien van Ham. 2013. "The Embarrassment of Riches? A Meta-Analysis of Individual-Level Research on Voter Turnout." *Electoral Studies* 32 (2): 344–59. https://doi.org/10.1016/j.electstud.2012.12.006.

Stephenson, L. B., and J. Crete. 2011. "Studying Political Behavior: A Comparison of Internet and Telephone Surveys." *International Journal of Public Opinion Research* 23 (1): 24–55. https://doi.org/10.1093/ijpor/edq025.

Sterrett, David, Dan Malato, Jennifer Benz, Trevor Tompson, and Ned English. 2017. "Assessing Changes in Coverage Bias of Web Surveys in the United States." *Public Opinion Quarterly* 81 (S1): 338–56. https://doi.org/10.1093/poq/nfx002.

Stevens, Daniel, John Sullivan, Barbara Allen, and Dean Alger. 2008. "What's Good for the Goose Is Bad for the Gander: Negative Political Advertising, Partisanship, and Turnout." *The Journal of Politics* 70 (2): 527–41. https://doi.org/10.1017/S0022381608080481.

Struminskaya, Bella, Edith De Leeuw, and Lars Kaczmirek. 2016. "Mode System Effects in an Online Panel Study: Comparing a Probability-based Online Panel with Two Face-to-Face Reference Surveys." *Methods* data (July): 54 Pages. https://doi.org/10.12758/MDA.2015.001.

Trangucci, Rob, Imad Ali, Andrew Gelman, and Doug Rivers. 2018. "Voting Patterns in 2016: Exploration Using Multilevel Regression and Poststratification (MRP) on Pre-Election Polls." arXiv. https://doi.org/10.48550/ARXIV.1802.00842.

Valliant, Richard, and Jill A. Dever. 2011. "Estimating Propensity Adjustments for Volunteer Web Surveys." *Sociological Methods & Research* 40 (1): 105–37. https://doi.org/10.1177/0049124110392533.

Wagner, J. 2012. "A Comparison of Alternative Indicators for the Risk of Nonresponse Bias." *Public Opinion Quarterly* 76 (3): 555–75. https://doi.org/10.1093/poq/nfs032.

Wagner, Markus, Julian Aichholzer, Jakob-Moritz Eberl, Thomas M. Meyer, Nicolai Berk, Nico Büttner, Hajo Boomgaarden, Sylvia Kritzinger, and Wolfgang C. Müller. 2018. "AUTNES Online Panel Study 2017 (SUF Edition)." AUSSDA. https://doi.org/10.11587/I7QIYJ.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31 (3): 980–91. https://doi.org/10.1016/j.ijforecast.2014.06.001.

Wang, Xuefeng. 2014. "Firth Logistic Regression for Rare Variant Association Tests." *Frontiers in Genetics* 5 (June). https://doi.org/10.3389/fgene.2014.00187.

Wass, Hanna. 2007. "The Effects of Age, Generation and Period on Turnout in Finland 1975–2003." *Electoral Studies* 26 (3): 648–59. https://doi.org/10.1016/j.electstud.2007.06.002.

Wolf, Christof, Pablo Christmann, Tobias Gummer, Christian Schnaudt, and Sascha Verhoeven. 2021. "Conducting General Social Surveys as Self-Administered

Mixed-Mode Surveys." *Public Opinion Quarterly* 85 (2): 623–48. https://doi.org/10.1093/poq/nfab039.

Yeager, David S., Jon A. Krosnick, Linchiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." *Public Opinion Quarterly* 75 (4): 709–47. https://doi.org/10.1093/poq/nfr020.

Yin, Jingjing. 2017. "Using the ROC Curve to Measure Association and Evaluate Prediction Accuracy for a Binary Outcome." *Biometrics & Biostatistics International Journal* 5 (3). https://doi.org/10.15406/bbij.2017.05.00134.

Yoo, Sung-jin. 2010. "Two Types of Neutrality: Ambivalence Versus Indifference and Political Participation." *The Journal of Politics* 72 (1): 163–77. https://doi.org/10.1017/s0022381609990545.

3 Exploring the Feasibility of Recruiting Respondents and Collecting Web Data via Smartphone: A Case Study of Text-to-Web Recruitment for a General Population Survey in Germany[1]

Abstract

The widespread usage of smartphones, as well as their technical features, offers many opportunities for survey research. As a result, the importance and popularity of smartphone surveys is steadily increasing. To explore the feasibility of a new text-to-Web approach for surveying people directly via their smartphones, we conducted a case study in Germany in which we recruited respondents from a mobile random digit dialing sample via text messages that included a link to a Web survey. We show

[1]The study was conducted with Matthias Sand and has already been published in JSSAM; see (Bucher and Sand, 2022).

that, although this survey approach is feasible, it is hampered by a number of issues, namely a high loss of numbers at the invitation stage, and a high rate of implicit refusals on the landing page of the survey.

3.1 Introduction

For many years, text messaging on mobile phones has been a common form of communication (Battestini et al., 2010; Church and de Oliveira, 2013; Hall et al., 2015). Moreover, especially in developed economies such as the United States and Western Europe, most of the population is now reachable via smartphone (Silver, 2019). In Germany, for example, the smartphone penetration rate at household level rose to 81.6 percent in 2019 (destatis, 2019), and in the United States, 81 percent of adults surveyed by the Pew Research Center in early 2019 reported that they owned a smartphone (Pew, 2019). The spread of smartphones is changing the way we connect to the Internet: people increasingly use smart- phones for online access, not only while "on the go" but also at home (Pew, 2019).

The widespread use of smartphones, and the fact that they are both mobile phones and Internet-enabled devices, offers new possibilities for survey research because multiple ways of contacting people via smartphone can be combined. For example, text messages can be used for recruitment (Bosnjak et al., 2008; Dal Grande et al., 2016), and respondents can then be surveyed directly via smartphone (Couper, 2017).

In our research, we combine these two possibilities: in a case study conducted in Germany in November 2018, we recruited participants from a mobile random digit dialing (RDD) sample via text messages that included a link to a Web survey. To the best of our knowledge, the present study is the first to investigate and systematically describe this approach and to illustrate potential caveats and challenges.

To explore the feasibility of our text-to-Web approach, the present study focuses on three steps in the data collection process: sampling, invitation via text message, and Web surveying. In what follows, we first describe the challenges and outcomes of each of these steps. In conclusion, we discuss the implications of our findings for data collection practice and further research.

3.2 Data collection process

3.2.1 Sampling

We used mobile RDD sampling to generate a random selection of mobile phone numbers of the general population. As mobile RDD sampling is prone to producing a substantial proportion of unassigned numbers (Häder and Sand, 2019), we used an existing sampling frame (see appendix). Furthermore, home location register (HLR) lookups were carried out to reduce the proportion of unassigned numbers in our sample (Sand, 2017, see also appendix).

When using mobile RDD sampling for a smartphone survey for the general population, consideration should be given on two different aspects: first, it must be considered that this method does not allow for sampling adults who are not reachable via cellphone, which is accompanied with the problem of "undercoverage" (i.e., the existence of elements in the target population "that do not, or cannot, appear in the sampling frame"; Groves et al. (2011), p.72). Second, using RDD sampling for a smartphone survey raises the problem of not being able to contact some of the units covered by the sampling frame, for example, adults who own a cellphone but not a smartphone.

Because the share of cellphones that are not smartphones in the general population in Germany is small (Silver, 2019), and no other possibility of drawing a random sample of smartphone numbers exists, we decided to implement this method of sampling in our study.

We drew a simple random sample of 30,102 cellphone numbers from our sampling frame. However, the HLR lookups for two of the four German providers (Telekom D1 and E-Plus) failed in the case of an improbably high proportion of numbers (59 and 24 percent, respectively) to provide information about whether the number was (un)assigned. Although it is known that the outcomes of HLR lookups may vary depending on the HLR lookup provider (Sand, 2017), we are not aware of any study using mobile RDD sampling where HLR lookups produced defective results for entire networks. Therefore, we assume that the problem was due to the way in which the sampled numbers were verified by the HLR lookup provider. We were unable to solve this problem, even after we contacted the institute that conducted the HLR lookups. As we had to pay for every invitation short message service (SMS) sent (and not only for those that were successfully delivered), we decided to exclude the cellphone numbers of the aforementioned networks from our study to maximize the proportion of delivered invitations within our given budget.

However, as these two networks together accounted for a relatively high proportion of all assigned numbers in Germany in 2018[2] (Bundesnetzagentur, 2019), this decision

[2]At the time of data collection in November 2018, the Telekom network's share of the German

may have had a profound impact on the composition of our sample because excluding specific networks can lead to systematic sample bias if the users of these networks differ systematically from those of other networks. After the HLR lookups, 13,338 (44.3 percent) numbers were left in our sample.

3.2.2 Text message invitation

To send invitations via text messages, we used the SMS protocol. The advantage of SMS is that it is a standard installation on every smartphone, and, therefore, all smartphone owners can receive this kind of text message (Dal Grande et al., 2016). The main disadvantage is the limit of 160 characters per message (Bosnjak et al., 2008). Due to cost restrictions, we could send only one invitation message per number, and we could not send any reminders, which made it impossible to provide detailed information on the survey mentioned in the invitation.

To increase the salience of the message and arouse the interest of potential respondents, we added the generic subject line "Life in Germany" to the SMS (Porter and Whitcomb, 2005; Sappleton and Lourenço, 2016). The name of the survey sponsor "GESIS" was displayed as the sender of the SMS, as we assumed that the fact that the sponsor was a public institution would increase the response rate (Dillman et al., 2014). The text message consisted of a short introductory sentence and the link to the survey in form of a tiny URL. The English translation of the invitation SMS is found in figure 3.1.

---

mobile network was 32.9 percent. Since the merger of E-Plus and Telefonica O2 in October 2014, the federal network agency (Bundesnetzagentur) reports only joint data for these networks. However, as E-Plus was the third-largest network provider in Germany at the time of the merger, its share of the German mobile network was still substantial in 2018.

Some problems arose with the delivery of the invitation messages. Whenever an SMS was successfully delivered, it was designated as such in the delivery report of the bulk SMS provider. Therefore, we were able to distinguish between delivered and undelivered SMS. In our case, however, only 6,016 of the 13,338 SMS sent could be successfully delivered. Despite our investigations into why this happened (see appendix), the reason for the low delivery rate remains unclear.

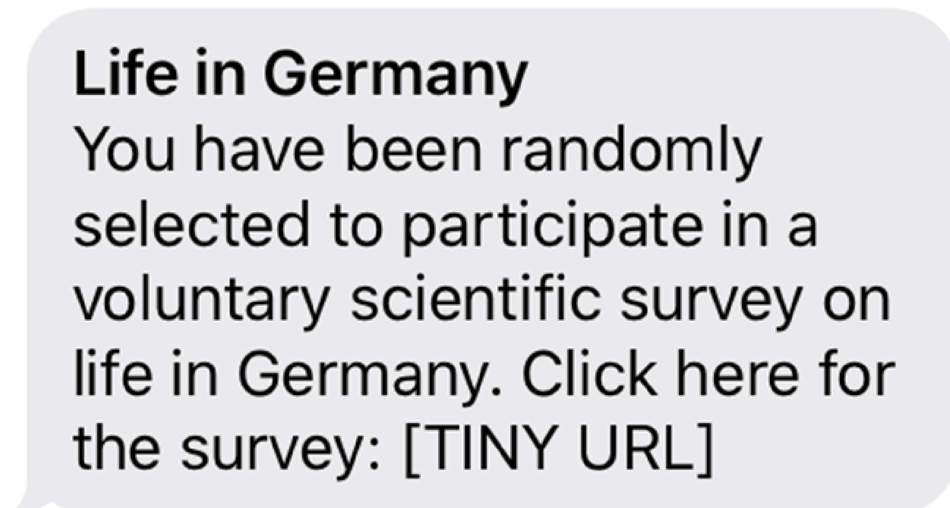

Figure 3.1: Text message invitation to participate in the survey

3.2.3 Web survey

By clicking on the link embedded in the invitation message, invitees with a smartphone were directed to the landing page of our survey on the smartphone's Internet browser. As the SMS invitation was limited to 160 characters, we used the landing page to obtain consent, which resulted in a lengthy page.

The questionnaire itself consisted of 36 items on sociodemographic characteristics, political attitudes, and smartphone usage. To ensure the best possible presentation of the survey on smartphones, we used a mobile-friendly design. We also collected server-side paradata that included information on the device used by the respondent and on response behavior (e.g., user agent (UA), response time, date of first/last access).

The landing page of our survey was accessed 998 times, which is only about one-sixth of the number of successfully delivered invitation SMS. However, only 161 accesses also started the survey — that is answered at least one question. To get a deeper understanding of this high rate of implicit refusals rate (AAPOR, 2016), we analyzed the UAs of the landing page accesses (see appendix). This analysis showed that most of the accesses that did not start the survey came from automated, non-human visits to the landing page and could be attributed to link previews (n = 673). As a UA is automatically transmitted when a link preview is displayed on a smartphone, it does not mean that the landing page of the survey was accessed by the invitee. A total of 44 landing page accesses were found to have been made by bots.

Of the 281 humans who intentionally visited the landing page, 120 did not take part in our survey. As can be seen from table 3.1, a relatively high proportion of the respondents who answered at least one question (n = 161) also completed the survey (n = 102) — that is, answered >80 percent of all applicable questions (AAPOR, 2016). Fifty respondents were classified as "break-offs" (AAPOR, 2016) — that is, they answered less than 50 percent of all applicable questions. Nine respondents were classified as "partials" (AAPOR, 2016) — that is, they answered between 50 and 80 percent of all applicable questions.

3.3 Results

3.3.1 Calculation of outcome rates

To be able to report outcome rates that are comparable with those of other surveys, we used standardized final disposition codes proposed by American Association for Public Opinion Research [AAPOR (2016)][3]. The AAPOR levels of eligibility and completion to which cases were assigned, and the number of cases assigned to each level, are shown in table 3.1. We did not include the accesses made by bots and link previews in the calculation of outcome rates as no numbers were generated for these accesses and no SMS were sent to them.

Following American Association for Public Opinion Research (AAPOR, 2016),

> "Response Rate 1 (RR1), or the minimum response rate, is the number of complete interviews [I] divided by the number of interviews (complete [I] plus partial [P]) plus the number of non-interviews (refusal and break-off [R] plus non-contacts [NC] plus others [O]) plus all cases of unknown eligibility (unknown if housing unit [UH] plus unknown, other [UH])" (AAPOR, 2016, p.61).

[3] Furthermore, we have drawn upon the work of Callegaro et al. (2007), who adapted the AAPOR final disposition codes for smartphone surveys.

Table 3.1: Assignment of cases to American Association for Public Opinion Research (AAPOR) (2016) levels of eligibility and completion

Data collection steps	Status	AAPOR description	AAPOR classification	AAPOR code	Category assigned to caluclate the response rate	n(\%)
Sampling	Generated cellphone numbers				Total sample used	30,102 (100%)
	Status of HLR lookup: unassigned number	Non-working/disconnected number	Not eligible	4.30		16,764 (55.7%)
Text message invitation	Status of HLR lookup: assigned number, but undelivered message	Cannot be classified as either eligible or ineligible, Callegaro et al. (2007)	Unknown eligibility	3.25	UO	7,322 (24.3%)
	Status of HLR lookup: assigned number, delivered message, but no access to survey landing page	Delivered text message should count as a call attempt, Callegaro et al. (2007)	Non-contact	2.20	NC	5,735 (19.1%)

Table 3.1: Assignment of cases to American Association for Public Opinion Research (AAPOR) (2016) levels of eligibility and completion *(continued)*

Web survey	Access to survey landing page but no survey participation	Visit to the Internet survey URL, but no participation	Implicit refusal	2.10	R	120 (0.4%)
	Participation, but <50% of the questionnaire completed	Answers to <50% of all applicable questions	Break-off	2.12	R	50 (0.2%)
	Participation, but only 50–80% of the questionnaire completed	Answers to 50–80% of all applicable questions	Partial interview	2.11	P	9 (0.03%)
	Participation, >80% of the questionnaire completed	Answers to >80% of all applicable questions	Complete interview	1.00/1.10	I	102 (0.3%)

Note: I = complete interview; NC = non-contact; P = partial interview; R = refusal and break-off; UO = unknown, other, (AAPOR,2016, p. 61).

Based on the classification of cases (see table 3.1), the following RR 1 can be reported for our survey:

$$RR1 = \frac{1}{(I+P)+(R+NC+O)+(UH+UO)} = 0.8\%$$

,

The response rate was extremely low. Although other mobile phone surveys conducted in Germany have shown low response rates[4], this result indicates that using our approach of text-to-Web recruitment for survey research can be expected to lead to greater and more far-reaching eligibility issues and a higher number of refusals than other survey research methods that use smartphones.

3.3.2 Characteristics of the survey respondents

The persons who took part in our survey were primarily young and highly educated. Over 50 percent of the respondents were under 30 years of age, and most of the respondents had a high level of education. In addition, we found that our respondents differed from those in other survey modes in terms of their political attitudes. From the answers to the question about voting intention, it became apparent that supporters of nontraditional parties on the political margins (DIE LINKE, AfD) were clearly overrepresented in our sample, compared with two other surveys that were conducted around the same time—namely, the Politbarometer 2018 (Forschungsgruppe Wahlen, 2019), a telephone survey, and the German Longitudinal Election

[4]For example, the response rate (RR1) for the mobile phone sub-sample of the CATI survey conducted as part of the German Longitudinal Election Study 2017 (GLES, 2019) was 4.2% (this is our own calculation based on disposition codes provided by GLES).

Study (GLES) Panel survey, which is a mixed-mode panel study (Roßteutscher et al., 2020). Figure 3.2 gives an overview of the sociodemographic characteristics of our survey respondents compared with census data (for the underlying figures, see tables 6.3, 6.4, and 6.5 of the appendix). The voting intentions of our survey respondents compared with those of the Politbarometer and GLES respondents are shown in figure 3.3 (for the underlying figures, see table 6.6 of the appendix).

