# RECRUITMENT STRATEGIES FOR A PROBABILITY-BASED ONLINE PANEL: EFFECTS OF INTERVIEW LENGTH, QUESTION SENSITIVITY, INCENTIVES AND INTERVIEWERS

Inauguraldissertation

zur Erlangung des akademischen Grades

einer Doktorin der Sozialwissenschaften

der Universität Mannheim

Vorgelegt von Ines Schaurer

Hauptamtlicher Dekan der Fakultät für Sozialwissenschaften:

Prof. Dr. Michael Diehl

Gutachter/ Betreuer: Prof. Dr. Michael Bošnjak

Gutachterin/ Betreuerin: Prof. Annelies Blom, Ph.D.

Gutachter: Prof. Dr. Marek Fuchs

Tag der Disputation: 29. September 2017

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ACKNOWLEDGEMENTS

# 1. Introduction and problem statement

Probability-based online panels of the general population are widely discussed in the scientific community as an alternative to interviewer-administered studies of the general population (Bosnjak, Das, & Lynn, 2016). They are increasingly used throughout Europe and the United States (Blom, Bosnjak, et al., 2016; Bosnjak et al., 2016; Hays, Liu, & Kapteyn, 2015).

Probability-based online panels represent a pool of potential respondents that are easily accessible for several surveys (Callegaro, Baker, et al., 2014). The respondent selection is based on a probability sample of the general population, resulting in a known inclusion probability for each sampled member. As no sampling frame is available for online surveys, researchers have to rely on traditional sampling methods. In consequence, respondents need to be recruited in an offline mode (i.e., using face-to-face, telephone, or postal surveys) and are switched to the online mode for the regular surveys (Couper, 2000). Thus, the recruitment process of probability-based online panels is characterized by a sequential mixed-mode strategy (Bosnjak, 2002). Furthermore, probability-based online panels need to include those parts of the general population that do not use the internet (Couper, 2000) to avoid coverage errors associated with data collection via the internet (Baker et al., 2010). The panel members are regularly invited to participate in internet surveys (Couper, 2000) that do not have to be thematically related (Callegaro, Baker, et al., 2014). In contrast to traditional panels of the social sciences, the main aim is not to study social change (Lynn, 2009) but to provide a selection of respondents that can be invited to several surveys (Callegaro, Baker, et al., 2014).

Surveys based on a probability-based online panel have several advantages compared to conventional interviewer-administered surveys. They can be realized relatively fast (Scherpenzeel & Bethlehem, 2010) and represent a cost-effective alternative to face-to-face and telephone surveys if recruited panel members participate in several subsequent surveys (Lynn, 2013). Furthermore, a comprehensive set of demographic background information can be collected over time (Lensvelt-Mulders, Lugtig, & Hox, 2008). Additionally, probability-based online panels benefit from the fact that the social desirability bias is lower in self-administered surveys compared to interviewer-administered surveys (Kreuter, Presser, & Tourangeau, 2008; Tourangeau, Conrad, & Couper, 2013).

Examples of probability-based online panels are the Dutch LISS (Longitudinal Internet Studies for the Social Sciences) Panel (Das, 2012; Scherpenzeel & Toepoel, 2012), the French ELIPSS Panel (Revilla, Cornilleau, Cousteaux, Legleye, & Pedraza, 2015), the German Internet Panel (Blom, Gathmann, & Krieger, 2015) and the GESIS Panel (Bosnjak et al., 2017) in Germany, and the Amerispeak Panel (Dennis, 2017) and the Understanding America Study (Hays et al., 2015) in the United States.

Notwithstanding of different recruitment protocols, future panelists go through several stages before becoming an active panel member (Vehovar, Lozar Manfreda, Zaletel, & Batagelj, 2002).

Due to this multi-step approach, probability-based online panels are prone to nonresponse (Bartsch, 2011; Bosnjak, 2002; Lensvelt-Mulders et al., 2008), resulting in a cumulative response rate across stages that is low compared to more traditional modes of data collection (Callegaro & DiSogra, 2008; Tourangeau et al., 2012, p. 41).