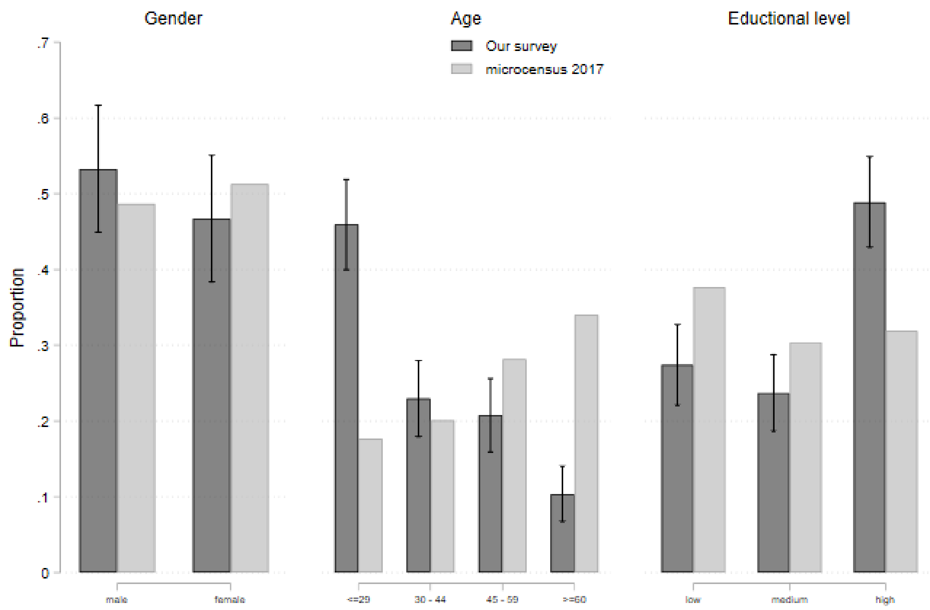

Figure 3.2: Sociodemographic characteristics of our survey respondents compared with data from the German Microcensus 2017

3.4 Discussion

The aim of the present study was to explore the feasibility of recruiting respondents and conducting Web data collection via smartphone. For this purpose, we carried

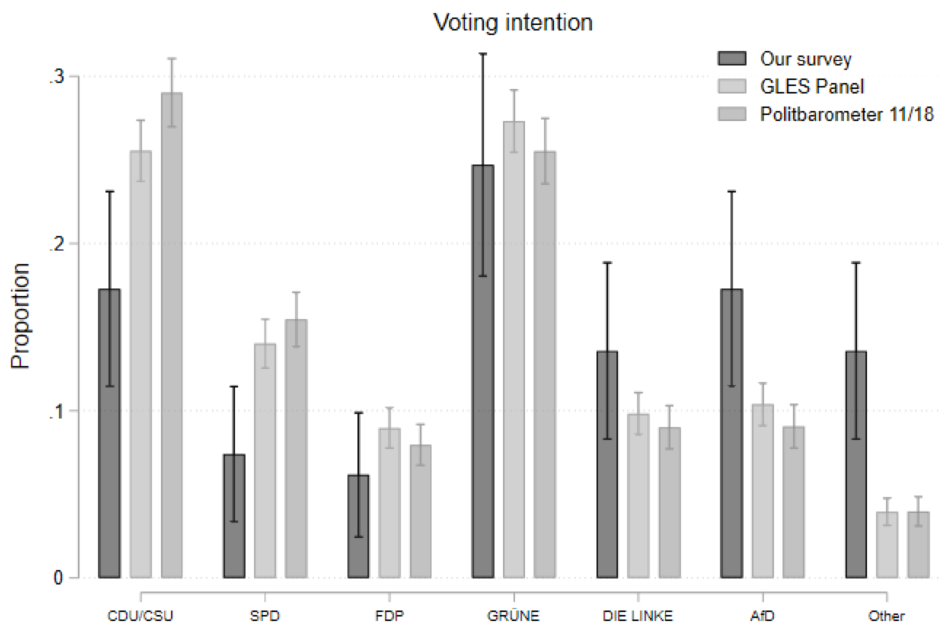

Figure 3.3: Voting intentions of our respondents compared with those of Politbarometer and GLES Panel respondents

out a case study in Germany in November 2018. A central finding of our study is that, although recruiting respondents for Web surveys via text messaging to smartphones is feasible, this approach is hampered by considerable issues relating to its implementation and the resulting data. On the positive side, RDD mobile sampling is easy and quick to implement.

However, in our study, we had to exclude two network providers because the HLR lookups failed in the case of an improbably high proportion of numbers to provide information about whether the number was (un)assigned. This exclusion may have had far-reaching implications in terms of data quality as it potentially introduced biased estimates based on the sample. This potential for bias is due to the fact that users of different network providers may differ systematically from each other. However, the extent to which this is the case is unclear, as no data are available to date on users of different network providers.

This is a task for further research. Furthermore, the coverage of network providers differs across Germany. Especially in some rural areas, only one provider can provide a service, which may further impact composition of a survey sample (See Bundesnetzagentur, 2020). In addition, RDD mobile sampling does not allow a random sample consisting only of smartphone numbers to be generated. And finally, our text-to-Web approach did not work for all elements in our sample, as cellphones that were not smartphones did not offer the possibility of clicking on the link embedded in our invitation SMS.

The sending of survey invitations via text messages is fast, relatively cheap and does not require a large operational effort. However, when combining this invitation mode with mobile RDD sampling, it must be assumed that there also will be a high rate of undelivered text messages—in our study, only half of the SMS messages sent could be successfully delivered. As we were unable to determine why the text messages were not delivered, our study does not provide much information on the implications of this high rate of undelivered SMS for data quality. One possible explanation for the high number of undelivered messages is that our invitation may have been blocked or marked as spam by network operators because we sent a considerable number of SMS messages using the same bulk SMS provider. We would therefore recommend that future studies use several different providers to send the invitations to prevent the SMS messages being blocked or marked as spam by the network operators.

A challenge that arose during the data collection process was that the Web survey landing page was only rarely accessed by humans (n = 281), and only 161 of these persons actually started the survey. Thus, the question arises as to whether the low willingness to participate was due to the design of our invitation, or whether the use of text message invitations led to high nonresponse. To answer this question, further research must experimentally test different invitation designs and content. Furthermore, the willingness to participate could be increased by sending reminders (Cook et al., 2016). In our study, this was not possible due to the limited budget. Despite these challenges, the increasing use of smartphones in the general population, and the technical versatility of these devices, means that they offer new opportunities for survey research. To our knowledge, our study is the first to explore and illustrate the various stages of conducting a text-to-Web survey combined with mobile RDD sampling. By detailing the different steps and potential pitfalls, we have thus laid the foundation for future studies.

An issue that was not addressed in this study is who can be reached via smartphone. Although smartphone ownership is increasing, some demographic subgroups still cannot be adequately reached via this device. Overall, some evidence seems to indicate that people who can be reached via smartphone differ systematically from those who do not own such a device (Couper, 2017; Couper et al., 2017; Baier et al., 2019; Keusch et al., 2019). Therefore, the effect that facilitating the recruitment of smartphone users has on the overall quality of the survey data remains open (Couper et al., 2017). Future investigations should be undertaken to focus on questions regarding the generalizability of findings based on data collected by using text-to-Web recruitment for general population studies.

References

AAPOR. 2016. "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys." The American Association for Public Opinion Research.

Baier, Tobias, Anke Metzler, and Marek Fuchs. 2019. "Coverage Error in Smartphone Surveys Across European Countries." General Online Research. Köln.

Battestini, Agathe, Vidya Setlur, and Timothy Sohn. 2010. "A Large Scale Study of Text-Messaging Use." In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '10*, edited by Marco de Sá, Luís Carriço, and Nuno Correia, 229. New York, New York, USA: ACM Press. https://doi.org/10.1145/1851600.1851638.

Bosnjak, Michael, Wolfgang Neubarth, Mick P. Couper, Wolfgang Bandilla, and Lars Kaczmirek. 2008. "Prenotification in Web-Based Access Panel Surveys." *Social Science Computer Review* 26 (2): 213–23. https://doi.org/10.1177/0894439307305895.

Bucher, Hannah, and Matthias Sand. 2022. "Exploring the Feasibility of Recruiting Respondents and Collecting Web Data via Smartphone: A Case Study of Text-To-Web Recruitment for a General Population Survey in Germany." *Journal of Survey Statistics and Methodology* 10 (4): 886–97. https://doi.org/10.1093/jssam/smab006.

Bundesnetzagentur. 2019. "Teilnehmerentwicklung Im Mobilfunk."

———. 2020. "Kartenansicht-Funkloch."

Callegaro, Mario, Charlotte Steeh, Trent D. Buskirk, Vasja Vehovar, Vesa Kuusela, and Linda Piekarski. 2007. "Fitting Disposition Codes to Mobile Phone Surveys: Experiences from Studies in Finland, Slovenia and the USA." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (3): 647–70. https://doi.org/10.1111/j.1467-985X.2006.00461.x.

Church, Karen, and Rodrigo de Oliveira. 2013. "What's up with Whatsapp?" In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services - MobileHCI '13*, edited by Michael Rohs, Albrecht Schmidt, Daniel Ashbrook, and Enrico Rukzio, 352. New York, New York, USA: ACM Press. https://doi.org/10.1145/2493190.2493225.

Cook, David A., Christopher M. Wittich, Wendlyn L. Daniels, Colin P. West, Ann M. Harris, and Timothy J. Beebe. 2016. "Incentive and Reminder Strategies to Improve Response Rate for Internet-Based Physician Surveys: A Randomized Experiment." *Journal of Medical Internet Research* 18 (9): e244. https://doi.org/10.2196/jmir.6318.

Couper, Mick P. 2017. "New Developments in Survey Data Collection." *Annual Review of Sociology* 43 (1): 121–45. https://doi.org/10.1146/annurev-soc-060116-053613.

Couper, Mick P., Christopher Antoun, and Aigul Mavletova. 2017. "Mobile Web

Surveys." In *Total Survey Error in Practice*, edited by Paul P. Biemer, Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West, 28:133–54. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119041702.ch7.

Dal Grande, Eleonora, Catherine Ruth Chittleborough, Stefano Campostrini, Maureen Dollard, and Anne Winifred Taylor. 2016. "Pre-Survey Text Messages (SMS) Improve Participation Rate in an Australian Mobile Telephone Survey: An Experimental Study." *PloS One* 11 (2): e0150231. https://doi.org/10.1371/journal.pone.0150231.

destatis. 2019. "Daten Aus Den Laufenden Wirtschaftsrechnungen (LWR) Zur Ausstattung Privater Haushalte Mit Informationstechnik."

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th edition. Hoboken: Wiley.

Forschungsgruppe Wahlen, Mannheim. 2019. "Politbarometer 2018 (Kumulierter Datensatz)." GESIS Data Archive. https://doi.org/10.4232/1.13420.

GLES. 2019. "Rolling Cross-Section Campaign Survey with Post-election Panel Wave (GLES 2017)Rolling Cross-Section-Wahlkampfstudie mit Nachwahl-Panelwelle (GLES 2017)." GESIS Data Archive. https://doi.org/10.4232/1.13213.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2011. *Survey Methodology*. 2nd ed. Hoboken: John Wiley & Sons.

Häder, Sabine, and Matthias Sand. 2019. "Telefonstichproben." In *Telefonumfragen in Deutschland*, edited by Sabine Häder, Michael Häder, and Patrick Schmich, 39:113–51. Schriftenreihe Der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23950-36.

Hall, Amanda K., Heather Cole-Lewis, and Jay M. Bernhardt. 2015. "Mobile Text Messaging for Health: A Systematic Review of Reviews." *Annual Review of Public Health* 36: 393–415. https://doi.org/10.1146/annurev-publhealth-031914-122855.

Keusch, Florian, Mariel M. Leonard, Christoph Sajons, and Susan Steiner. 2019. "Using Smartphone Technology for Research on Refugees: Evidence from Germany." *Sociological Methods & Research* 1 (3): 004912411985237. https://doi.org/10.1177/0049124119852377.

Pew. 2019. "Mobile Fact Sheet." https://www.pewresearch.org/inter- net/fact-sheet/mobile/.

Porter, Stephen R., and Michael E. Whitcomb. 2005. "E-Mail Subject Lines and Their Effect on Web Survey Viewing and Response." *Social Science Computer Review* 23 (3): 380–87. https://doi.org/10.1177/0894439305275912.

Roßteutscher, Sigrid, Rüdiger Schmitt-Beck, Harald Schoen, Bernhard Weßels, and Christof Wolf. 2020. "GLES Panel 2016-2019, Wellen 1-10." GESIS Data Archive. https://doi.org/10.4232/1.13475.

Sand, Matthias. 2017. "Evaluierung von HLR-Lookup-Verfahren." In *Methodische Probleme von Mixed-Mode-Ansätzen in Der Umfrageforschung*, edited by Stefanie Eifler and Frank Faulbaum, 211–37. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15834-79.

Sappleton, Natalie, and Fernando Lourenço. 2016. "Email Subject Lines and Response Rates to Invitations to Participate in a Web Survey and a Face-to-Face Interview: The Sound of Silence." *International Journal of Social Research Methodology* 19 (5): 611–22. https://doi.org/10.1080/13645579.2015.1078596.

Silver, L. 2019. "Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally: In Emerging Economies, Technology Use Still Much More Common Among Young People and the Well-Educated."

4 Enhancing Model-Based Adjustments of Nonprobability Surveys: Selecting Auxiliary Variables Based on Theoretical Assumptions about their Association with Survey Participation and Variables of Interest[1]

Abstract

Nonprobability surveys, have become increasingly popular for social science research based on observational data. Since nonprobability samples are based on a non-random (self-)selection of respondents, they are prone to selection bias, that is, variables affecting the selection of units in the sample as well as the outcome of variables

[1]The study was conducted with Joss Roßmann.

of interest in the survey (Elwert and Winship, 2014). To reduce this bias, we need powerful model-based adjustments. However, the auxiliary variables used for adjustments are frequently limited to a small subset of sociodemographic variables. Recent research suggests these variables hardly help minimize bias in nonprobability surveys. One proposed explanation for the often-poor performance of survey adjustments is that variables used for these adjustments hardly correlating with survey participation, as well as variables of interest. In our study, we argue that survey researchers should select and include questions on auxiliary variables in their surveys based on theoretical assumptions about the links to survey participation and substantive variables of interest and introduce a six-step theory-based approach for selecting adjustment variables. Further, we proceed with two empirical examples in that we applied our six-step approach to developing post-surveys adjustment for nonprobability surveys based on theoretical consideration. To evaluate the effectiveness of the post-survey adjustments, we compare external benchmarks to estimates derived from models based on (1) unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights. Although promising in theory, we saw that the approach did not perform as desired concerning reducing bias in estimates obtained from the nonprobability survey data in both cases.

4.1 Introduction

Nonprobability online surveys have gained popularity in survey research as they are easy to conduct, cheap, and fast. Especially nonprobability surveys based on self-selection samples (Bethlehem, 2016) where respondents usually volunteer to participate in surveys, for instance, by registering on an Internet platform, become more widespread in survey research. Here, respondents self-select into samples rather than being randomly selected (Marsden and Wright, 2010). From a theoretical point of view, inference to the target population is only applicable for these kinds of surveys under additional distributional assumptions on variables in the target population as they do not rely on probability-based sampling methods (Mercer et al., 2017). Importantly, owing to the lack of control over the process of selection of respondents into the survey, the risk of selection bias is high for nonprobability surveys. Selection bias occurs when variables that have an impact on the survey participation of units also affects the outcome of variables of interest (Elwert and Winship, 2014; Rohrer, 2018; Wysocki et al., 2022). Thus, the quality of data from nonprobability surveys is uncertain and researchers must rely on statistical models to control for selection bias when they intend to generalize on a defined target population (Mercer et al., 2017).

Recently, researchers have made efforts to develop such statistical models that aim at reducing selection bias in nonprobability surveys and allow for inferences to be made about target populations (Loosveldt and Sonck, 2008; Valliant and Dever, 2011; Ghitza and Gelman, 2013a; Wang et al., 2015; Trangucci et al., 2018; Kennedy and Gelman, 2021). Crucial for the success of these models in reducing selection bias in nonprobability surveys is the selection of adjustment variables. These adjustment variables need to affect both selection into the sample as well as the variables of interest. If the selected adjustment variables do not fulfill this requirement and the models applied to control for selection bias do not include the variables that affect sample selection, as well as variables of interest, then estimates from the survey will likely be biased. As a result, any conclusions drawn from the survey about the target population may be biased[2].

However, up to the present rather little attention has been paid to theoretical considerations in selecting variables for use in adjustment models. Also, empirical tests of whether the chosen variables affect both, selection into a sample, and the outcomes of interest have been sparse.

In the present study, we address this shortcoming by highlighting the importance of providing theoretical considerations about relationships between variables used in adjustments, and the need for empirical examinations on whether the underlying assumptions hold before running post-survey adjustment models. For this purpose, we introduce a six-step theory-based approach for selecting adjustment variables. We further conduct two empirical examples of developing post-surveys adjustment for nonprobability surveys based on our six-step theory-based approach. The goal of our study is to contribute to the further enhancement of post-survey adjustments for nonprobability surveys.

[2]The variables of interest being independent of the selection of respondents into the survey sample conditional upon a set of covariates is also called "missing at random" (MAR)(Rubin, 1976; Schouten, 2007; Little et al., 2020). Including adjustment variables in models that correlate with survey participation, as well as variables of interest, we assume that selection bias can be approximate to be "missing completely at random" (MCAR), which means that selection in the sample is independent of the survey questions (Schouten, 2007).

4.2 Model-based adjustments to reduce selection bias in nonprobability surveys

There is a growing body of literature investigating how selection bias in nonprobability surveys can be reduced by post-survey adjustments (Park et al., 2004; Schonlau et al., 2007; Loosveldt and Sonck, 2008; Valliant and Dever, 2011; Ghitza and Gelman, 2013a; Pasek, 2016; Elliott and Valliant, 2017; MacInnis et al., 2018; Iachan et al., 2019; Kennedy and Gelman, 2021; Little et al., 2020). A common and widely used method that aims at correcting for selection bias is the computation of adjustment weights for use in statistical analyses (Wolf et al., 2016). This post-survey adjustment methods balances the actual distribution of variables in a survey sample to the known distribution of those variables in the target population. This approach is also referred to as “superpopulation modeling” (Valliant, 2020).

A variety of statistical procedures can be applied to compute adjustment weights, for instance, multilevel regression and poststratification (MRP) (Park et al., 2004; Ghitza and Gelman, 2013a; Kennedy and Gelman, 2021), propensity matching (Valliant and Dever, 2011; MacInnis et al., 2018), raking (Iachan et al., 2019), or - more recently - machine learning approaches (Pasek, 2016; Ferri-García et al., 2021; Kim et al., 2021). In most cases, the weights are computed based on a selection of socio-demographic variables, such as age, gender, educational attainment, ethnicity (Loosveldt and Sonck, 2008; Pasek, 2016), or geographical information such as state (Park et al., 2004; Ghitza and Gelman, 2013a; MacInnis et al., 2018). A primary reasons for the frequent use of socio-demographics is that this information is commonly available for both respondents and nonrespondents. Further, some studies also included health-related information in adjustment models (Valliant and Dever, 2011; Iachan et al., 2019).

Focusing merely on the statistical computation of survey weights to correct for selection bias in nonprobability surveys, most of these approaches lack links to theory in the selection of adjustment variables. However, post-survey adjustments are a promising tool to overcome selection bias in nonprobability surveys only when the variables used in the models influence both sample selection and the survey outcomes of interest (Little, 1986; Little and Vartivarian, 2005; Groves, 2006; Valliant and Dever, 2018). Groves et al. (1992) take this argumentation even one step further by stating that post-survey adjustment models "are themselves theories of survey participation but are typically the result of"making do" with what variables have been measured in the survey, not measures of the causes of participation in the survey at hand", and "that the specification of such theories should inform the adjustment process" (Groves et al., 1992, p. 476). Building on this line of argumentation, we propose that more effort should be invested in justifying the selection of variables for use in post-survey adjustment models. Also, we intend to raise awareness for the challenge involved in a theory-based selection of adjustment variables for nonprobability surveys. Accordingly, we suggest that survey researchers should invest effort in presenting the assumed theoretical interrelations between adjustment variables, the selection mechanism, and the survey outcomes of interest. Further, we want to encourage researchers to empirically test whether their assumptions about these interrelations hold.

An essential challenge in the selection of adjustment variables is the existence and availability of essentially error-free estimates for these parameters in the target population (cf., Biemer, 2010). However, even if such influential adjustment variables exist, error-free estimates for them are often unavailable (Park et al., 2004; Schonlau et al., 2009; Valliant and Dever, 2011), or they are just not considered for the calculation of post-survey adjustments, such as weighting factors (Peytchev et al., 2018).

To summarize, the variables used in selection bias adjustment models for nonprobability surveys need to satisfy three requirements: First, adjustment variables need to be correlated with the mechanism of selection into the sample. Second, adjustment variables need to be related to the outcome variables of interest. And third, essentially error-free estimates of the adjustment variables must be available for all elements of the target population. Importantly, as surveys differ with respect to their topics, there will be no universal set of adjustment variables that fully satisfies all three requirements equally well: Any adjustment variable will inevitably differ in its relationship with the selection mechanism and the large universe of variables of interest from different topical areas. Thus, the central challenge for survey research lies in the identification of adjustment variables that propose to satisfy these requirements best concerning the specific survey that is being conducted. Consequentially, this implies that survey researchers have to adapt their adjustment models to the specifics of the survey they conduct (i.e., the survey's topic, the central variables of interest, or the survey's context). Figure 4.1 gives an overview of the requirements that have to be satisfied by adjustment variables.

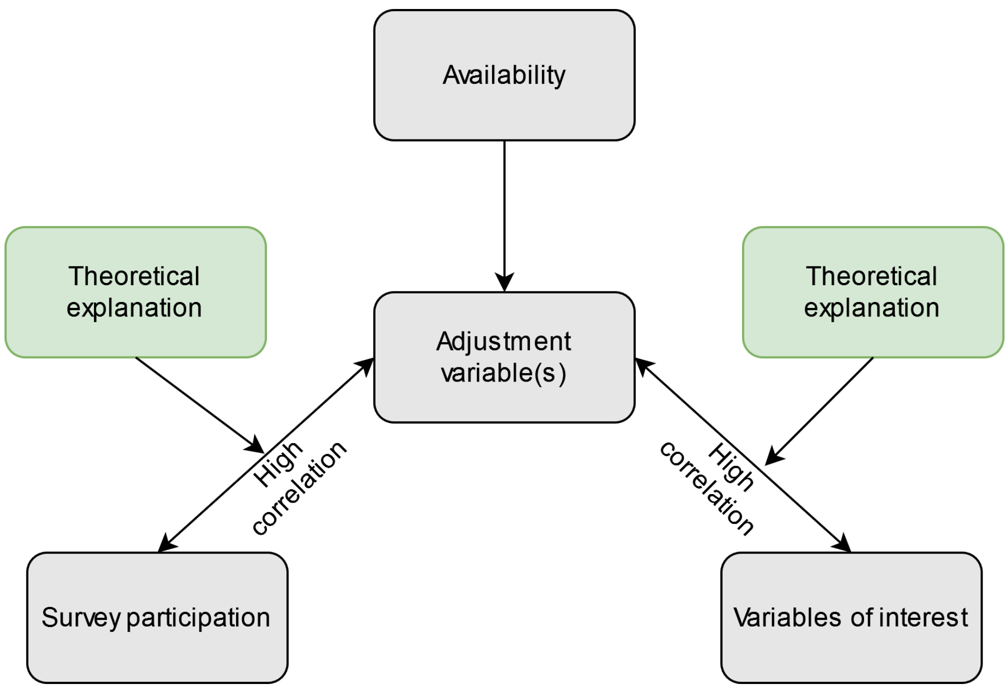

Figure 4.1: Requirements for selection bias adjustment variables

4.3 Enhancing model-based adjustments for nonprobability surveys: A six step approach

We propose a theory-based strategy for selecting adjustment variables that gears towards strong correlations of these variables with survey participation, as well as substantive survey questions. In what follows, we present a six step approach that practically guides survey researchers in implementing the approach. Table 4.1 gives an overview of each step of the approach.

Table 4.1: Enhancing model-based adjustments for nonprobability surveys: A six step approach

1.	Select adjustment variables based on theoretical arguments on the proposed correlation with survey participation and variables of interest.
2.	Check whether there is data available to adjust the survey data. Is there benchmark data that can provide error-free estimates of the adjustment variables?
3.	Include the adjustment variables in the survey program.
4.	Collect data.
5.	Empirically check the assumptions: Do the adjustment variables correlate highly with (1) survey participation and (2) substantial variables of interest?
6.	If the assumed correlations can be found in the data, use the theoretically selected variables to compute the post-survey adjustments.

4.3.1 STEP 1. Select adjustment variables based on theoretical arguments on the proposed correlation with survey participation and variables of interest

This first step is the most crucial in implementing the proposed approach. Specifically, the difficulty is finding variables that highly correlate with, first, survey outcomes and, second, survey participation.

First, as surveys widely differ in their topics, adjustment variables should be selected with regard to the concepts of interest that are measured in a survey. Therefore, a profound knowledge about a survey's topic is a prerequisite for finding suitable adjustment variables. We suggest close cooperation between survey methodologists and researchers with a substantial interest in the surveys topic to find adjustment variables that are strongly related to the central concepts in a survey.

Second, theoretical frameworks of survey participation should inform the selection of adjustment variables: The prominent general framework by Groves et al. (1992) distinguishes between societal-level factors, survey design attributes, and the sampled person's characteristics. As the impact of these three factors on survey participation may vary across different setting, we propose that their relation to potential adjustment variables should be considered for each individual survey setting. In addition, other theoretical frameworks address the cognitive and psychological processes affecting survey response. To provide surveys researchers with guidance in considering suitable adjustment variables, we briefly outline the central theoretical statements of the most popular frameworks of survey participation and give some examples of their application methodological survey research (see Table 4.2).

According to Albaum and Smith (2012) (see also Albaum et al. (1998); Keusch (2015)), popular theories of survey participation include social-exchange (Blau, 1964), cognitive dissonance (Festinger, 2001), self-perception (Bem, 1972), and commitment/involvement theory (Becker, 1960). In addition, the theory of compliant behavior (Groves et al., 1992; Groves and Couper, 1998), leverage-saliency theory (Groves et al., 2000), and the theory of planned behavior (Hox et al., 1995; Bosnjak et al., 2005; Heerwegh and Loosveldt, 2009) have been applied to explaining survey participation. Finally, Voogt and Saris (2003) presented a theoretical explanation for participation in surveys on political attitudes and behavior rather than participation in surveys in general.

Table 4.2: Theoretical frameworks of survey participation

Theoretical framework	Central statement	Exemplary applications
Social exchange: Blau (1964)	Perceived benefits exceed the perceived costs of participation	Dillman (1991); Groves and Couper (1998); Dillman et al. (2014); Dillman (2022); Greenberg and Dillman (2023)
Cognitive dissonance: Festinger (2001)	Individual´s motivation to participate stems from avoiding negative feelings associated with nonresponse	Hackler and Bourgette (1973); Furse and Stewart (1982); Furse and Stewart (1984)
Self-perception: Bem (1972)	Individuals participate to keep their behavior consistent with their favorable self-perception	Allen et al. (1980); Hansen (1980); Evangelista et al. (2012)
Commitment/ involvement: Becker (1960)	Individuals who are highly committed to responding to surveys are more likely to respond	Albaum et al. (1998); Evangelista et al. (2012)
Planned behavior: Ajzen (1985); Ajzen (1991)	Participation as consequence of a behavioral intention to participate	Hox et al. (1995); Bosnjak et al. (2005); Heerwegh and Loosveldt (2009)
Compliant behavior: Cialdini and Goldstein (2004)	Response-acquiescence to a particular kind of communication - a request	Groves et al. (1992); Groves and Couper (1998)
Leverage-saliency: Groves et al. (2000)	Attributes of the survey are salient and a leverage for decision to respond	Marcus et al. (2007); DeCamp and Manierre (2016)
Social isolation and attachment to society: Voogt and Saris (2003)	Participation in a political survey is driven by social involvement and attachment to society	Sciarini and Goldberg (2015); Sciarini and Goldberg (2016); Walgrave et al. (2016)

Of note, a common feature of these theoretical frameworks is that they were developed for explaining survey participation in traditional sample surveys, often cross-sectional surveys that rely on a random selection of respondents[3]. That said, it

[3] E.g.,the examples we mentioned for the application of there theoretical frameworks all apply them to probability samples (Hackler and Bourgette, 1973; Allen et al., 1980; Hansen, 1980; Furse and

remains unclear to what extent these traditional approaches can be adopted for explaining participation in nonprobability surveys that sample their respondents from self-selective opt-in online panels. Processes of decisions for or against participation in a specific survey is most likely very different for these kinds of nonprobability surveys than for surveys that rely on established probability-based sampling methods.

4.3.2 STEP 2. Check whether there is data available to adjust your survey data to

After having identified adjustment variables that - from a theoretical perspective - should be correlated with survey participation as well as the variables of interest, researchers then need to find error-free estimates of these adjustment variables for the target population. Data from official statistics such as administrative data or data from mandatory censuses offer a best choice data source as these data are usually much less biased by sampling, coverage, nonresponse, or measurement errors. In addition, they are often available for many different target populations.

Stewart, 1982, 1984; Dillman, 1991; Groves et al., 1992; Hox et al., 1995; Albaum et al., 1998; Groves and Couper, 1998; Bosnjak et al., 2005; Marcus et al., 2007; Heerwegh and Loosveldt, 2009; Evangelista et al., 2012; Dillman et al., 2014; Sciarini and Goldberg, 2015; DeCamp and Manierre, 2016; Sciarini and Goldberg, 2016; Walgrave et al., 2016; Dillman, 2022; Greenberg and Dillman, 2023).

4.3.3 STEP 3. Include the adjustment variables in your survey program

To avoid differences in measurement between benchmark data and survey data, we recommend using an identical or at least very similar operationalization for the selected construct as in the data collection of the source (e.g., official registry data, census data).

4.3.4 STEP 4. Collect data

Collect survey data as planned.

4.3.5 STEP 5. Empirically check your assumptions: Are the adjustment variables suffiecently correlated with (1) survey participation and (2) substantial variables of interest?

In the fifth step, we propose to empirically check whether the assumed correlations of the selected adjustment variables with survey participation and variables of interest can be found in the data. There are a few studies examining the strength of correlations between adjustment variables and survey variables, as well as survey participation and the resulting ability to reduce bias. All these studies found that the ability to reduce bias increases with the strength of the correlation [Little and Vartivarian (2005); Geuzinge et al. (2000); Kreuter et al. (2007); Yan and Raghunathan (2007); Maitland et al. (2008); Kreuter and Olson (2011);]. Further, some studies examined the possibility of reducing bias for different kinds of correlations between adjustment variables and survey participation, as well as survey variables.

First, a simulation study by Little and Vartivarian (2005) shows adjustments do not decrease bias if the adjustment variables correlate strongly with survey questions but not so with survey participation. Second, another simulation demonstrates that already moderate associations between adjustment variables, survey variables, and survey participation are sufficient to decrease bias in tendency (Kreuter and Olson, 2011, p. 327). However, if the direction of this association differs between survey participation and survey variables, and the model does not include all relevant adjustment variables, this is not the case anymore (Kreuter and Olson, 2011). Thus, researchers not only have to consider the strength of associations between adjustment variables, survey variables, and survey participation, but also whether these associations are in the same direction.

To summarize, previous simulation studies show that the strength and direction of correlations between adjustment variables, survey variables, and survey participation affects the ability of post-survey adjustments to reduce bias.

4.3.6 STEP 6. If the assumed correlations can be found in your data, use the theoretically selected variables to compute your post-survey adjustments and weight the data

The last step includes the computation of survey weights. We recommend that variables to be included in the adjustment models should at least be moderately correlated with indicators of survey participation and the survey variables of interest.

When calculating post-survey adjustments, different methods can be applied (Valliant and Dever, 2018; Valliant et al., 2018). However, the present study is concerned with a strategy for selecting suitable adjustment variables as opposed to properties of different statistical procedures for computing weights. Thus, we will not give general recommendations on methods for the calculation of adjustment weights.

4.4 Part II: Empirical applications of the six step approach

In the following, we present two empirical applications of the six-step strategy: Both studies rely on surveys about political attitudes and behavior in Germany, based on samples from a German opt-in online panel (Respondi AG). We apply the six-step approach for selecting adjustment variables in both studies. To investigate whether the adjustments successfully correct for selection bias, we compare estimates for the outcome of two elections in Germany to (1) the unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights. We hypothesize that bias in these estimates is lowest for the survey data adjusted with the enhanced, theory-based weights for both studies. In what follows, we give a detailed description of how we implemented the adjustment variable selection process.

4.4.1 STUDY I: Using COVID-19 vaccination status for adjusting a nonprobabilty survey on politics and elections

In our first study, we used the respondents' COVID-19 vaccination status as an adjustment variable for adjusting a nonprobability survey on political attitudes and behaviors in Germany.

4.4.1.1 STEP 1. Proposed correlation of COVID-19 vaccination with survey participation and variables of interest

Concerning vaccinations against COVID-19, we assume complex relations with factors that drive the process of survey participation on the one hand, and political attitudes and behaviors on the other hand.

Let us first dive into to the assumed association of COVID-19 vaccination with survey participation. We assume a strong correlation here: We propose that voluntary participation in a survey and vaccination against COVID-19 can be attributed to compliant behavior(Cialdini and Goldstein, 2004; Wetzel and Hünteler, 2022).

First, both represent a particular kind of response to a particular request. Getting vaccinated was supported and recommended by many national governments. Also, the German government initiated a comprehensive campaign to foster COVID-19 vaccinations in the population[4]. With this, citizens were directly confronted with the request to get vaccinated. With receiving an invitation to participate in a survey also representing a request (Groves et al., 1992) to a particular kind of response,

[4]https://www.bundesgesundheitsministerium.de/presse/pressemitteilungen/neue-kampagne-gibt-84-gruende-fuer-corona-schutz.html

namely participation in a survey, we assume that both survey participation and vaccination for COVID-19 can be understood as compliant behavior. Second, Cialdini and Goldstein (2004) elaborate that norms play a pivotal role in compliant behavior. Regarding getting vaccinated against COVID-19, we assume norms impact an individual's decision (Agranov et al., 2021; Baeza-Rivera et al., 2021; Ryoo and Kim, 2021; Jaffe et al., 2022; Rabb et al., 2022). During the second phase of the Corona pandemic in 2021, the World Health Organization (WHO)) advised the general public to get vaccinated against COVID-19[5]. This recommendation was supported by many national governments and accompanied by the introduction of temporary rules and laws restricting social contacts and activities - such as visiting public places, going to the cinema/theater, or even restaurants- to those vaccinated against COVID-19. With these rules, non-vaccinated adults were sanctioned by being prohibited from participating in many areas of public life.

In the theory of compliant behavior, adopting a particular behavior to suffer from being sanctioned for non-behaving is defined as an injunctive norm (Cialdini and Goldstein, 2004). With this, getting vaccinated against COVID-19 can be understood as injunctive norm. Concerning survey participation, we also assume norms to influence the participation decision. According to the theory of compliant behavior, respondents tend to participate in surveys based on the assumption that they are obliged to because (1) other individuals would also participate if they had been asked to (social norms) and (2) they do not want to get sanctioned for non-participating (injunctive norms) (Groves and Couper, 1998; Cialdini and Goldstein, 2004). To

[5]The WHO thereby declared that "getting vaccinated is one of the most important things you can do to protect yourself against COVID-19, help end the pandemic and stop new variants emerging" (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines/advice).

summarize, we assume a strong correlation between survey participation and getting vaccinated against COVID-19, representing a request for a particular behavior and being based on norms.

Second, we also assumed strong associations between the likelihood of getting the COVID-19 vaccination and political attitudes and behavior. Getting vaccinated against COVID-19 or not was a highly polarized issue during the second phase of the Corona pandemic in Germany but also in other nations. The controversy was carried out in the public and political sphere (Sabahelzain et al., 2021). While many democratic national governments supported vaccination campaigns and enacted regulations that made participation in public activities dependent on vaccination status, others, often more populist governments, vehemently opposed getting vaccinated. Thus, we assumed that vaccinating for COVID-19 is related to higher support for democracy, higher satisfaction with democratic leaders, and individual voting decisions (El-Mohandes et al., 2021; Killgore et al., 2021; Latkin et al., 2021; Travis et al., 2021; Roberts et al., 2022).

In addition, previous research has consistently shown that attitudes toward COVID-19 vaccination were correlated with political attitudes (Edwards et al., 2021; El-Mohandes et al., 2021; Killgore et al., 2021; Latkin et al., 2021; Travis et al., 2021; Roberts et al., 2022; Schernhammer et al., 2022; Serani, 2022). Mostly, these studies provided an indication that adults holding conservative social attitudes also reported being less likely to get vaccinated compared to adults holding liberal-social attitudes (El-Mohandes et al., 2021; Killgore et al., 2021; Latkin et al., 2021; Travis et al., 2021; Roberts et al., 2022). Further, some studies suggested that a lack of vaccine acceptance was related to a lack of trust in the government (Schernhammer et al., 2022; Yuen, 2022), authorities and scientists (Lindholt et al., 2021; Travis et al., 2021), and positively connected to populist attitudes (Edwards et al., 2021). Finally, some studies found a correlation between attitudes toward COVID-19 vaccination and actual voting behavior with voters of populist parties are less likely to get vaccinated (Sabahelzain et al., 2021; Serani, 2022). Taken together, these studies provided consistent indications for a correlation between political attitudes and COVID-19 vaccinations across the globe[6]. We assumed that these findings also apply to the German context. Thus, we expected that COVID-19 vaccination status is strongly correlated to political attitudes and voting decisions.

4.4.1.2 STEP 2. Available registry data for COVID-19 vaccination

To calculate adjustment weights based on COVID-19 vaccination status, we use data provided by the Robert-Koch-Institute (RKI), the leading bio-medical research institution in Germany. In their "digital vaccination rate monitoring", the RKI published indicators of the vaccination rate in Germany per age group and federal state. Since this data is available daily, we computed the mean proportion of adults vaccinated over the data collection period of the nonprobability online survey.

An essential limitation of the RKI Data for our purposes was that they were only available for adults living in Germany and not the German electorate, which was the target population of our survey. However, due to the substantial overlap between the German electorate and the resident population[7], we assumed that the effect on the accuracy of the survey adjustment weights should be rather neglectable.

[6]For studies in US see: El-Mohandes et al. (2021); Killgore et al. (2021); Latkin et al. (2021); Roberts et al. (2022), Europe see: Lindholt et al. (2021); Schernhammer et al. (2022), Asia see:Yuen (2022), Australia see: Edwards et al. (2021).

[7]for more information, see: https://www.bundeswahlleiter.de/info/presse/mitteilungen/bundestagswahl-2021/01_21_wahlberechtigte-geschaetzt.html

4.4.1.3 STEP 3. Questions on COVID-19 vaccination included in our survey program

COVID-19 vaccination status was measured with the following question: “Have you been vaccinated against COVID-19 or not? By this we mean that you have been vaccinated with at least one dose of one of the vaccines approved in Germany.” Vaccination status was dummy coded with 1 meaning that respondents did get vaccinated and 0 that they did not get vaccinated.

4.4.1.4 STEP 4. Data collection

The data for this study comes from the German Longitudinal Election Study (GLES). Specifically, we use data from the GLES control group III, an online survey with 1,232 respondents from a commercial German opt-in online panel operated by Respondi.AG. The online panel members were recruited via advertisements placed on various Internet platforms (GLES, 2022). The respondents for the GLES survey were selected using quotas for gender, age, and level of education. Data collection started right after the Federal Election in Germany in 2021, and the fieldwork period was two weeks (2021-09-29 - 2021-10-12).

4.4.1.5 STEP 5. Correlations of COVID-19 vaccination with survey participation and political attitudes and behavior in our survey data

To test whether our assumption about the strong association between COVID-19 vaccination with survey participation and political attitudes and behavior holds, we compute pearson correlation coefficients, using the R command cor.

Following our theoretical argumentation, we assume COVID-19 vaccination status to correlate with a range of political attitudes, namely holding conservative or populist attitudes, as well as trust in parliament/institutions [8].

For the correlation between COVID-19 vaccination status and survey participation, we are not able to directly measure whether these two variables correlate with each other. We will measure this assumed correlation indirectly. As stated above, we assume COVID-19 vaccination status to correlate with participation in a scientific survey on political attitudes and behavior since both variables are associated with adhering to norms. Therefore, we assume adhering to norms to be higher among vaccinated adults. We use the item whether individuals believe that voting generally is a civic duty (Voogt and Saris, 2003), and whether respondents followed our instructions in an implemented attention check (Silber et al., 2022) to empirically check whether our assumption holds [9]. Table 4.3 shows the correlations between COVID-19 vaccination status with survey participation, as well as political attitudes and behavior.

Altogether, this table shows that the assumed correlations between being vaccinated against COVID-19 with survey participation and political attitudes and behavior are rather low in our data set. Concerning survey participation, the correlation between following the instructions on the attention check and being vaccinated against COVID-19 is almost zero (0.03). Further, to being vaccinated against COVID-19 and adhering to norms, we find a moderate positive correlation (0.22).