The typical multi-step recruitment approach for a probability-based online panel and the potential selection steps is illustrated are Figure 1.1.

*Figure 1.1: Schematic recruitment process and selection steps for a probability-based online panel*
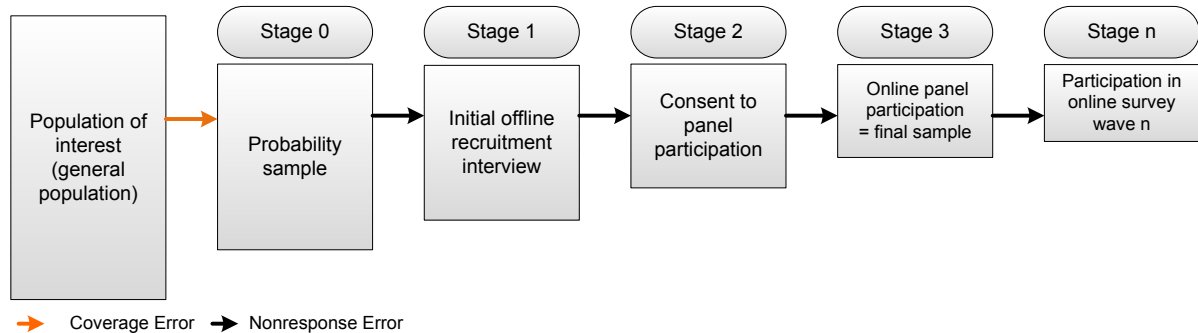


Nonresponse at the selection step from the sample to the interview (stage 0 → stage 1) occurs either because respondents cannot be contacted or because they are not willing to participate in a personal recruitment interview, a problem that is not unique to probability-based online panels. Given a successful interview (stage 1), respondents who were asked for further panel participation may either provide consent (stage 2) or refuse to participate. At the transition from expressed consent (stage 2) to survey participation (stage 3), respondents who consented can decide whether to become an active panel member or not. As an active panel member (stage 3), respondents have to decide about participation each time they are invited to a survey (stage n).

The determinants of survey participation (stage 0 → stage 1) in general have received extensive scholarly attention (see for example Groves & Couper, 1998; Lynn, 2008; Schnell, 1997). For probability-based online panels, however, selection steps 2 (stage 1 → stage 2) and 3 (stage 2 → stage 3) are critical for success (Callegaro, Baker, et al., 2014). The success of the recruitment process is crucial to the success of the panel operation. However, research on nonresponse and selectivity in probability-based online panels is still scarce (Bosnjak et al., 2016).

When studying nonresponse and selectivity of probability-based online panels, the focus can be either on the overall evaluation of the resulting sample or on factors of the survey design that affect response on the various selection steps.

To evaluate the overall outcome of the recruitment process, the extent of bias in the resulting sample is of importance. These are the magnitude of coverage bias and the extend of nonresponse bias (e.g., Biemer & Lyberg, 2010; Groves 2004). Coverage bias is a particular concern in online surveys as parts of the general population cannot participate in online surveys because of missing internet access and competencies (Couper, 2000; Eckman, 2016). However, probability-based online panels cope with

this problem by recruiting respondents offline and by including non-internet users (Callegaro & DiSogra, 2008).

The issue of nonresponse and the risk of nonresponse bias are of particular importance for probability-based online panels. Nonresponse bias is defined as the product of the response rate and the discrepancy between the survey estimates of respondents and nonrespondents[1] (Biemer, 2010a). Nonresponse bias in probability-based online panels was addressed by several studies in recent years (Blom, Herzing, et al., 2016; Bosnjak et al., 2017; Chang & Krosnick, 2009; Knoef & Vos, 2008; Revilla & Saris, 2010; Struminskaya, Kaczmirek, Schaurer, & Bandilla, 2014; Yeager et al., 2011). The studies show that the sample composition and the quality of estimates reach a good representation compared to benchmark surveys. However, the sample composition differs on some indicators from the benchmark surveys. For example, elderly respondents are often underrepresented (Bandilla, Kaczmirek, Blohm, & Neubarth, 2009; Blom, Herzing, et al., 2016; Knoef & Vos, 2008), and respondents with a high level of education are overrepresented (Blom, Herzing, et al., 2016; Bosnjak et al., 2017). In sum, the cited studies focused on aspects of the quality of the resulting sample and not on factors that affect nonresponse during the different stages of the recruitment process. The rationale behind design decisions about specific features of the recruitment process is often based on best-practice knowledge (Bosnjak et al., 2016), but rarely on experimental data.