For political attitudes, we find weak negative correlations between being vaccinated against COVID-19 and conservative (-0.16), as well as populist attitudes (-0.14), in-

[8] For a detailed description of the operationalization of these variables, see appendix (C.2).
[9] For a detailed description of the operationalization of these variables, see appendix (C.2).

Table 4.3: Correlations of COVID-19 vaccination with survey participation and variables of interest.

Variable	Coefficient
Survey Participation	
Attention check	0.03
Voting norm	0.22
Political Attitudes	
Conservative	-0.16
Populist	-0.14
Trust in institutions	0.33
Political Behavior	
Voting: Populist	-0.32

Note:
Correlation with being vaccinated against COVID-19 (yes); Coefficient = Pearson correlation coefficient.

dicating that vaccinated individuals are less likely to hold conservative or populist attitudes. We further find a positive, moderate correlation between trust in institutions and being vaccinated against COVID-19 (0.33), indicating that vaccinated individuals are more likely to trust institutions than non-vaccinated. These findings are -in a directional sense- in line with what we expected. However, the correlations are weaker than we expected. The findings are in line in what we expected for voting for the populist party. We find a moderate negative correlation (-0.32) which indicates that vaccinated individuals are less likely to cast a ballot for the populist party.

Overall, the correlations are primarily weak and partly do not correspond to our assumptions. Accordingly, we suspect that the resulting enhanced, theory-based weights may be less effective in minimizing bias. Following our six-step approach we would, based on these weak correlations, recommend not to use the theoretically selected variables to calculate adjustment weights. However, the aim of this chapter is to present an empirical application of the six-step strategy. Therefore, we we will calculate the weights in the next step, to provide a detailed description of our approach.

4.4.1.6 STEP 6. Calculation of COVID-19 vaccination post-survey adjustments

We calculate all post-survey adjustment weights using the same approach as the GLES survey program, post-stratification (GLES, 2022). The actual post-stratification weighting factors are computed using iterative proportional fitting (IPF) (Deming and Stephan, 1940; Bergmann, 2011). This method gradually adjusts the actual distribution of the individual cells to the respective target distribution of the weighting variables[10].

Standard socio-demographic weights are provided by the GLES (GLES, 2022a). They are computed using the information on age, gender, and level of education. The actual distribution of variables in the data set is adjusted to the distribution of the reference from the German Microcensus 2019.

To compare the bias in estimates of (1) the unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights, we use two types of bias measurement: First, we calculate the absolute bias for individual variables as the difference between the estimate and the benchmark. Second, we calculate the absolute relative bias, which

[10]The process of adjustment is finished when the difference between the weighted marginal distribution of all factors and the target distribution undercuts the abort criterion of 0.057. In order to prevent extremely large weighting factors; the factors are trimmed to the quadruple mean value of the weighting variable (thus five) after every step of the iteration process.

is a measure of the average difference between the estimates in the survey and the outcomes of the election. Mathematically, the absolute relative bias is defined as:

$$AbsoluteRelativeBias(Y) = \sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)}{n}$$

, where

$$n$$

denotes the number of variables considered,

$$y_i$$

denotes the values of these variables observed in the target population, and

$$\hat{y}_i$$

denotes the values of these variables estimated by the survey. To calculate the absolute relative bias, we dichotomized the variables. As population parameters, we use validated benchmark data that covers voting behavior (turnout, party vote).

Figure 4.2 graphically illustrates the differences between the values of the variables in the benchmark data and (1) the unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights. The benchmark values of the variables estimated with the survey data are shown utilizing a horizontal line, whereas estimates with 95% confidence intervals are displayed for the survey data.

Altogether, we see hardly any differences in the three estimates. Regarding our hypothesis, namely that the bias of these estimates is lowest for the survey data adjusted with the enhanced, theory-based weights, let us first compare the estimates of survey data adjusted with the enhanced, theory-based weights with the unadjusted survey data. Here, we see that for estimating turnout, the proportion of voters of the left-wing party (DIE LINKE) and the social democratic party (SPD) this hypothesis fits with our data. However, the differences are minor. Further, we see that for some variables adjusting survey data with the enhanced, theory-based weights even increases the bias compared to the estimates of the unadjusted survey data. This is true for estimating the proportion of green party voters, voters of the populist party (AfD), and small splinter parties (other party). The picture is similar when comparing estimates based on the (2) survey data adjusted with standard socio-demographic weights and (3) survey data adjusted with the enhanced, theory-based weights. Here, for estimating the proportion of green party voters, left-wing party voters (DIE LINKE), voters of the social democratic party (SPD), and turnout, the enhanced, theory-based weights perform better than the standard socio-demographic weights. However, for estimating the proportion of voters of the conservative party (CDU/CSU), the liberal party (FDP), the right-wing party (AfD), as well as small splinter parties (other party), the standard socio-demographic weights perform better.

Table 4.4 quantifies the differences by reporting the (absolute) relative bias in the estimates for (1) the unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights. We find that none of the adjustment weights effectively helps in reducing the absolute relative bias. Adjusting survey data with the enhanced, theory-based weights does not impact the absolute relative bias. With this finding, we must

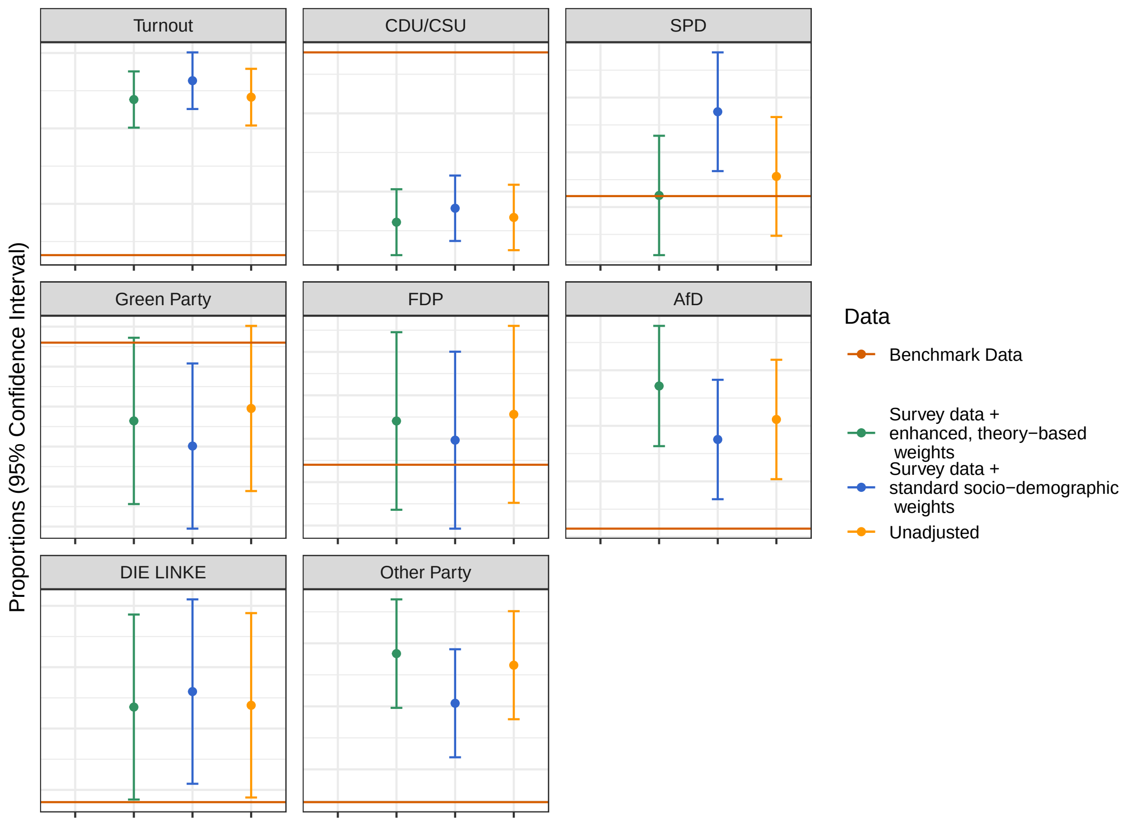

Figure 4.2: Point estimates in survey data compared to benchmark statistics

Table 4.4: Bias across the different survey weights

Variable	Unadjusted survey data	Survey data + standard socio-demographic weights	Survey data + enhanced,theory-based weights
Turnout	0.105	0.116	0.103
SPD	0.009	0.039	0.000
CDU/CSU	-0.106	-0.100	-0.109
Green Party	-0.016	-0.026	-0.020
FDP	0.012	0.006	0.010
AfD	0.039	0.032	0.051
DIE LINKE	0.016	0.018	0.016
Other Party	0.054	0.039	0.059
Total			
Absolute relative bias	0.014	0.015	0.014

reject our hypothesis that the bias is the slightest for the survey data adjusted with the enhanced, theory-based weights. We suggest this finding may be attributed to low correlations between COVID-19 vaccination status with variables of interest and survey participation, an empirical finding that is not in line with our theoretical expectation and findings from previous studies.

Moreover, adjusting survey data with standard socio-demographic weights increases the absolute relative bias. This finding cautions us that misspecification in weighting models can worsen things.

4.4.2 STUDY II: Using Internet usage patterns for adjusting a nonprobability survey on politics and elections

In our second study, we used information on the different ways of private Internet usage for adjusting a nonprobability survey on political attitudes and behaviors in Germany. With patterns of Internet usage, we refer to the manifold ways in which people make use of Internet-based services such as social media, online participation in political activities and processes, or searching for information in Web-based newspapers. In the following, we elaborate on why we assume that different patterns of Internet usage are related to survey participation, as well as political attitudes and behavior.

4.4.2.1 STEP 1. Proposed correlation of Internet usage patterns with survey participation and variables of interest

Concerning Internet usage patterns, we assume complex relations to survey participation on the one hand and political attitudes and behaviors, on the other hand.

First, we assume a positive association of survey participation with different Internet usage patterns. Concerning the assumed association of survey participation and social media usage, this assumption is based on the social isolation hypothesis (Voogt and Saris, 2003). According to this theory, social involvement is one main predictor of survey participation in political surveys. Voogt and Saris (2003) suggested that people involved in many social groups should be intensely interested in politics and highly likely to participate in surveys. Social involvement takes place offline and online, as social media can be understood as a specific kind of social group (Park et al., 2009; Skoric et al., 2009; Gil de Zúñiga and Valenzuela, 2011; Skoric and Zhu, 2016). Therefore, we assume that participating in social networks is linked to political interest and a higher probability of participating in a survey about political attitudes and behavior.

Further, the German online panel provider our respondents are recruited from advertises participating in its online panel on, among others, social media platforms[11]. Thus, we assume that social media usage is linked to a higher probability of being invited to become an online panel member.

Second, we also assume a positive association between online participation in political activities and processes and survey participation, as well as searching for Information in Web-based newspapers. Participating online in political activities and processes, as well as seeking Information in Web-based newspapers, can be understood as a particular pattern of political behavior, and one explanatory factor of political behavior, also in the online realm, is interest in politics (Gil de Zúñiga et al., 2014; Saldaña et al., 2015; Boulianne, 2016; Wolfsfeld et al., 2016). Political interest, again, represents a political survey interest in the survey topic. As it is well established that topic interest is a significant factor that affects survey participation, we suggest that the probability of participating in a survey on politics and elections is positively related to a person's interest in politics (Hansen, 1980; Furse and Stewart, 1984; Ajzen, 1985; Dillman, 1991; Poon et al., 1999; Groves et al., 1992, 2000; Voogt and Saris, 2003).

Regarding political attitudes and behavior, we also assume strong associations with different Internet usage patterns. Altogether, we can identify three theoretical arguments for why and how different patterns of Internet usage influence political participation (for a summary, see: Boulianne, 2019).

First, the Internet offers individuals information on political issues and current or upcoming (political) events. Thus, participating in social media, as well as seeking

[11] for a detailed information on the recruitment strategy of the panel provider, see: https://www.respondi.com/wp-content/uploads/2021/03/respondi_panel-quality-2021.pdf?cf_id=3397

information in Web-based newspapers, may raise awareness for political issues or events and may also positively impact a user's likelihood to engage in politics in the offline realm (Gil de Zúñiga et al., 2014; Saldaña et al., 2015; Boulianne, 2016; Wolfsfeld et al., 2016).

Second, actively involved individuals may be more likely to engage in politics, as being part of a network increases the likelihood of being asked to participate in politics (Park et al., 2009; Skoric et al., 2009; Gil de Zúñiga and Valenzuela, 2011; Skoric and Zhu, 2016). Social media platforms are tools for building both informal and formal personal networks, we assume a positive association between political and social network participation.

Third, the Internet offers the opportunity to express personal opinions and discussing political issues. By this, social media platforms as well as online participation in political activities and processes may enhance political knowledge (Saldaña et al., 2015) and increase the chance that people share their opinions on politics with others the offline realm (Kushin and Yamamoto, 2010; Jung et al., 2011; Gil de Zúñiga et al., 2014; Lu et al., 2016).

Building on these arguments, we expect that Internet usage patterns should positively affect political participation. However, some authors (Gil de Zúñiga et al., 2010; Jung et al., 2011; Chan and Guo, 2013; Chan, 2016) have argued those ways in which people make use of Internet-based services such as social media, online participation in political activities and processes, or searching for information in Web-based newspapers only indirectly influence online political participation. According to this position, the association is mediated by increased political efficacy. We, therefore, included measures on political efficacy in our data analysis (Gil de Zúñiga et al., 2010; Jung et al., 2011; Chan and Guo, 2013; Chan, 2016).

4.4.2.2 STEP 2. Available registry data for Internet usage patterns

We use data from the German Microcensus 2021 to calculate the enhanced, theory-based weights for different Internet usage patterns, As our dataset consists of respondent from Lower Saxony, we use benchmark data provided by the statistical office of Lower Saxony. Since 2021 the German Microcensus has included measures of private Internet usage as part of the use of the information technologies module. We used data on social media usage, reading online newspapers or magazines, and online political participation of adults with German citizenship aged 18 and older.

4.4.2.3 STEP 3. Measures of Internet usage included in our survey program

To investigate how and to what extent survey respondents use the Internet for private purposes, we included the following question in our questionnaire: “For what private purposes did you use the Internet in the last three months?” We further added the following note to clarify what we mean with Internet usage for private purposes: “Please, also think about app usage with any Internet-enabled devices (e.g., desktop PC, laptop, tablet, smartphone, game console, e-book reader)”. The same note was also included in the questionnaire of the microcensus. We asked whether respondents used the Internet for (a) participating in social networks (e.g., creating a user profile, posting messages or other posts on Facebook, Twitter, Instagram, Snapchat, or other social networking sites); (b) reading online news, online newspapers, or online magazines; and (c) socio-political participation (writing opinions on political or social issues on Websites, participating in deliberations, or voting on political or social, or community issues online). The variables were dummy coded with 1, stating that respondents used the Internet for the mentioned purpose and 0, stating that respondents did not use the Internet for the mentioned purpose.

4.4.2.4 STEP 4. Data collection

We used data from the GLES Panel, a multi-wave panel initiated in 2016. Respondents were recruited in the online panel Respondi.AG, using quotas on gender, age, and education (GLES, 2023). The initial sample comprised 15,802 respondents. Our study used data from the 23rd wave, conducted between October 12 and 25, 2022, after the state election in Lower Saxony. As benchmark data was only available for Lower Saxony, we ran our analyses on a subsample of 1,221 complete interviews from respondents living in this state.

4.4.2.5 STEP 5. Correlations of Internet usage patterns with survey participation and political attitudes and behavior in our survey data

To empirically check whether our assumptions holds that Internet usage patterns correlate with (1) survey participation and (2) political attitudes and behavior, we calculate pearson correlation coefficients, using the R command cor.

For the correlation between different Internet usage patterns and survey participation, we cannot directly measure whether there is an association. However, we measure this assumed correlation indirectly. As stated above, we assume that different Internet usage patterns are associated with political interest, which is, in turn, related to the probability of participating in a survey on politics and elections. We measure the respondents' interest in politics to investigate whether this assumption holds.

Regarding variables of interest, we calculate correlation coefficients for each category of Internet usage and political participation. To measure political participation, we use data on voting behavior, namely whether respondents voted in the previous federal or state election. In addition, we use voting intention as a more general measure of voting behavior. Further, as some authors state a more indirect effect of social media usage and political participation mediated by internal political efficacy, we also compute correlation coefficients between each category of Internet usage and internal political efficacy[12].

Table 4.5: Correlations of Internet usage with survey participation and variables of interest

Internet usage	Political attitudes/behavior			
	political interest	internal efficacy	voting intention	voted in previous election
Social media usage	0.03	0.00	-0.01	0.00
Online news consumption	0.24	0.25	0.15	0.15
sharing opinion	0.28	0.28	0.13	0.11

Note: Coefficient = Pearson correlation coefficient.

Table 4.5 presents the correlations between Internet usage patterns and measures of survey participation as well as political attitudes and behavior. We find moderate positive correlations between political interest and sharing one's opinion about politics online and online news consumption. These findings indicate that higher Internet usage is related to a more substantial political interest. Looking at the

[12]For a detailed description of the operationalization of these variables, see appendix (C.3).

correlation between social media usage and political attitudes and behavior, we do not find a notable association. This finding can be attributed to the assumption that many individuals do not use social media for political purposes. Some researchers argue that using social media alone does not imply individuals getting in contact with political content (Vissers and Stolle, 2014; Skoric and Zhu, 2016; Boulianne and Theocharis, 2020) and, therefore, does not provide a sufficient condition to encourage political participation. What matters more is whether individuals actively use these platforms for political communication or to inform themselves about politics (Vissers and Stolle, 2014; Skoric and Zhu, 2016; Boulianne and Theocharis, 2020). Looking at sharing one's opinion about politics online and online news consumption - two variables that directly target using the Internet for political purposes- the picture is somewhat different. Here, we see positive medium strong correlations with internal political efficacy. However, the correlations with political behavior (voting intention and voted in the previous election) are weak.

Altogether, these findings show that the assumed correlations between different Internet usage patterns and substantial variables of interest, as well as survey participation, are weaker than we expected. As in study I, we would therefore - following our six-step, recommend not to use the theoretically selected variables to calculate adjustment weights. However, to present an empirical application of the six-step strategy, we we will calculate the weights in the next step.

4.4.2.6 STEP 6. Calculation of Internet usage post-Survey adjustments

As in study I, we compute post-stratification survey weights (GLES, 2023). In contrast, however, we have to rely on the Information provided by the statistical office of Lower Saxony, which includes gender and education but, unfortunately, not age. Thus, the standard socio-demographic weight does not adjust for age.

As before, we use the absolute and the absolute relative bias as measures of bias in the estimates across different survey weights. Figure 4.3 shows the estimates for different variables with 95 percent confidence intervals for (1) the unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights. As population parameters, we use benchmark data on voting behavior (turnout, party vote, depicted as horizontal lines).

As in study I, we do not find substantive differences in the estimates across adjusted and unadjusted survey data. Regarding the estimates for voter turnout and the proportion of voters of the green party, the results align with the hypothesis that the bias is lowest if the data is adjusted with the enhanced, theory-based weight. However, the differences are minor. Further, we see that adjusting survey data with the enhanced, theory-based weights increases the bias compared to the estimates derived from the unadjusted survey data for some variables. This is true for the proportion of voters for the left-wing party (DIE LINKE) voters, the social democratic party (SPD), and splinter parties (other party). Comparing estimates derived from the (2) survey data adjusted with standard socio-demographic weights and (3) survey data adjusted with the enhanced, theory-based weights, the picture is similar: The bias is lower for estimates of the voters of the left-wing party (DIE LINKE), splinter parties (other party), and voter turnout if data is adjusted with the enhanced, theory-based weights. However, for the proportion of voters of the liberal party (FDP) and the social democratic party (SPD), the standard socio-demographic weights result in more accurate estimates, although the differences are again negligible.

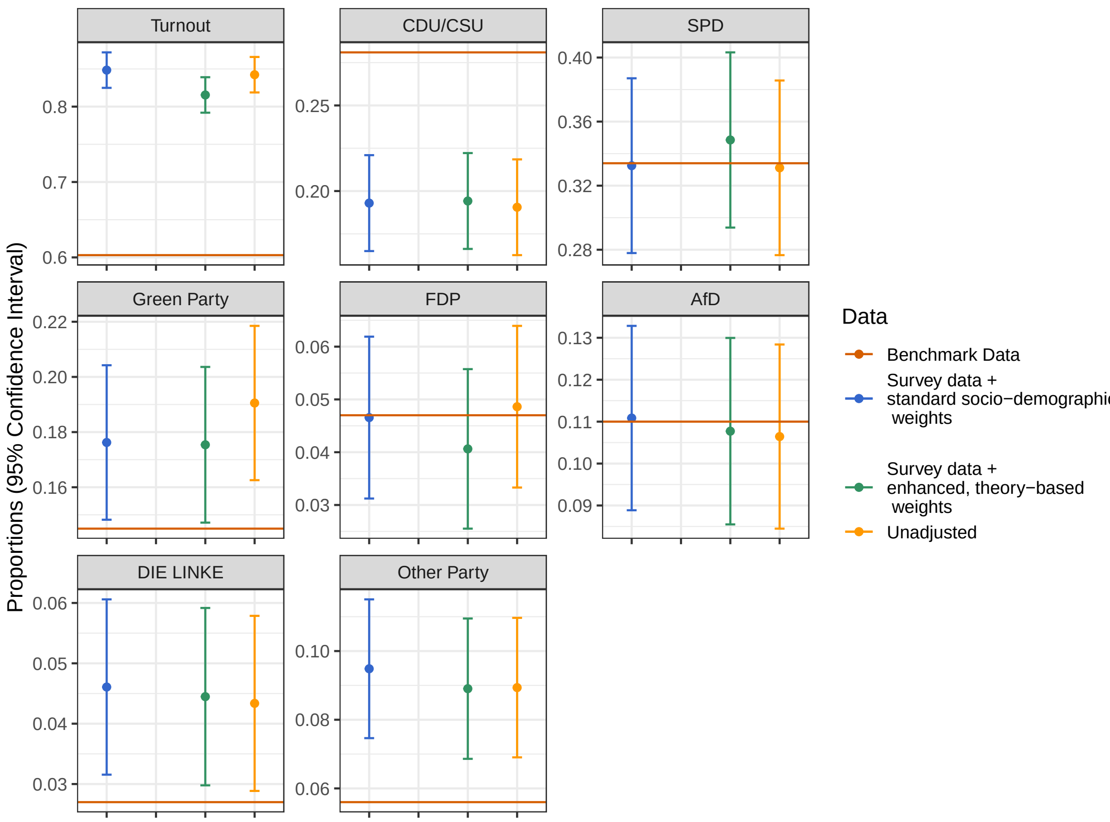

Figure 4.3: Point estimates in survey data with different weights applied

Table 4.6 reports the differences in the (absolute) relative bias for the (1) unadjusted survey data, (2) survey data adjusted with standard socio-demographic weights, and (3) survey data adjusted with the enhanced, theory-based weights. As in study I, none of the adjustment weights effectively reduces the absolute relative bias. Although the results support the hypothesis that the bias is lowest for the survey data adjusted with the enhanced, theory-based weights, the relative bias reduction is minimal (0.3 percentage points). As in study I, one possible explanation for this finding is weak correlations between the patterns of Internet usage and the variables of interest, as well as survey participation. This observation is inconsistent with our theoretical expectations and contradicts findings from previous studies.

Moreover, adjusting the data with standard socio-demographic weights amplifies the absolute relative bias in estimates of voting behavior. Like in study I, we take this finding as a warning that misspecification in adjustment models can exacerbate biases in estimates and lead to erroneous conclusions.

Table 4.6: Bias across the different survey weights.

Variable	Unadjusted survey data	Survey data + standard socio-demographic weights	Survey data + enhanced,theory-based weights
Turnout	0.239	0.245	0.212
SPD	-0.003	-0.002	0.015
CDU	-0.090	-0.088	-0.087
Green Party	0.046	0.031	0.030
FDP	0.002	0.000	-0.006
AfD	-0.004	0.001	-0.002
DIE LINKE	0.016	0.019	0.017
Other Party	0.033	0.039	0.033
Total			
Absolute relative bias	0.030	0.031	0.027

4.5 Discussion

This contribution aimed to introduce an approach that rests on a theoretically and empirically informed selection of variables to be used in adjusting data from nonprobability online surveys. By introducing a six-step strategy, we suggested that considering variables that are associated with both survey participation and substantive variables of interest increase the effectiveness of post-survey adjustments in reducing selection biases. However, the results from two studies, in that we implemented the proposed six-step strategy, consistently showed that adjusting the data from nonprobability online surveys with enhanced, theory-based weights did not substantively reduce selection biases in estimates of voting behavior compared to the unadjusted data and data adjusted with standard socio-demographic weights. While using COVID-19 vaccination status as an adjustment ultimately failed to reduce biases in the estimates in study I, adjusting the nonprobability survey data on Internet usage patterns resulted in a minimal reduction of the absolute relative bias. A reasonable explanation for the poor performance of the enhanced, theory-based weights were the weak correlations of the selected adjustment variables with measures of (1) survey participation, and (2) political attitudes and behavior.

One rationale might explain the weak correlation of the selected adjustment variables with survey participation that we observed in our data: To choose adjustment variables which are substially related to survey response, we referred to traditional theoretical frameworks of survey participation. In thesen frameworks, Factors such as as adhering to norms and interest in the survey´s topic play a pivotal role in explaining the the decision to participate in a survey. Particularly, methodological studies on participation in surveys on politics and elections often stress the importance of peoples' general interest in politics (e.g. Sciarini and Goldberg, 2016). In study I, we used indicators of norm adherence (voting as norm, complying with the instructions of an attention check) and interest in the survey's topic (interest in politics) as proxies for the willingness to participate in the survey. However, we take the finding of weak correlations with COVID-19 vaccination status as an indication that other factors, such as, for example, financial motives, may be more relevant in explaining survey participation by members of nonprobability opt-in online panels.

To date, only a few studies have sought to understand what brings people to join opt-in online panels and how they decide whether or not to participate in specific surveys (Brüggen et al., 2011; Keusch et al., 2014). The findings from these studies suggest that factors influencing the decision to participate in a survey substantially differ between respondents recruited by random selection methods and respondents recruited from self-selective online panels. For the latter, enjoying answering surveys (Brüggen et al., 2011; Keusch et al., 2014) and monetary incentives (Brüggen et al., 2011; Keusch et al., 2014) are very important motives that positively influence the decision to participate in surveys. Moreover, Keusch et al. (2014) found that in online panels, a survey's topic positively affected response by fresh online panel members but did not influence paticipation decisions by experienced panel members (Keusch et al., 2014). Thus, we conclude that traditional theoretical explanations of survey participation may not be transferable to nonprobability surveys with respondents from online panels without restrictions.

The goal of enhancing model-based adjustments for nonprobability surveys, therefore, necessitates the further development of theoretical frameworks for explaining the multifaceted selection mechanisms that affect participation in these kinds of surveys. Unfortunately, to the best of our knowledge, a comprehensive theoretical framework for nonprobability survey participation is still a desideratum in the area of survey research. Such a framework would have to clearly disentangle the multiple steps involved in nonprobability survey participation: (1) It needs to provide rationales for why people join opt-in online panels, and (2) it needs to explain on what basis panel members decide for participating in a given survey or not. Relatedly, further research should be undertaken to enhance our understanding of recruitment for opt-in online panels and participation in nonprobability surveys, and how these processes induce selection biases in survey outcomes. Ultimately, this research should foster the theory-guided enhancement of post-survey adjustments for nonprobability surveys.

Notwithstanding the limitation that applying the six step strategy in empirical case studies did not succeed in reducing selection biases in estimates obtained from two nonprobability surveys, two findings should be acknowledged. First, adjusting the nonprobability survey data with basic socio-demographic weights did not succeed in decreasing biases in estimates of voting behavior. Even worse, applying such standard weights, as they are often distributed by survey agencies, amplified biases in the estimates of voting behavior compared to estimates from the unadjusted survey data. We take this finding as another cautionary note to survey researchers that misspecification in adjustment models can exacerbate biases in estimates, which may ultimately lead to erroneous conclusions from survey data analyses.

Eventually, the results of our study demonstrate the key role of extending our understanding of selection processes in participation decisions for advancing our knowledge on how to improve adjustments for nonprobability surveys and, as a consequence, to decrease biases in estimates from these data.

References

Agranov, Marina, Matt Elliott, and Pietro Ortoleva. 2021. "The Importance of Social Norms Against Strategic Effects: The Case of Covid-19 Vaccine Uptake." *Economics Letters* 206 (September): 109979. https://doi.org/10.1016/j.econlet.2021.109979.

Ajzen, Icek. 1985. "From Intentions to Actions: A Theory of Planned Behavior." In *Action Control*, edited by Julius Kuhl and Jürgen Beckmann, 11–39. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69746-3_2.

Albaum, Gerald, Felicitas Evangelista, and Nila Medina. 1998. "Role of Response Behavior Theory in Survey Research." *Journal of Business Research* 42 (2): 115–25. https://doi.org/10.1016/S0148-2963(97)00108-2.

Albaum, Gerald, and Scott M. Smith. 2012. "Why People Agree to Participate in Surveys." In *Handbook of Survey Methodology for the Social Sciences*, edited by Lior Gideon, 179–93. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3876-2_11.

Allen, C. T., C. D. Schewe, and G. Wijk. 1980. "More on Self-Perception's Foot Techniques in the Pre-Call/Mail Survey Setting." *Journal of Marketing Research* 1980 (17): 498–502.

Baeza-Rivera, María José, Camila Salazar-Fernández, Leslie Araneda-Leal, and Diego Manríquez-Robles. 2021. "To Get Vaccinated or Not? Social Psychological Factors Associated with Vaccination Intent for COVID-19." *Journal of Pacific Rim Psychology* 15 (January): 183449092110517. https://doi.org/10.1177/18344909211051799.

Becker, Howard S. 1960. "Notes on the Concept of Commitment." *American Journal* *of Sociology* 66 (1): 32–40. https://doi.org/10.1086/222820.

Bem, Daryl J. 1972. "Self-Perception Theory." In *Advances in Experimental Social Psychology*, 6:1–62. Elsevier. https://doi.org/10.1016/S0065-2601(08)60024-6.

Bergmann, Michael. 2011. "Ipfweight: Stata Module to Create Adjustment Weights for Surveys."

Bethlehem, Jelke. 2016. "Solving the Nonresponse Problem With Sample Matching?" *Social Science Computer Review* 34 (1): 59–77. https://doi.org/10.1177/0894439315573926.

Biemer, P. P. 2010. "Total Survey Error: Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74 (5): 817–48. https://doi.org/10.1093/poq/nfq058.

Blau, Peter Michael. 1964. *Exchange and Power in Social Life*. New York, NY: Wiley.

Bosnjak, Michael, Tracy L. Tuten, and Werner W. Wittmann. 2005. "Unit (Non)Response in Web-based Access Panel Surveys: An Extended Planned-Behavior Approach." *Psychology and Marketing* 22 (6): 489–505. https://doi.org/10.1002/mar.20070.

Boulianne, Shelley. 2016. "Online News, Civic Awareness, and Engagement in Civic and Political Life." *New Media & Society* 18 (9): 1840–56. https://doi.org/10.1177/1461444815616222.

———. 2019. "Revolution in the Making? Social Media Effects Across the Globe." *Information, Communication & Society* 22 (1): 39–54. https://doi.org/10.1080/1369118X.2017.1353641.

Boulianne, Shelley, and Yannis Theocharis. 2020. "Young People, Digital Media, and Engagement: A Meta-Analysis of Research." *Social Science Computer Review* 38 (2): 111–27. https://doi.org/10.1177/0894439318814190.

Brüggen, Elisabeth, Martin Wetzels, Ko De Ruyter, and Niels Schillewaert. 2011.

"Individual Differences in Motivation to Participate in Online Panels: The Effect on Reponse Rate and Reponse Quality Perceptions." *International Journal of Market Research* 53 (3): 369–90. https://doi.org/10.2501/IJMR-53-3-369-390.

Chan, Michael. 2016. "Social Network Sites and Political Engagement: Exploring the Impact of Facebook Connections and Uses on Political Protest and Participation." *Mass Communication and Society* 19 (4): 430–51. https://doi.org/10.1080/15205436.2016.1161803.

Chan, Michael, and Jing Guo. 2013. "The Role of Political Efficacy on the Relationship Between Facebook Use and Participatory Behaviors: A Comparative Study of Young American and Chinese Adults." *Cyberpsychology, Behavior, and Social Networking* 16 (6): 460–63. https://doi.org/10.1089/cyber.2012.0468.

Cialdini, Robert B., and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55 (1): 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015.

DeCamp, Whitney, and Matthew J. Manierre. 2016. "'Money Will Solve the Problem': Testing the Effectiveness of Conditional Incentives for Online Surveys." *Survey Practice* 9 (1): 1–9. https://doi.org/10.29115/SP-2016-0003.

Deming, W. Edwards, and Frederick F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." *The Annals of Mathematical Statistics* 11 (4): 427–44. https://doi.org/10.1214/aoms/1177731829.

Dillman, Don A. 1991. "The Design and Administration of Mail Surveys." *Annual Review of Sociology* 17 (1): 225–49. https://doi.org/10.1146/annurev.so.17.080191.001301.

———. 2022. "Fifty years of survey innovation." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 154 (1): 9–38. https://doi.org/10.

1177/07591063221088317.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th edition. Hoboken: Wiley.

Edwards, Ben, Nicholas Biddle, Matthew Gray, and Kate Sollis. 2021. "COVID-19 Vaccine Hesitancy and Resistance: Correlates in a Nationally Representative Longitudinal Survey of the Australian Population." Edited by Francesco Di Gennaro. *PLOS ONE* 16 (3): e0248892. https://doi.org/10.1371/journal.pone.0248892.

Elliott, Michael R., and Richard Valliant. 2017. "Inference for Nonprobability Samples." *Statistical Science* 32 (2): 249–64. https://doi.org/10.1214/16-STS598.

El-Mohandes, Ayman, Trenton M. White, Katarzyna Wyka, Lauren Rauh, Kenneth Rabin, Spencer H. Kimball, Scott C. Ratzan, and Jeffrey V. Lazarus. 2021. "COVID-19 Vaccine Acceptance Among Adults in Four Major US Metropolitan Areas and Nationwide." *Scientific Reports* 11 (1): 21844. https://doi.org/10.1038/s41598-021-00794-6.

Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40 (1): 31–53. https://doi.org/10.1146/annurev-soc-071913-043455.

Evangelista, Felicitas, Patrick Poon, and Gerald Albaum. 2012. "Using Response Behaviour Theory to Solicit Survey Participation in Consumer Research: An Empirical Study." *Journal of Marketing Management* 28 (9-10): 1174–89. https://doi.org/10.1080/0267257X.2011.619148.

Ferri-García, Ramón, Luis Castro-Martín, and María del Mar Rueda. 2021. "Evaluating Machine Learning Methods for Estimation in Online Surveys with Superpopulation Modeling." *Mathematics and Computers in Simulation* 186 (August): 19–28. https://doi.org/10.1016/j.matcom.2020.03.005.

Festinger, Leon. 2001. *A Theory of Cognitive Dissonance.* Reissued by Stanford Univ. Press in 1962, renewed 1985 by author, [Nachdr.]. Stanford, Calif: Stanford Univ. Press.

Furse, David H., and David W. Stewart. 1982. "Monetary Incentives Versus Promised Contribution to Charity: New Evidence on Mail Survey Response." *Journal of Marketing Research* 19 (3): 375–80. https://doi.org/10.1177/002224378201900311.

———. 1984. "Manipulating Dissonance to Improve Mail Survey Response." *Psychology and Marketing* 1 (2): 79–94. https://doi.org/10.1002/mar.4220010208.

Geuzinge, Linda, Johan van Rooijen, and Bart F. Bakker. 2000. "The Use of Administrative Registers to Reduce Non-Response Bias in Household Surveys." Edited by Statistics Netherlands. *Integrating Administrative Registers and Household Surveys* Special Issue (15): 32–39.

Ghitza, Yair, and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups." *American Journal of Political Science* 57 (3): 762–76. https://doi.org/10.1111/ajps.12004.

Gil de Zúñiga, Homero, Logan Molyneux, and Pei Zheng. 2014. "Social Media, Political Expression, and Political Participation: Panel Analysis of Lagged and Concurrent Relationships." *Journal of Communication* 64 (4): 612–34. https://doi.org/10.1111/jcom.12103.

Gil de Zúñiga, Homero, and Sebastián Valenzuela. 2011. "The Mediating Path to a Stronger Citizenship: Online and Offline Networks, Weak Ties, and Civic Engagement." *Communication Research* 38 (3): 397–421. https://doi.org/10.1177/0093650210384984.

Gil de Zúñiga, Homero, Aaron Veenstra, Emily Vraga, and Dhavan Shah. 2010. "Digital Democracy: Reimagining Pathways to Political Participation." *Journal* *of Information Technology & Politics* 7 (1): 36–51. https://doi.org/10.1080/19331680903316742.

GLES. 2022a. "GLES Cross-Section 2021, Pre-Election GLES Querschnitt 2021, Vorwahl." GESIS Data Archive. https://doi.org/10.4232/1.13860.

GLES. 2022b. "GLES Panel 2021, Control Group III (to Panel Wave 20, Sample A)GLES Panel 2021, Kontrollquerschnitt III (zu Welle 20, Sample A)." GESIS. https://doi.org/10.4232/1.14053.

———. 2023. "GLES Panel 2022, Wave 23 GLES Panel 2022, Welle 23." GESIS. https://doi.org/10.4232/1.14064.

Greenberg, Pierce, and Don Dillman. 2023. "Mail Communications and Survey Response: A Test of Social Exchange Versus Pre-Suasion Theory for Improving Response Rates and Data Quality." *Journal of Survey Statistics and Methodology* 11 (1): 1–22. https://doi.org/10.1093/jssam/smab020.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70 (5): 646–75. https://doi.org/10.1093/poq/nfl033.

Groves, Robert M., Robert B. Cialdini, and Mick P. Couper. 1992. "Understanding The Decision to Participate in a Survey." *Public Opinion Quarterly* 56 (4): 475. https://doi.org/10.1086/269338.

Groves, Robert M., and Mick Couper. 1998. *Nonresponse in Household Interview Surveys.* Wiley Series in Probability and Statistics. New York: Wiley.

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation." *Public Opinion Quarterly* 64 (3): 299–308. https://doi.org/10.1086/317990.

Hackler, James C., and Patricia Bourgette. 1973. "Dollars, Dissonance, and Survey Returns." *Public Opinion Quarterly* 37 (2): 276. https://doi.org/10.1086/

268085.

Hansen, Robert A. 1980. "A Self-Perception Interpretation of the Effect of Monetary and Nonmonetary Incentives on Mail Survey Respondent Behavior." *Journal of Marketing Research* 17 (1): 77–83. https://doi.org/10.1177/002224378001700110.

Heerwegh, Dirk, and Geert Loosveldt. 2009. "Explaining the Intention to Participate in a Web Survey: A Test of the Theory of Planned Behaviour." *International Journal of Social Research Methodology* 12 (3): 181–95. https://doi.org/10.1080/13645570701804235.

Hox, Joop, Edith de Leeuw, and Harrie Vorst. 1995. "Survey Participation as Reasoned Action; a Behavioral Paradigm for Survey Nonresponse?" *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 48 (1): 52–67. https://doi.org/10.1177/075910639504800109.

Iachan, Ronaldo, Lewis Berman, Tonja M. Kyle, Kelly J. Martin, Yangyang Deng, Davia N. Moyse, Deirdre Middleton, and Audie A. Atienza. 2019. "Weighting Nonprobability and Probability Sample Surveys in Describing Cancer Catchment Areas." *Cancer Epidemiology, Biomarkers & Prevention* 28 (3): 471–77. https://doi.org/10.1158/1055-9965.EPI-18-0797.