Regarding factors of the survey design, the main focus of experiments on panel recruitment has so far been on the recruitment mode (e.g., Rao, Kaminska, & McCutcheon, 2010; Scherpenzeel & Toepoel, 2012), the amount and timing of incentive at the different nonresponse stages (e.g., Blom et al., 2015; Scherpenzeel & Toepoel, 2012), and on the logistics around the recruitment interview, for instance, the use and timing of a reminder letter (e.g., Martinsson & Riedel, 2015; Rao et al., 2010). To date, there is a lack of experimental evidence on the design of the recruitment interview and its effects on panel quality across the various recruitment stages.

In this dissertation, I aim to fill the gap and contribute experimental evidence about the recruitment process of a probability-based online panel. The overall objective is to identify ways to optimize the telephone recruitment process of a probability-based online panel in Germany and derive practical recommendations. Referring to the Total Survey Error perspective (Groves & Lyberg, 2010) optimal is defined in the sense of maximizing the recruitment probability and online participation probability and minimizing the selection bias under given budget constraints.

Selected features of the recruitment process were experimentally varied during the telephone recruitment process of a probability-based online panel in Germany. The features in question were selected for theoretical reasons based on the framework of survey participation by Groves & Couper (1998): they are the length of the recruitment interview, the inclusion of a sensitive question in the recruitment interview, and the amount of promised incentive for subsequent online panel participation.

---

[1]Nonresponse bias $\hat{B}$ is defined as: $\hat{B} = \hat{\gamma}_{NR} (\bar{Y}_R - \bar{Y}_{NR})$ with $\hat{\gamma}_{NR}$ being the amount of nonresponse, $\bar{Y}_R$ the average survey estimate of the respondents and $\bar{Y}_{NR}$ the average survey estimate of those who do not respond (Biemer, 2010a, p. 841).

To assess the quality of the recruitment process, I use (1) the proportion of respondents that provided consent to participate in the online panel, and (2) the proportion of respondents that participated in the panel. Furthermore, I examine whether the experimental treatments affect the composition of the respondents in the different treatment groups at the stage of recruitment and the stage of online participation. For this purpose, I use (3) the selection bias that is introduced by experimental variation as a third quality criterion.

The empirical analyses in this dissertation are based on the GESIS Online Panel Pilot. The pilot project started in 2010 and ended in 2012 and was hosted by GESIS - Leibniz Institute of the Social Sciences. The nature of a pilot study gave the unique opportunity to implement various experiments during the recruitment process as well as during the panel operation (Struminskaya et al., 2014; Struminskaya, 2014, 2016).

The remainder of this thesis is arranged as follows. In the first part of Chapter 2, I present the framework of survey participation (Groves & Couper, 1998) that provides the theoretical basis and discuss further theories of survey participation. This part is followed by an overview of probability-based online panel projects and a review of what is known in the literature on probability-based online panel recruitment and ends with the formulation of the leading research questions. In Chapter 3, I introduce the GESIS Online Panel Pilot project, describe the database that is used throughout the dissertation, and define the central dependent variables of the analyses. Furthermore, the preparation of the data set and the strategy of dealing with missing values are described. The subsequent chapters explore the effect of variations in the recruitment process on the overall success of the recruitment process. Chapter 4 looks at the effect of the length of the recruitment interview. Chapter 5 analyzes the effect of asking about income - assumed to be a sensitive question - in the recruitment survey. Experimental variation of the amount of promised monetary incentive for panel participation is discussed in Chapter 6. In the final empirical Chapter 7, I additionally examine the role of interviewers in carrying out the experiments. I quantify the amount of interviewer variance on the recruitment, identify determinants of interviewer effects, and analyze whether different interviewers cope differently with the varying experimental treatments. Chapter 8, closes with a summary of the findings of this dissertation, discusses limitations and implications of the work and sketches paths of future research.