Jaffe, Anna E., Scott Graupensperger, Jessica A. Blayney, Jennifer C. Duckworth, and Cynthia A. Stappenbeck. 2022. "The Role of Perceived Social Norms in College Student Vaccine Hesitancy: Implications for COVID-19 Prevention Strategies." *Vaccine* 40 (12): 1888–95. https://doi.org/10.1016/j.vaccine.2022.01.038.

Jung, Nakwon, Yonghwan Kim, and Homero Gil de Zúñiga. 2011. "The Mediating Role of Knowledge and Efficacy in the Effects of Communication on Political Participation." *Mass Communication and Society* 14 (4): 407–30. https://doi.

org/10.1080/15205436.2010.496135.

Kennedy, Lauren, and Andrew Gelman. 2021. "Know Your Population and Know Your Model: Using Model-Based Regression and Poststratification to Generalize Findings Beyond the Observed Sample." *Psychological Methods* 26 (5): 547–58. https://doi.org/10.1037/met0000362.

Keusch, Florian. 2015. "Why Do People Participate in Web Surveys? Applying Survey Participation Theory to Internet Survey Data Collection." *Management Review Quarterly* 65 (3): 183–216. https://doi.org/10.1007/s11301-014-0111-y.

Keusch, Florian, Bernad Batinic, and Wolfgang Mayerhofer. 2014. "Motives for Joining Nonprobability Online Panels and Their Association with Survey Participation Behavior." In *Online Panel Research*, edited by Mario Callegaro, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas, 171–91. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118763520.ch8.

Killgore, William D. S., Sara A. Cloonan, Emily C. Taylor, and Natalie S. Dailey. 2021. "The COVID-19 Vaccine Is Here—Now Who Is Willing to Get It?" *Vaccines* 9 (4): 339. https://doi.org/10.3390/vaccines9040339.

Kim, Jae Kwang, Seho Park, Yilin Chen, and Changbao Wu. 2021. "Combining Non-Probability and Probability Survey Samples Through Mass Imputation." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184 (3): 941–63. https://doi.org/10.1111/rssa.12696.

Kreuter, Frauke, Michael Lemay, and Carolina Casas-Cordero. 2007. "Using Proxy Measures of Survey Outcomes in Post-Survey Adjustments: Examples from the European Social Survey (ESS)." In *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass*, 3142–49.

Kreuter, Frauke, and Kristen Olson. 2011. "Multiple Auxiliary Variables in Nonresponse Adjustment.” *Sociological Methods & Research* 40 (2): 311–32. https://doi.org/10.1177/0049124111400042.

Kushin, Matthew James, and Masahiro Yamamoto. 2010. “Did Social Media Really Matter? College Students’ Use of Online Media and Political Decision Making in the 2008 Election.” *Mass Communication and Society* 13 (5): 608–30. https://doi.org/10.1080/15205436.2010.516863.

Latkin, Carl, Lauren A. Dayton, Grace Yi, Arianna Konstantopoulos, Ju Park, Catherine Maulsby, and Xiangrong Kong. 2021. “COVID-19 Vaccine Intentions in the United States, a Social-Ecological Framework.” *Vaccine* 39 (16): 2288–94. https://doi.org/10.1016/j.vaccine.2021.02.058.

Lindholt, Marie Fly, Frederik Jørgensen, Alexander Bor, and Michael Bang Petersen. 2021. “Public Acceptance of COVID-19 Vaccines: Cross-National Evidence on Levels and Individual-Level Predictors Using Observational Data.” *BMJ Open* 11 (6): e048172. https://doi.org/10.1136/bmjopen-2020-048172.

Little, Roderick J A, Brady T West, Philip S Boonstra, and Jingwei Hu. 2020. “Measures of the Degree of Departure from Ignorable Sample Selection.” *Journal of Survey Statistics and Methodology* 8 (5): 932–64. https://doi.org/10.1093/jssam/smz023.

Little, Roderick J. A. 1986. “Survey Nonresponse Adjustments for Estimates of Means.” *International Statistical Review / Revue Internationale de Statistique* 54 (2): 139. https://doi.org/10.2307/1403140.

Little, Roderick, and Sonya Vartivarian. 2005. “Does Weighting for Nonresponse Increase the Variance of Survey Means?” *Survey Methodology* 2 (31).

Loosveldt, Geert, and Nathalie Sonck. 2008. “An Evaluation of the Weighting Procedures for an Online Access Panel Survey.” *Survey Research Methods* 2 (2): 93–105. https://doi.org/10.18148/srm/2008.v2i2.82.

Lu, Yanqin, Kyle A. Heatherly, and Jae Kook Lee. 2016. "Cross-Cutting Exposure on Social Networking Sites: The Effects of SNS Discussion Disagreement on Political Participation." *Computers in Human Behavior* 59 (June): 74–81. https://doi.org/10.1016/j.chb.2016.01.030.

MacInnis, Bo, Jon A. Krosnick, Annabell S. Ho, and Mu-Jung Cho. 2018. "The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension." *Public Opinion Quarterly* 82 (4): 707–44. https://doi.org/10.1093/poq/nfy038.

Maitland, Aaron, Carolina Casas Cordero, and Frauke Kreuter. 2008. "An Exploration into the Use of Paradata for Nonresponse Adjustment in a Health Survey." In *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods*, 2250–55.

Marcus, Bernd, Michael Bosnjak, Steffen Lindner, Stanislav Pilischenko, and Astrid Schütz. 2007. "Compensating for Low Topic Interest and Long Surveys: A Field Experiment on Nonresponse in Web Surveys." *Social Science Computer Review* 25 (3): 372–83. https://doi.org/10.1177/0894439307297606.

Marsden, Peter V., and James D. Wright. 2010. *Handbook of Survey Research*. 2nd ed. Bingley, UK: Emerald.

Mercer, Andrew W., Frauke Kreuter, Scott Keeter, and Elizabeth A. Stuart. 2017. "Theory and Practice in Nonprobability Surveys." *Public Opinion Quarterly* 81 (S1): 250–71. https://doi.org/10.1093/poq/nfw060.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12 (4): 375–85. https://doi.org/10.1093/pan/mph024.

Park, Namsu, Kerk F. Kee, and Sebastián Valenzuela. 2009. "Being Immersed in Social Networking Environment: Facebook Groups, Uses and Gratifications, and Social Outcomes." *CyberPsychology & Behavior* 12 (6): 729–33. https://doi.org/10.1089/cpb.2009.0003.

Pasek, Josh. 2016. "When Will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence." *International Journal of Public Opinion Research* 28 (2): 269–91. https://doi.org/10.1093/ijpor/edv016.

Peytchev, Andy, Stanley Presser, and Mengmeng Zhang. 2018. "Improving Traditional Nonresponse Bias Adjustments: Combining Statistical Properties with Social Theory." *Journal of Survey Statistics and Methodology* 6 (4): 491–515. https://doi.org/10.1093/jssam/smx035.

Poon, Patrick, Gerald Albaum, and Felicitas Evangelista. 1999. "An Empirical Test of Alternative Theories of Survey Response Behaviour." *Market Research Society. Journal.* 41 (2): 1–20. https://doi.org/10.1177/147078539904100201.

Rabb, Nathaniel, Jake Bowers, David Glick, Kevin H. Wilson, and David Yokum. 2022. "The Influence of Social Norms Varies with 'Others' Groups: Evidence from COVID-19 Vaccination Intentions." *Proceedings of the National Academy of Sciences* 119 (29): e2118770119. https://doi.org/10.1073/pnas.2118770119.

Roberts, Hannah A., D. Angus Clark, Claire Kalina, Carter Sherman, Sarah Brislin, Mary M. Heitzeg, and Brian M. Hicks. 2022. "To Vax or Not to Vax: Predictors of Anti-Vax Attitudes and COVID-19 Vaccine Hesitancy Prior to Widespread Vaccine Availability." Edited by Anat Gesser-Edelsburg. *PLOS ONE* 17 (2): e0264019. https://doi.org/10.1371/journal.pone.0264019.

Rohrer, Julia M. 2018. "Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data." *Advances in Methods and Practices in Psychological Science* 1 (1): 27–42. https://doi.org/10.1177/2515245917745629.

Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63 (3): 581–92. https://doi.org/10.1093/biomet/63.3.581.

Ryoo, Yuhosua, and WooJin Kim. 2021. "Using Descriptive and Injunctive Norms to Encourage COVID-19 Social Distancing and Vaccinations." *Health Communication*, September, 1–10. https://doi.org/10.1080/10410236.2021.1973702.

Sabahelzain, Majdi M, Kenneth Hartigan-Go, and Heidi J Larson. 2021. "The Politics of Covid-19 Vaccine Confidence." *Current Opinion in Immunology* 71 (August): 92–96. https://doi.org/10.1016/j.coi.2021.06.007.

Saldaña, Magdalena, Shannon C. McGregor, and Homero Gil De Zúñiga. 2015. "Social Media as a Public Space for Politics: Cross-National Comparison of News Consumption and Participatory Behaviors in the United States and the United Kingdom." *International Journal of Communication* 9: 3304–26.

Schernhammer, Eva, Jakob Weitzer, Manfred D Laubichler, Brenda M Birmann, Martin Bertau, Lukas Zenk, Guido Caniglia, Carlo C Jäger, and Gerald Steiner. 2022. "Correlates of COVID-19 Vaccine Hesitancy in Austria: Trust and the Government." *Journal of Public Health* 44 (1): e106–16. https://doi.org/10.1093/pubmed/fdab122.

Schonlau, Matthias, Arthur H. O. van Soest, and Arie Kapteyn. 2007. "Are 'Webographic' or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring?" *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1006108.

Schonlau, Matthias, Arthur van Soest, Arie Kapteyn, and Mick Couper. 2009. "Selection Bias in Web Surveys and the Use of Propensity Scores." *Sociological Methods & Research* 37 (3): 291–318. https://doi.org/10.1177/0049124108327128.

Schouten, Barry. 2007. "A Selection Strategy for Weighting Variables Under a Not-Missing-at-Random Assumption." *Journal of Official Statistics* 23 (1): 51.

Sciarini, Pascal, and Andreas C. Goldberg. 2015. "Lost on the Way: Nonresponse and Its Influence on Turnout Bias in Postelection Surveys." *International Journal of Public Opinion Research*, December, edv049. https://doi.org/10.1093/ijpor/edv049.

———. 2016. "Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting." *Journal of Survey Statistics and Methodology* 4 (1): 110–37. https://doi.org/10.1093/jssam/smv039.

Serani, Danilo. 2022. "The Covid Pandemic Enters the Ballot Box: The Impact of Conspiracy Theories on Italians' Voting Behaviour During the COVID-19 Crisis." *Italian Political Science Review/Rivista Italiana Di Scienza Politica*, January, 1–18. https://doi.org/10.1017/ipo.2021.56.

Silber, Henning, Joss Roßmann, and Tobias Gummer. 2022. "The Issue of Noncompliance in Attention Check Questions: False Positives in Instructed Response Items." *Field Methods* 34 (4): 346–60. https://doi.org/10.1177/1525822X221115830.

Skoric, Marko M., Deborah Ying, and Ying Ng. 2009. "Bowling Online, Not Alone: Online Social Capital and Political Participation in Singapore." *Journal of Computer-Mediated Communication* 14 (2): 414–33. https://doi.org/10.1111/j.1083-6101.2009.01447.x.

Skoric, Marko M., and Qinfeng Zhu. 2016. "Social Media and Offline Political Participation: Uncovering the Paths From Digital to Physical." *International Journal of Public Opinion Research* 28 (3): 415–27. https://doi.org/10.1093/ijpor/edv027.

Trangucci, Rob, Imad Ali, Andrew Gelman, and Doug Rivers. 2018. "Voting Patterns in 2016: Exploration Using Multilevel Regression and Poststratification (MRP) on Pre-Election Polls." arXiv. https://doi.org/10.48550/ARXIV.1802.

00842.

Travis, Justin, Scott Harris, Tina Fadel, and Ginny Webb. 2021. "Identifying the Determinants of COVID-19 Preventative Behaviors and Vaccine Intentions Among South Carolina Residents." Edited by Anat Gesser-Edelsburg. *PLOS ONE* 16 (8): e0256178. https://doi.org/10.1371/journal.pone.0256178.

Valliant, Richard. 2020. "Comparing Alternatives for Estimation from Nonprobability Samples." *Journal of Survey Statistics and Methodology* 8 (2): 231–63. https://doi.org/10.1093/jssam/smz003.

Valliant, Richard, and Jill A. Dever. 2011. "Estimating Propensity Adjustments for Volunteer Web Surveys." *Sociological Methods & Research* 40 (1): 105–37. https://doi.org/10.1177/0049124110392533.

———. 2018. *Survey Weights: A Step-by-Step Guide to Calculation.* First edition. College Station, Texas: Stata Press, A Stata Press Publication, StataCorp LLC.

Valliant, Richard, Jill A. Dever, and Frauke Kreuter. 2018. *Practical Tools for Designing and Weighting Survey Samples.* Statistics for Social and Behavioral Sciences. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-93632-1.

Vissers, Sara, and Dietlind Stolle. 2014. "The Internet and New Modes of Political Participation: Online Versus Offline Participation." *Information, Communication & Society* 17 (8): 937–55. https://doi.org/10.1080/1369118X.2013.867356.

Voogt, Robert J. J., and Willem E. Saris. 2003. "To Participate or Not to Participate: The Link Between Survey Participation, Electoral Participation, and Political Interest." *Political Analysis* 11 (2): 164–79. https://doi.org/10.1093/pan/mpg003.

Walgrave, Stefaan, Ruud Wouters, and Pauline Ketelaars. 2016. "Response Problems in the Protest Survey Design: Evidence from Fifty-One Protest Events in Seven Countries*." *Mobilization: An International Quarterly* 21 (1): 83–104. https://doi.org/10.17813/1086/671X-21-1-83.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31 (3): 980–91. https://doi.org/10.1016/j.ijforecast.2014.06.001.

Wetzel, Martin, and Bettina Hünteler. 2022. "The Blind Spot: Studying the Association Between Survey Nonresponse and Adherence to COVID-19 Governmental Regulations in a Population-Based German Web-Survey." *Survey Research Methods*, December, 267–281 Pages. https://doi.org/10.18148/SRM/2022.V16I3.7901.

Wolf, Christof, Dominique Joye, Tom Smith, and Yang-chih Fu. 2016. *The SAGE Handbook of Survey Methodology.* 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd. https://doi.org/10.4135/9781473957893.

Wolfsfeld, Gadi, Moran Yarchi, and Tal Samuel-Azran. 2016. "Political Information Repertoires and Political Participation." *New Media & Society* 18 (9): 2096–2115. https://doi.org/10.1177/1461444815580413.

Wysocki, Anna C., Katherine M. Lawson, and Mijke Rhemtulla. 2022. "Statistical Control Requires Causal Justification." *Advances in Methods and Practices in Psychological Science* 5 (2): 251524592210958. https://doi.org/10.1177/25152459221095823.

Yan, Ting, and Trivellore Raghunathan. 2007. "Using Proxy Measures of the Survey Variables in Post-Survey Adjustments in a Transportation Survey." In *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass*, 3349–55.

Yuen, Vera Wing Han. 2022. "Political Attitudes and Efficacy of Health Expert Communication on the Support for COVID-19 Vaccination Program: Findings from a Survey in Hong Kong." *Vaccine* 40 (15): 2282–91. https://doi.org/10.1016/j.vaccine.2022.02.086.

5 General Conclusion and Discussion

Looking back, survey sampling has undergone significant changes since Neyman's groundbreaking article in 1934 (Neyman, 1934). Notwithstanding decreasing response rates, rapidly increasing costs for fielding probability based surveys (Brick, 2011; Bethlehem, 2016), and technological changes that gave rise to nonprobability based approaches to sampling (Baker et al., 2010; Callegaro et al., 2014; Cornesse et al., 2020), probability sampling remain the fundamental framework for survey sampling (Groves et al., 2009; Brick, 2011). Probability sampling gained its dominant position in social sciences through the combination of survey sampling and statistical inference and these qualities remain its strength (Smith, 1976; Groves et al., 2009; Brick, 2011; Bethlehem, 2016). But the challenges posed to the framework of probability sampling led to a coexistence of both probability and nonprobability sampling approaches in survey practice. In this dissertation, I addressed this coexistence and the methodological challenges that arise from the dual framework by providing three key contributions.

In chapter two, I compared estimates derived from four surveys that differed in survey mode and survey sampling across a wide range of statistical measures and analyses. The results consistently showed that decisions made by survey researchers on mode and sampling affect the outcomes of data analysis and thus, have significant implications for the results derived therefrom. Placing this finding within the wider context of the coexistence of probability and nonprobability sampling approaches, this study highlights the importance of investigating differences in sample selection and the potential introduction of errors resulting from the use of different sampling strategies. The empirical evidence presented underscores the necessity of methodological improvements in both probability and nonprobability sampling approaches.

Chapters three and four tackle these needs. Chapter three addresses some of the challenges faced by probability sampling, such as high survey costs and the inability to directly sample individuals for Web surveys. The chapter presented the text-to-Web recruitment approach as a cost-effective and easily implemented method for directly sampling, inviting, and surveying respondents via their smartphones.

Our findings indicate the feasibility of this approach, but also its limitations. Notably, there was a significant loss of numbers during the RDD sampling process and a low response rate. Consequently, our implementation of a text-to-Web recruitment may not be universally applicable for general population surveys. However, it could serve as a baseline study to tackle the persistent challenges of rising survey costs, declining response rates, and the need for an efficient way to integrate probability sampling with Web surveys.

Further research is needed to fully explore the potential of text-to-Web recruitment. For instance, this approach holds great promise in cases where individuals are sampled from a list that includes cell phone numbers, potentially addressing the challenges encountered in the RDD sampling process. When sampling individuals from a list, researchers may have access to background information about both respondents and non-respondents. This presents an opportunity to investigate two important questions: 1) Is it feasible to invite and survey respondents directly via their smartphone when sampling from a list? 2) Does this survey invitation mode work better for certain subgroups of the population than others? To explore these questions, a suitable database is the nation-wide voter file for the US, which is accessible via a commercial vendor. This database includes data from all voting registries in the US and provides additional variables, including cell phone numbers for registered voters. To investigate the feasibility of the text-to-Web approach and explore who participates in such surveys, researchers could use the following analysis strategy: First, they could systematically describe the approach and highlight potential challenges. Second, they could compare respondents from the text-to-Web survey with non-respondents from the sampling frame using a range of sociodemographic variables and variables that capture political behavior. This would allow for an exploration of response patterns, correlations with responding to text messages, and possible nonresponse bias. Overall, such a study would contribute to our understanding of the generalizability of data collected using text-to-Web recruitment for list-based population studies.

In Chapter four, we proposed a six-step selection strategy for adjustment variables that accounts for high correlations with survey participation and questions. This strategy aimed to improve inference based on nonprobability samples for general population surveys. While promising, our findings revealed that the method did not achieve the desired outcome of minimizing bias in estimates derived from nonprobability survey data. Our results consistently demonstrated that applying enhanced, theory-based weights did not significantly mitigate selection biases in the estimates. The poor performance of the enhanced, theory-based weights could be explained by the weak correlations of the selected adjustment variables with measures of survey participation. The more fundamental problem here may be that traditional survey participation theories used to justify the selection of adjustment variables are not applicable to nonprobability samples with recruited respondents from opt-in online panels.

This insight highlights the borader point that nonprobability sampling lacks a well-defined theoretical explanation of survey participation. To enhance the applicability of nonprobability surveys, further research is needed to understand why individuals choose to become members of opt-in online panels and participate in surveys. To date, only a few studies have sought to understand what brings people to join opt-in online panels and how they decide whether or not to participate in specific surveys (Brüggen et al., 2011; Keusch et al., 2014). However, a consolidated theoretical model that explains survey participation in nonprobability samples is missing. A theory aimed at explaining survey participation for nonprobability surveys must clearly distinguish between two steps: (1) why people join opt-in online panels and (2) why they decide to participate in a given survey.

One possible approach to developing a theory aimed at explaining survey participation for nonprobability surveys could be split into two steps: First, one could apply a qualitative research approach and conduct focus group interviews with individuals who volunteered to become members of opt-in online panels by registering on an Internet platform and non-members of opt-in online panels who did see the advertisement, but did not register. With this, one could evaluate basic motives for and against becoming a member of an opt-in online panel. Further, it would be informative to conduct those interviews with online panel members who differ in their survey participation behavior. This would allow to get insights into motives not only for becoming a member in an online panel, but also for the decision to participate in a given survey. For this purpose, one could include members of the online panel who participate in many surveys and members of the online panel who participate in few surveys in the focus groups. These interviews could serve as a basis for developing a theory of response behavior in online panel nonprobability surveys.

Subsequently, the theory should be tested empirically. For this purpose, it would be useful to compare the newly developed theory with traditional survey response theories. A possible study design could be to include new-developed items explaining survey participation in nonprobability surveys and items explaining survey response from traditional theories in one survey. Both respondents recruited from self-selected opt-in online panels and respondents recruited using probabilistic random sampling should be invited to participate in this survey. This approach allows to systematically compare the extent to which response behavior in nonprobability surveys differs from response behavior in traditional surveys. This yields new insights into why individuals (1) become members of online panels and (2) participate in surveys.

To summarize, this dissertation revealed that the differences between probability and nonprobability sampling not only exist in theory but can be observed empirically as estimates obtained from different survey samples substantially differ (chapter two). It highlighted the need for methodological enhancement in both probability and nonprobability sampling. In response, we introduced a new approach for probability sampling to reduce survey costs and make it more applicable to Web surveys (chapter three). At the same time, the findings highlighted the need to better understand self-selection mechanisms in nonprobability sampling to minimize selection bias (chapter four). Thus, this dissertation advances the development of both probability and nonprobability sampling, making them more suitable for the challenges faced by survey sampling today and in the future.

But what will this future look like?
I conclude with a quote of Brick (2011) who attempts to answers this question:

> "One thing we are confident about is that the future of sampling will be dynamic. Our society has continually expanding needs, and the scientific progress in society is extraordinary. Survey sampling is surely going to undergo changes, and the changes might be larger in scope than anticipated here. While change is certain to occur, we are much less confident about the specific path the future will follow" (Brick, 2011, p. 886).

References

Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. Dillman, et al. 2010. "Research Synthesis: AAPOR Report on Online Panels." *Public Opinion Quarterly* 74 (4): 711–81. https://doi.org/10.1093/poq/nfq048.

Bethlehem, Jelke. 2016. "Solving the Nonresponse Problem With Sample Matching?" *Social Science Computer Review* 34 (1): 59–77. https://doi.org/10.1177/0894439315573926.

Brick, J. M. 2011. "The Future of Survey Sampling." *Public Opinion Quarterly* 75 (5): 872–88. https://doi.org/10.1093/poq/nfr045.

Brüggen, Elisabeth, Martin Wetzels, Ko De Ruyter, and Niels Schillewaert. 2011. "Individual Differences in Motivation to Participate in Online Panels: The Effect on Reponse Rate and Reponse Quality Perceptions." *International Journal of Market Research* 53 (3): 369–90. https://doi.org/10.2501/IJMR-53-3-369-390.

Callegaro, Mario, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas, eds. 2014. *Online Panel Research*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118763520.

Cornesse, Carina, Annelies G Blom, David Dutwin, Jon A Krosnick, Edith D De Leeuw, Stéphane Legleye, Josh Pasek, et al. 2020. "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research." *Journal of Survey Statistics and Methodology* 8 (1): 4–36. https://doi.org/10.1093/jssam/smz041.

Groves, Robert M., Floyd J. Fowler, Mick Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau, eds. 2009. *Survey Methodology*. 2. ed. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley.

Keusch, Florian, Bernad Batinic, and Wolfgang Mayerhofer. 2014. "Motives for Joining Nonprobability Online Panels and Their Association with Survey Participation Behavior." In *Online Panel Research*, edited by Mario Callegaro, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas, 171–91. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118763520.ch8.

Neyman, Jerzy. 1934. "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection." *Journal of the Royal Statistical Society* 97 (4): 558. https://doi.org/10.2307/2342192.

Smith, T. M. F. 1976. "The Foundations of Survey Sampling: A Review." *Journal of the Royal Statistical Society. Series A (General)* 139 (2): 183. https://doi.org/10.2307/2345174.

6 Appendix

A. Would electoral resarch show different findings if we replaced probability face-to-face surveys with cheaper alternatives of data collection?

A.1 Data availability and replication materials

The data underlying chapter two are available at the GESIS data repository and can be accessed at [gesis.org] (GLES, 2019a,b, 2022a,b). The replication materials can be accessed through the Open Science Framework (OSF) repository at [https://osf.io/c54pz/files/osfstorage]. Within this repository, a dedicated folder contains comprehensive information on the data processing, variable operationalization, and data analysis procedures.

A.2 Overview of replicated models

I utilized the classification framework proposed by Smets and van Ham (2013) to categorize the models according to the broad theoretical models of individual-level voter turnout. To facilitate this process, I obtained the coding scheme from Smets and van Ham (2013). Table 6.1 provides an overview of the models replicated in Chapter 2 (Green and Shachar, 2000; Lyons and Alexander, 2000; Highton and Wolfinger, 2001; Holbrook et al., 2001; Clarke et al., 2002; Goldstein and Ridout, 2002; Mughan and Lacy, 2002; Mutz, 2002; Perea, 2002; Aarts and Semetko, 2003; Jackson, 2003; Blais et al., 2004; Heath, 2004; Rubenson et al., 2004; Anduiza–Perea, 2005; Chong and Rogers, 2005; Adams et al., 2006; Leighley and Nagler, 2007; Malhotra and Krosnick, 2007; Sanders et al., 2007; Wass, 2007; Killian et al., 2008; Pattie and Johnston, 2009; Stevens, 2009; Yoo, 2010). Additionally, it presents the independent variables as well as the control variables employed in each model[1].

[1] A detailed description on the operationalization of the variables can be accessed at [https://osf.io/c54pz/files/osfstorage].

Table 6.1: Overview of replicated models

Study	Broad theoretical model	DV	IV	CV	Hypothesis	GLES:DV	GLES:IV	GLES:CV
Aarts and Semetko (2003)	ressource	reported turnout	television use	age	How media use influences voter turnout	voting intention	reading popular newspapers	age
				education			reading high quality newspapers	education
				political interest			watching public TV	political interest
							watching private TV	
							watching TV news	
Adams et al. (2006)	socialization	alienation		black	How the perception of candidates influences voter turnout	voting intention		political efficacy
				political efficacy				voted in previous election
				voted in previous election				education

Table 6.1: Overview of replicated models *(continued)*

Anduiza–Perea (2005)	mobilization	reported turnout	attention to content	age	How the attention to election camapign influences voter turnout	voting intention	interest in campaign	Age
			attention to media	interest in politics			reading news	interest in politics
			talks about election	party closeness			watching TV news	left-right Self placement
			evaluation of campaign	left-right Self placement			talking about politics	evaluation of economic situation
			contacted by party	evaluation of political situation			evaluation of campaign	
			changes in evaluation of political situation	evaluation of economic situation			changes in evaluation of economic situation	
			changes in evaluation of economic situation					

Table 6.1: Overview of replicated models *(continued)*

Blais et al. (2004)	ressource	reported turnout	age		How the affiliation to a specific generation influences voter turnout	voting intention	babyboomer	
			age suqared				generation 60s	
			babyboomer				generation 70s	
			generation 60s				age	
			generation 70s				age squared	
			period post 90					
			seasonal elections					
Blais et al. (2004)	ressource	reported turnout	babyboomer	sense of duty	How the affiliation to a specific generation influences voter turnout	voting intention	babyboomer	sense of duty
			generation 60s	attention to politics			generation 60s	attention to politics
			generation 70s				generation 70s	
Blais et al. (2004)	ressource	reported turnout	age	education	How the affiliation to a specific generation influences voter turnout	voting intention	age	education

Table 6.1: Overview of replicated models *(continued)*

			age2	gender			age2	gender
			babyboomer	union member			babyboomer	union member
			generation 60s	religiosity			generation 60s	religiosity
			generation 70s	region			generation 70s	region
			period post 90	foreign born				foreign born
			seasonal elections	Non european orgin				married
				married				income
				income				
Chong and Rogers (2005)	ressource	voted in primary	helped voter registration		How campaign activities influences voter turnout	voting intention	attended a demonstration	
			went to political meeting				signed a petition	
			gave money to candidate				picketed, boycoted	
			campaigned for black candidate				worked for party	
			worked for party					0
			contacted a public official					0
			signed a petition					0

Table 6.1: Overview of replicated models *(continued)*

			attended a demonstration					0
			picketed, boycoted					0
Clarke et al. (2002)	rational choice	turnout	general incentives	age	How influence and benefits of voting influence voter turnout	voting intention	efficacy	age
			influence-discounted benefits	disability			party-positive	gender
			costs	ethnicity			party-negative	region
			personal benefits	gender			civic duty	
			democracy satisfaction	region			socal class	
			social norms				education	
			civic duty				election interest	
			cognitive engagement				democracy satisfaction	
			election interest					
			political knowledge					
			civic voluntarism					
			education					

Table 6.1: Overview of replicated models *(continued)*

			socal class					
			party mobilization					
			equity fairness					
			relative deprivation					
			social capital					
			social trust					
Goldstein and Ridout (2002)	mobilization	voting	contacted by party	south	How contact during electoral campaign influence voter turnout	voting intention	contacted by party	income
				income				union member
				union member				strength of party ID
				strength of party ID				married
				married				interest in politics
				interest in politics				political efficiacy
				political efficiacy				voted in previous election

Table 6.1: Overview of replicated models *(continued)*

				voted in previous election				region
				house competiveness				
				senate competiveness				
Green and Shachar (2000)	rational choice	voter turnout	turnout previous election	discussion	How voting in previous election influence voter turnout	voting intention	turnout previous election	net-feelings
				group membership				worked for party
				political efficacy				contacted by party
				personal trust				discussion
				length of residence				organization member
				education				political efficacy
				family income				education
				race				family income
				marital status				marital status
				home ownership				region
				south/nonsouth				age
				age				age2

Table 6.1: Overview of replicated models *(continued)*

				age2				
Heath (2004)	ressource	voting	Ethnic	age	How perceived costs and benefits of participation influence political participation	voting intention	political efficacy	age
		0	costs	age2			uni	age2
		0	benefits	gender			alevel	gender
		0	uni				GCSE	
		0	alevel				interest in politics	
		0	GCSE				Party ID	
		0	knowledge				satisfaction with democracy	
		0	SOCCAP					
		0	attention					
		0	influence					
		0	Party ID					
		0	political trust					
		0	democracy satisfaction					

Table 6.1: Overview of replicated models *(continued)*

Highton and Wolfinger (2001)	ressource	turnout (<25)	residential stability	gender	How life-cycle effects influence voter turnout among young adults (<25y)	voting intention (u25)	married	gender
			marital status	family income			labor force status	family income
			home ownership	education			student	education
			labor force status	family income			age-18	
			student status	education			age-20	
			living with parent(s)				age-22	
			Age					
Holbrook et al. (2001)	psychological	voter turnout	attitude: both positive	political efficacy	How favorable and unfavorable beliefs on attitudes toward presidential candidates influence voter turnout	voting intention	attitude: both positive	education
			attitude: both negative	age			attitude: both negative	political efficacy

Table 6.1: Overview of replicated models *(continued)*

	|Attitude1 - Attitude 2|	age2	|Attitude1 - Attitude 2|	age
	|Attitude1 - Attitude 2|xboth positive	black	|Attitude1 - Attitude 2|xboth positive	age2
	|Attitude1 - Attitude 2|xboth negative	mexican-american	|Attitude1 - Attitude 2|xboth negative	region
		puerto-rican		income
		southern		party id strength
		border state		contacted by a party
		income		care about election
		home ownership		employed
		years lived in community		unemployed
		employed		
		unemployed		
		party id strength		
		perception of closeness of election		

Table 6.1: Overview of replicated models *(continued)*

				contacted by a party				
				care about election				
Jackson (2003)	ressource	turnout		2nd income quartile	How deficit on socioeconomic status and social-connectedness influence vote turnout	voting intention		2nd income quartile
				3rd income quartile				3rd income quartile
				4th income quartile				4th income quartile
				some college				some college
				college				college
				advanced degree				advanced degree
				female				female
				age				age
				age2				age2
			0	1-2 year resident				unemployed
				3-4 year resident				not in labor force

Table 6.1: Overview of replicated models *(continued)*

	5+ year resident	
	home owner	
	unemployed	
	retired or disabled	
	not in labor force	
	pre 1980s immigrant	
	1980s immigrant	
	1990s immigrant	
	African-American	
	American Indian	
	Aisna-American	
	Mexican-American	
	Puerto Rican	
	Cuban American	
	Central or South American	

Table 6.1: Overview of replicated models *(continued)*

				Other spanish				
Killian et al. (2008)	rational choice	turnout	pocketbook evaluations	political efficacy	How pocketbook and sociotrophic considerations influence vote turnout	voting intention	pocketbook evaluations	political efficacy
			sociotropic evaluations	strength of partisanship			sociotropic evaluations	strength of partisanship
				education				education
				age				age
				income				income
				race				gender
				gender				
Leighley and Nagler (2007)	mobilization	turnout	union member	family income	How union membership and union strength influence vote turnout	voting intention	union member	family income
			union density	education				education
			union strength	age				age
				age2				age2
				male				male
				black				married
				married				

Table 6.1: Overview of replicated models *(continued)*

Lyons and Alexander (2000)	ressource	turnout	Membership in cohort one (born prior to 1932)	south	How cohort memebership influence vote turnout	voting intention	Membership in cohort one (born prior to 1932)	age
				nonwhite				income third
				age				education beyond high School
				income third				number of contacts initiated by parties
				education beyond high School				number of likes and dislikes about the candidates
				number of contacts initiated by parties				number of likes and dislikes about the parties
				number of likes and dislikes about the candidates				strength of party identification
				number of likes and dislikes about the parties				amount of interest in the election

Table 6.1: Overview of replicated models *(continued)*

				strength of party identification				care about the outcome of presidential election
				amount of interest in the election				region
				care about the outcome of presidential election				
Malhotra and Krosnick (2007)	ressource	vote	party ID	0	How survey mode and sampling influence reported turnout	voting intention	party ID	
			presidential job approval				sociotropic retrispective	
			Iraq worth it				pocketbook retrospective	
			sociotropic retrispective				gender	
			pocketbook retrospective				education	
			male				age	
			black					
			high school					
			some college					

Table 6.1: Overview of replicated models *(continued)*

			college					
			graduate					
			age					
Mughan and Lacy (2002)	rational choice	abstain	insecure about American jobs		How job insecurity influence vote turnout	voting intention	insecure about own job	party
			insecure about own job	republican			national economy worse	trade-union
			national economy worse	democrat			personal finances worse	gender
			personal finances worse	conservative				education
			disapprove of clinton on economy	liberal				income
				union household				age (18-29)
				female				age(30-44)
				white				age(45-64)
				high school graduate				
				some college				
				college graduate				
				income				

Table 6.1: Overview of replicated models *(continued)*

				age (18-29)				
				age(30-44)				
				age(45-64)				
Mutz (2002)	mobilization	intent to vote	cross-cutting exposure	political interest	How networks with political disagreement influence vote turnout	voting intention	cross-cutting exposure	political interest
			frequency of political talk	education			frequency of political talk	education
			size of network	republican			size of network	party membership
				democrat				age
				age				income
				income				gender
				race				
				gender				
				political knowledge				
				minor children				
Pattie and Johnston (2009)	mobilization	self-reported turnout	Number of political discussants	age group	How networks with political disagreement influence vote turnout	voting intention	Number of political discussants	age
			number of discussants disagreeing	class			number of discussants disagreeing	class

Table 6.1: Overview of replicated models *(continued)*

				education			0	education
				gender			0	gender
				partisanship			0	partisanship
				interest in election			0	interest in election
				when people like me vote, they can really change the way Britain is governed			0	
Perea (2002)	psychological	turnout	individual incentives		How individual characteristics and institutional incentives influence vote turnout	invote	individual incentives	
			compulsory voting				preference expression	
			voting facilities					
			preference expression					
Rubenson et al. (2004)	ressource	turnout	age		How the age gap influence vote turnout	voting intention	age	0

Table 6.1: Overview of replicated models *(continued)*

			income				income	0
			education				education	0
			new immigrant				religiosity	0
			religiosity				married	0
			married				male	0
			male				age x income	0
			age x income				age x Education	0
			age x Education					0
Rubenson et al. (2004)	mobilization	turnout	party identification	age	How the age gap influence vote turnout	voting intention	party identification	age
			contacted during campaign	income			contacted during campaign	income
			negative party Sentiment	education			negative party Sentiment	education
			political cynicism	new immigrant			political interest	religiosity
			ge x cynosism	religiosity			political information	married
			political information	married				male
			political interest	male				age x Income

Table 6.1: Overview of replicated models *(continued)*

				age x Income				age x Education
				age x Education				
Sanders et al. (2007)	institutional	turnout	efficacy x collective benefits		How the survey mode influence reported turnout	voting intention	political efficacy	
			personal benefits					
			costs				civic duty	
			civic duty				democracy dissatisfaction	
			democracy dissatisfaction				election interest	
			election interest				party mobilization	
			party mobilization				age	
			relative deprivation				education	
			social norms				gender	
			social trust				social class	
			age				region	
			education					
			ethnicity(white)					
			gender(male)					

Table 6.1: Overview of replicated models *(continued)*

			social class					
			region					
Stevens (2009)	ressource	turnout	logged volume of exposure to negative adds	age	How the proportion of negative advertising influence vote turnout	voting intention	news viewing	age
			proportion of negative ads to which exposed	african-american			newspaper reading	gender
			total spots in market	female				income
			local news viewing	income				strength of partisanship
			newspaper reading	education				mobilized
				political information				
				strength of partisanship				
				mobilized				
				house race competitive				
				senate race competitive				

Table 6.1: Overview of replicated models *(continued)*

				presidential race competitive				
Wass (2007)	ressource	turnout	age	gender	How age, generation and period influence vote turnout	voting intention	age	gender
			age2				age2	
			generation of reconstruction				generation-reconstruction	
			generation of transforma-tion				generation-transformation	
			generation of suburban				generation-suburban	
			generation of individual choice				generation-individual	
			period					
Yoo (2010)	psychological	turnout	ambivalence		How ambivalence and indifference influence vote turnout	voting intention	ambivalence	

Table 6.1: Overview of replicated models *(continued)*

	ambivalence-Squared	ambivalence-Squared
	PID strength	PID strength
	concern of Election Outcome	concern of Election Outcome
	political knowledge	political efficacy
	external Efficacy	education
	internal Efficacy	contacted by Party
	education	
	contacted by Party	
	perceived closeness of election	

Note: DV = dependent variable; IV = independent variable; CV = control variable. Classification of broad theoretical models according to Smets and van Ham (2013).

B. Exploring the feasibility of recruiting respondents and collecting Web data via smartphone: A case study of text-to-Web recruitment for a general population survey in Germany

B.1 Data availability and replication materials

The data underlying this study, and the software code with which the data were analyzed, are available in the Open Science Framework (OSF) repository and can be accessed at [https://osf.io/fjenk/]. Within this repository, a dedicated folder contains comprehensive information on the data processing, variable operationalization, and data analysis procedures.