# 2. Conceptual framework, literature review, and research gaps

In the following chapter, I present the framework of survey participation (Groves & Couper, 1998) that provides the theoretical basis of this work, discuss it in the light of additional theories of survey participation, and expand it to the multi-step recruitment process. This part is followed by an overview of probability-based online panel projects and a review of what is known in the literature on probability-based online panel recruitment and ends with the formulation of the leading research questions.

## 2.1 Framework of survey participation

In the framework of Total Survey Error "the goal of optimal survey design, […] is to minimize the total survey error […] within the constraints impose by survey resources" (Biemer, 2010b, p. 30). The primary error sources during the recruitment for a probability-based online panel are coverage and nonresponse error. This dissertation focuses on the nonresponse aspect.

When designing the recruitment process, researchers aim at minimizing nonresponse at the various stages. Nonresponse can occur at the stage of the recruitment interview, the stage of giving consent, and at the stage of online survey participation. The decision to consent to panel participation represents only an intention of behavior, whereas the decision of participation in the recruitment interview and the subsequent self-administered surveys represents actual behavior. The three selection steps encompass those phases of the recruitment process where researchers have to make decisions that are assumed to influence the final success of the recruitment in terms of the cumulative response rate (Callegaro & DiSogra, 2008). The conceptual framework that allows for the inclusion of all the factors that are relevant for participation at each stage of the recruitment process is that of Groves & Couper (1998). The framework of survey participation was developed in the context of household surveys, but it is transferrable to other survey contexts. The authors distinguish, first, factors that are out of researchers' control, and second, factors that are under researchers' control (see Figure 2.1).

*Figure 2.1: Conceptual framework of survey participation*
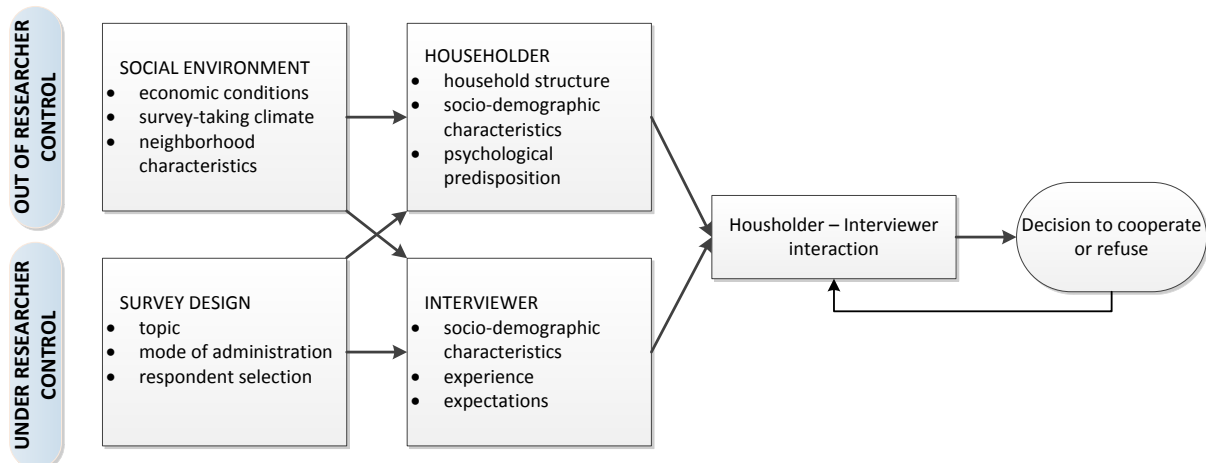


*Note.* Conceptual framework of survey participation, slightly adapted from Groves, R. M., & Couper, M. P. (1998). Nonresponse in household interview surveys. New York: Wiley. p. 120.