B.2 Sampling: Detailed information about the way we implemented mobile RDD sampling

To minimize the proportion of unassigned numbers in our sample, we implemented a two- step procedure when generating a random selection of cellphone numbers. First, we used the GESIS sampling frame, from which we chose a simple random sample of numbers. This sampling frame was developed for scientific purposes in Germany in 2009 by Häder and Gabler to improve the sample quality of mobile RDD samples, and it does not contain banks of numbers that belong entirely to businesses and technical services or numbers that have not yet been assigned to customers (Häder et al., 2009a; Häder and Sand, 2019). The sampling frame has been developed and improved continuously since then (for more information, see Häder et al., 2009b; Häder, 2016).

To further reduce the proportion of unassigned numbers, home location register (HLR) lookups were conducted as a second step. These HLR lookups enabled us to check whether a number was accessible without having to contact it (for detailed information on HLR lookups in the German telephone networks, see Sand, 2017). When a lookup is successful, more than 98% of the numbers can be classified as either assigned or unassigned. Earlier findings regarding HLR lookups in the German cellular networks show that about 50% of numbers are unassigned (Sand, 2017). This corresponds approximately to the proportion of unassigned numbers found in the successful HLR lookups conducted within the framework of our study: over 50% of the numbers (16,764) were removed from the sample because they were not assigned.

B.3 Text message invitation: Detailed information on the sending of the text messages

To try to find an explanation for why only 6,016 of the 13,338 SMS sent could be successfully delivered, we undertook some investigations. Generally speaking, there are a number of possible reasons why SMS messages cannot be delivered—for example, phones are out of area over a longer period, or numbers are unassigned (Steeh et al., 2007). In our study, however, we observed that the percentage of delivered text messages did not remain constant over time. Whereas between 82% and 55% of the SMS messages sent in each of the first eight tranches could be successfully delivered, this percentage dropped sharply to values between 24% and 20% between the ninth and the 14th tranche. Based on this observation, we suspect that there were other systematic reasons behind the large proportion of undelivered text messages. Figure 5.1 shows the proportion of delivered text messages over time.

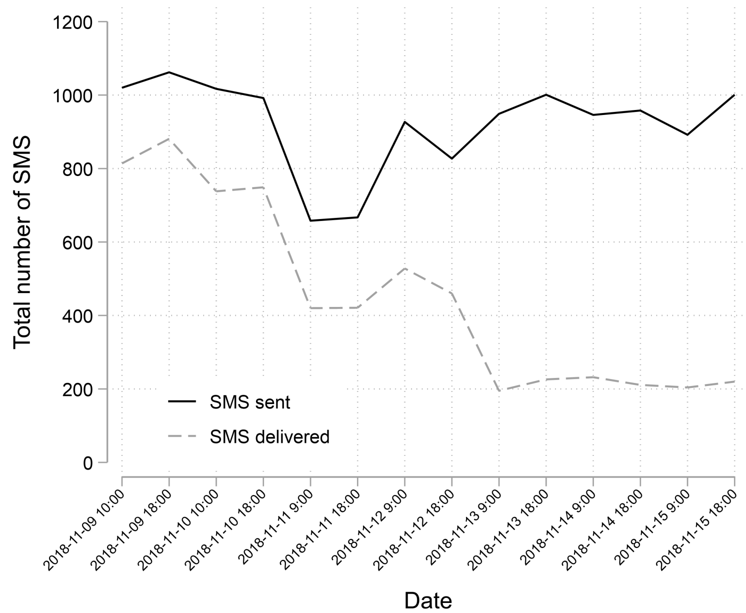

Figure 6.1: Proportion of delivered text messages over time.

One explanation for the high rate of undelivered text messages is that our invitation SMS were blocked by the network operators. As SMS have been used increasingly for advertising purposes in recent years, such “carrier filters” have become a common instrument employed by some network operators to filter spam SMS (for detailed information, see Carvalho, 2015). However, the assumption that our SMS were marked as “spam” could not be confirmed based on the information available to us, so that we could distinguish only between delivered and undelivered SMS.

As it is not known whether the 7,322 numbers to which the messages were not delivered were assigned to mobile phones, we decided to classify them in accordance with (AAPOR, 2016) as unknown eligibility (see table 3.1 in section 3.2.3). By contrast, because delivery reports can be regarded as indicators of working numbers (Callegaro et al., 2007), the 6,016 numbers to which invitations were delivered were classified as eligible units.

B.4 Web survey: Detailed information on the identification of non-human accesses to our survey landing page

To detect bots in our survey, we compiled a list of known bots and crawlers (Monperrus, 2019) and compared it with the user agents (UAs) collected in our survey. One explanation for the fact that the landing page of our survey was accessed by various bots is that we embedded an impersonalized link in the SMS. Therefore, access to our survey was not protected by password. Rather, the link was freely available on the Internet and could be accessed multiple times.

Furthermore, the analysis of the UAs showed that most of the accesses that did not start the survey came from one single UA (n = 673). As it seemed unlikely to us that such a high number of accesses were made by one single device, we undertook further investigations and discovered that this frequently occurring UA is automatically transmitted when the link preview for SMS messages on any kind of smartphone is activated. Including this link preview, we found that a total of 717 accesses were attributable to bots and link previews . Table 5.2 gives an overview of the identified non-human accesses to our survey and their UAs.

Table 6.2: User Agent Strings (UAs) of the identified non-human accesses

User Agent	Type	Classification	n
Mozilla/5.0 (X11; Ubuntu; Linux i686; rv:24.0) Gecko/20100101 Firefox/24.0	Non-human visit	Link preview	673
Mozilla/5.0 (Macintosh; Intel Mac OS X 10-11-1) AppleWebKit/601.2.4 (KHTML, like Gecko) Version/9.0.1 Safari/601.2.4 facebookexternalhit/1.1 Facebot Twitterbot/1.0	Bot	facebookexternalhit	29
Mozilla/5.0 (X11; Linux x86-64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/56.0.2924.87 Safari/537.36 Google (+https://developers.google.com/+/	Bot	Google+Snippet Fetcher	14
WhatsApp/2.18.341 A	Bot	WhatsApp	1
Total			717

B.5 Results: Detailed information on the demographic characteristics and voting intentions of the respondents

In this appendix, we provide the data underlying figures 3.2 and 3.3 in section 3.3.2. Only valid values are reported.

Table 6.3: Gender distribution of our respondents

	n	Proportion
Male	73	0.53
Female	64	0.47
Total	137	

Table 6.4: Age distribution of our respondents

	n	Proportion
29 years and younger	62	0.45
30 to 44 years	31	0.23
45 to 59 years	28	0.21
60 years and older	14	0.1
Total	135	

Table 6.5: Educational level of our respondents

	n	Proportion
Low	37	0.27
Medium	32	0.24
High	66	0.49
Total	135	

The classification of education into the categories low, medium, and high was based on the variable highest level of general education. The levels were as follows:

1) low: finished school without a school leaving certificate, obtained the lowest formal qualification in Germany's three-tier secondary school system after 8

or 9 years of schooling (Hauptschulabschluss, Volksschulabschluss), or still in general education;

2) medium: obtained an intermediary secondary qualification after 10 years of schooling (Mittlere Reife, Realschulabschluss or Polytechnische Oberschule mit Abschluss 10. Klasse);
3) high: obtained a qualification entitling the holder to study at a university of applied sciences (Fachhochschulreife) or a general higher education entrance qualification entitling the holder to study at any higher education institution (Hochschulreife).

Table 6.6: Voting intentions of our respondents

	n	Proportion
CDU/CSU (Christlich Demokratische Union / Christlich-Soziale Union)	14	0.17
SPD (Sozialdemokratische Partei Deutschlands)	6	0.07
FDP (Freie Demokratische Partei)	5	0.06
Bündnis 90/Die Grünen	20	0.25
DIE LINKE (die Linke)	11	0.14
AfD (Alternative für Deutschland)	14	0.17
Other	11	0.14
Total	81	

C: Enhancing model-based adjustments of nonprobability surveys: Selecting auxiliary variables based on theoretical assumptions about their association with survey participation and variables of interest

C.1 Data availability and replication materials

The data underlying chapter four are available at the GESIS data repository and can be accessed at [gesis.org] (GLES, 2022, 2023). The replication materials can be accessed through the Open Science Framework (OSF) repository at [https://osf.io/w3djp/files/osfstorage].

C.2 Study I: Operationalization of survey participation and political attitudes and behavior

We compute indicators with items available in the GLES data to operationalize political attitudes. To measure conservative attitudes as well as trust in institutions, we follow the operationalization by a recently published study that uses these indicators (Steiner et al., 2022) and also relies on GLES data. To measure whether a respondent holds conservative attitudes, we use items on the integration of immigrants, the European Union, gender equality, and climate change, measured on a 7 point scale. To measure trust in institutions, we use items on trust in the parliament, the Federal

Constitutional Court, the armed forces, and trust in politics, measured on a 5 point scale. We use principal component analysis to compute factors used in the analysis. To measure populist attitudes of the respondents, we use the approach introduced by Wuttke et al. (2020) and first compute the score of each subdimension of populist attitudes (antielitism, sovereignty, homogeneity) and then compute the product of each subdimension. For voting behavior, we dichotomized the reported voting behavior in voting for a conservative/populist party or not. A detailed overview of the operationalization and factor analysis can be accessed through the Open Science Framework (OSF) repository: [https://osf.io/w3djp/files/osfstorage].

For survey participation, we used the following item to measure whether respondents belief voting is a civic duty, measured on a five point scale (strongly agree - strongly disagree): " In a democracy, it is the duty of all citizens to vote regularly in elections." The attention check implemented was as follows: "To make sure that this survey is being filled out by a human, please click here on 'disagree'."

C.3 Study II: Operationalization of survey participation and political attitudes and behavior

To measure internal political efficacy, we use the short-scale introduced by Beierlein et al. (2014). For political knowledge, we do not have measures in our dataset. Therefore we cannot compute any correlation coefficients. For voting behavior, we dichotomized the reported voting behavior/voting intention in voted or not. A detailed overview of the operationalization and factor analysis can be accessed through the Open Science Framework (OSF) repository:

[https://osf.io/w3djp/files/osfstorage].

References

AAPOR. 2016. "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys." The American Association for Public Opinion Research.

Aarts, Kees, and Holli A. Semetko. 2003. "The Divided Electorate: Media Use and Political Involvement." *The Journal of Politics* 65 (3): 759–84. https://doi.org/10.1111/1468-2508.00211.

Adams, James, Jay Dow, and Samuel Merrill. 2006. "The Political Consequences of Alienation-Based and Indifference-Based Voter Abstention: Applications to Presidential Elections." *Political Behavior* 28 (1): 65–86. https://doi.org/10.1007/s11109-005-9002-1.

AnduizaPerea, Eva. 2005. "Campaign Effects in the Spanish Election of 2000." *Journal of Elections, Public Opinion and Parties* 15 (2): 215–36. https://doi.org/10.1080/13689880500178815.

Beierlein, C., C. J. Kemper, A. Kovaleva, and B. Rammstedt. 2014. "Political Efficacy Kurzskala (PEKS)." *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/ZIS34.

Blais, Andre, Elisabeth Gidengil, and Neil Nevitte. 2004. "Where Does Turnout Decline Come From?" *European Journal of Political Research* 43 (2): 221–36. https://doi.org/10.1111/j.1475-6765.2004.00152.x.

Callegaro, Mario, Charlotte Steeh, Trent D. Buskirk, Vasja Vehovar, Vesa Kuusela, and Linda Piekarski. 2007. "Fitting Disposition Codes to Mobile Phone Surveys:

Experiences from Studies in Finland, Slovenia and the USA." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (3): 647–70. https://doi.org/10.1111/j.1467-985X.2006.00461.x.

Carvalho, Edwin. 2015. "SMS 101: Why Messages Fail to Deliver."

Chong, Dennis, and Reuel Rogers. 2005. "Racial Solidarity and Political Participation." *Political Behavior* 27 (4): 347–74. https://doi.org/10.1007/s11109-005-5880-5.

Clarke, Harold D., David Sanders, Marianne C. Stewart, and Paul F. Whiteley. 2002. "Downs, Stokes and Modified Rational Choice: Modelling Turnout in 2001." *British Elections & Parties Review* 12 (1): 28–47. https://doi.org/10.1080/13689880208413068.

GLES. 2019a. "Langfrist-Online-Tracking T37 (GLES)." https://doi.org/10.4232/1.13295.

———. 2019b. "Vorwahl-Querschnitt (GLES 2017)." https://doi.org/10.4232/1.13234.

GLES. 2022a. "GLES Cross-Section 2021, Pre-Election GLES Querschnitt 2021, Vorwahl." GESIS Data Archive. https://doi.org/10.4232/1.13860.

GLES. 2022b. "GLES Panel 2021, Control Group III (to Panel Wave 20, Sample A)GLES Panel 2021, Kontrollquerschnitt III (zu Welle 20, Sample A)." GESIS. https://doi.org/10.4232/1.14053.

GLES. 2022c. "GLES Tracking September 2021, T50 GLES Tracking September 2021, T50." GESIS. https://doi.org/10.4232/1.14000.

GLES. 2023. "GLES Panel 2022, Wave 23 GLES Panel 2022, Welle 23." GESIS. https://doi.org/10.4232/1.14064.

Goldstein, Kenneth M., and Travis N. Ridout. 2002. "The Politics of Participation: Mobilization and Turnout over Time." *Political Behavior* 24 (1): 3–29. https:

//doi.org/10.1023/A:1020930604938.

Green, Donald P., and Ron Shachar. 2000. "Habit Formation and Political Behaviour: Evidence of Consuetude in Voter Turnou." *British Journal of Political Science* 30: 561–73.

Häder, Sabine. 2016. "Sampling in Practice." SDM-Survey Guidelines (GESIS Leibniz Institute for the Social Sciences). https://doi.org/10.15465/gesis-sgen014.

Häder, Sabine, Siegfried Gabler, and Christiane Heckel. 2009. "Stichprobenziehung Für Die CELLA-Studie." In *Telefonbefragungen Über Das Mobilfunknetz*, edited by Michael Häder and Sabine Häder, 21–49. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91490-93.

Häder, Sabine, Michael Häder, Jennifer Graeske, Tanja Kunz, and Götz Schneiderat. 2009. "Realisierung Der Stichprobe." In *Telefonbefragungen Über Das Mobilfunknetz*, edited by Michael Häder and Sabine Häder, 71–82. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91490-96.

Häder, Sabine, and Matthias Sand. 2019. "Telefonstichproben." In *Telefonumfragen in Deutschland*, edited by Sabine Häder, Michael Häder, and Patrick Schmich, 39:113–51. Schriftenreihe Der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23950-36.

Heath, Oliver. 2004. "Modelling the Components of Political Action in Britain: The Impact of Civic Skills, Social Capital and Political Support." *British Elections & Parties Review* 14 (1): 75–93. https://doi.org/10.1080/1368988042000258781.

Highton, Benjamin, and Raymond E. Wolfinger. 2001. "The First Seven Years of the Political Life Cycle." *American Journal of Political Science* 45 (1): 202. https://doi.org/10.2307/2669367.

Holbrook, Allyson L., Jon A. Krosnick, Penny S. Visser, Wendi L. Gardner, and John

T. Cacioppo. 2001. "Attitudes Toward Presidential Candidates and Political Parties: Initial Optimism, Inertial First Impressions, and a Focus on Flaws." *American Journal of Political Science* 45 (4): 930. https://doi.org/10.2307/2669333.

Jackson, Robert A. 2003. "Differential Influences on Latino Electoral Participation." *Political Behavior* 25 (4): 339–66. https://doi.org/10.1023/B:POBE.0000004062.12215.63.

Killian, Mitchell, Ryan Schoen, and Aaron Dusso. 2008. "Keeping Up with the Joneses: The Interplay of Personal and Collective Evaluations in Voter Turnout." *Political Behavior* 30 (3): 323–40. https://doi.org/10.1007/s11109-007-9051-8.

Leighley, Jan E., and Jonathan Nagler. 2007. "Unions, Voter Turnout, and Class Bias in the U.S. Electorate, 1964–2004." *The Journal of Politics* 69 (2): 430–41. https://doi.org/10.1111/j.1468-2508.2007.00541.x.

Lyons, William, and Robert Alexander. 2000. "A Tale of Two Electorates: Generational Replacement and the Decline of Voting in Presidential Elections." *The Journal of Politics* 62 (4): 1014–34. https://doi.org/10.1111/0022-3816.00044.

Malhotra, Neil, and Jon A. Krosnick. 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." *Political Analysis* 15 (3): 286–323. https://doi.org/10.1093/pan/mpm003.

Monperrus, Martin. 2019. "Syntatic Patterns of HTTP User-Agents Used by Bots / Robots / Crawlers / Scrapers / Spiders."

Mughan, Anthony, and Dean Lacy. 2002. "Economic Performance, Job Insecurity and Electoral Choice." *British Journal of Political Science* 32 (03): 513–33. https://doi.org/10.1017/S0007123402000212.

Mutz, Diana C. 2002. "The Consequences of Cross-Cutting Networks for Political Participation." *American Journal of Political Science* 46 (4): 838. https://doi.

org/10.2307/3088437.

Pattie, C. J., and R. J. Johnston. 2009. "Conversation, Disagreement and Political Participation." *Political Behavior* 31 (2): 261–85. https://doi.org/10.1007/s11109-008-9071-z.

Perea, Eva Anduiza. 2002. "Individual Characteristics, Institutional Incentives and Electoral Abstention in Western Europe." *European Journal of Political Research* 41 (5): 643–73. https://doi.org/10.1111/1475-6765.00025.

Rubenson, Daniel, André Blais, Patrick Fournier, Elisabeth Gidengil, and Neil Nevitte. 2004. "Accounting for the Age Gap in Turnout." *Acta Politica* 39 (4): 407–21. https://doi.org/10.1057/palgrave.ap.5500079.

Sand, Matthias. 2017. "Evaluierung von HLR-Lookup-Verfahren." In *Methodische Probleme von Mixed-Mode-Ansätzen in Der Umfrageforschung*, edited by Stefanie Eifler and Frank Faulbaum, 211–37. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15834-79.

Sanders, David, Harold D. Clarke, Marianne C. Stewart, and Paul Whiteley. 2007. "Does Mode Matter For Modeling Political Choice? Evidence From the 2005 British Election Study." *Political Analysis* 15 (3): 257–85. https://doi.org/10.1093/pan/mpl010.

Smets, Kaat, and Carolien van Ham. 2013. "The Embarrassment of Riches? A Meta-Analysis of Individual-Level Research on Voter Turnout." *Electoral Studies* 32 (2): 344–59. https://doi.org/10.1016/j.electstud.2012.12.006.

Steeh, Charlotte, Trent D. Buskirk, and Mario Callegaro. 2007. "Using Text Messages in U.S. Mobile Phone Surveys." *Field Methods* 19 (1): 59–75. https://doi.org/10.1177/1525822X06292852.

Steiner, Nils D., Christian H. Schimpf, and Alexander Wuttke. 2022. "Left Behind and United by Populism? Populism's Multiple Roots in Feelings of Lacking

Societal Recognition.” *Politische Vierteljahresschrift*, August. https://doi.org/10.1007/s11615-022-00416-4.

Stevens, Daniel. 2009. “Elements of Negativity: Volume and Proportion in Exposure to Negative Advertising.” *Political Behavior* 31 (3).

Wass, Hanna. 2007. “The Effects of Age, Generation and Period on Turnout in Finland 1975–2003.” *Electoral Studies* 26 (3): 648–59. https://doi.org/10.1016/j.electstud.2007.06.002.

Wuttke, Alexander, Christian Schimpf, and Harald Schoen. 2020. “When the Whole Is Greater Than the Sum of Its Parts: On the Conceptualization and Measurement of Populist Attitudes and Other Multidimensional Constructs.” *American Political Science Review* 114 (2): 356–74. https://doi.org/10.1017/S0003055419000807.

Yoo, Sung-jin. 2010. “Two Types of Neutrality: Ambivalence Versus Indifference and Political Participation.” *The Journal of Politics* 72 (1): 163–77. https://doi.org/10.1017/s0022381609990545.