As factors that are out of researchers' control, they identify 1) the "social environment," for instance, economic conditions, survey-taking climate, and neighborhood characteristics. These characteristics are assumed to be stable for a single survey; and 2) characteristics of the household/ sample person, for instance, household structure, socio-demographic characteristics, and psychological predisposition. These characteristics are included in models that aim to explain nonresponse and cooperation post hoc or models that are the base for a responsive design strategy (see for example Groves & Heeringa, 2006). A large number of articles identify characteristics that are associated with survey nonresponse for every survey mode (for web surveys, e.g., Vehovar et al. (2002); for CAPI, e.g., Goyder (1987), Olson (2006) and Schnell (1997); for mail surveys, e.g., Moore & Tarnai (2008) and Heberlein & Baumgartner (1978); telephone surveys, e.g., Groves & Lyberg (2001)).

Groves and Couper (1998) state that the identified characteristics are not causal to the decision of survey participation, but that they "tend to produce a set of psychological predispositions that affect the decision" (p. 32). The authors assume that most persons who are faced with a survey request do not have a strong opinion about survey participation. Potential respondents rely on a heuristic decision-making process. This assumption is in line with several other authors (Dillman, 1978; Esser, 1986; Groves, Singer, & Corning, 2000; Singer, 2011).

The second set of factors, those that are under researchers' control, is of interest for the design and implementation of the recruitment process of a probability-based online panel. Groves and Couper include two factors into their framework: the survey design and the interviewers. Examples of the survey design are, for instance, topic, mode of administration, and respondent selection. According to the authors, several features of the survey design that aim to maximize cooperation make use of principals of compliance heuristics that apply for survey participation as well (Cialdini, 1988; Groves, Cialdini, & Couper, 1992).

The six principals are reciprocation, authority, consistency, scarcity, social validation, and liking. Incentives, for example, make use of the heuristic of reciprocation. Evoking authority is a measure that

is often used in advance letters, when referring to a scientific sponsor, for instance. The principal of scarcity is used when the privilege of being chosen and the possibility to make "your" voice heard is emphasized.

The second factor that is under the control of the researcher is the interviewer. Interviewers differ, for instance, in their socio-demographic characteristics, their experience, and their expectations about potential success. Researchers have control over the recruitment, the training process, and the supervision during the field (Groves & McGonagle, 2001).

Groves et al. (2000) expand on the heuristic framework of survey participation in the leverage-salience theory. Accordingly, the influence of a survey feature depends on its importance for the individual respondent (leverage) and its prominence during the survey request (salience). The benefit-cost theory that is also based on cost-benefit assumptions emphasizes the multiplicity of factors that are assumed to influence the decision to participate in a survey represents (Singer, 2011). It "emphasizes the role of benefits in the decision-making process and the need for making these benefits salient to the respondent and achieving a favorable ratio of benefits to costs" (Singer, 2011, p. 388). In contrast, Dillmans' (1978) Tailored Design Method is not based on a rational behavior model. It is not primarily aimed at explaining survey participation, but at guiding the optimal design of a survey by a theory-based approach. He applies the social exchange theory to survey request and tries to connect several features of the survey design to maximize response (Dillman, Smyth, & Christian, 2014). Central concepts of the social exchange theory are reciprocity, trust, and altruism (Dillman et al., 2014, p. 25). The authors state that establishing trust in the survey organization/ interviewer represents the base of survey participation. Based on the primary aim of establishing trust, the authors derive recommendations for the survey design that minimize costs of survey participation and maximize benefits.

All of the summarized concepts that aim to explain the decision of survey participation have in common that they assume the decision of potential respondents is made quickly and is based on heuristics rather than based on an explicit cost-benefit calculation. A second commonality is the assumption of a multitude of factors, which affect the decision. Based on the knowledge of the heuristic decision making process several practical recommendations are derived for survey researchers. These are stressing the survey features that are assumed to be of benefit for the respondents and make them more prominent than the features that are expected to be burdensome. Burden is associated with interview length (Dillman et al., 2014) and questions that are sensitive or inconvenient for the respondent (Bogen, 1996; Bradburn, 1978; Tourangeau & Yan, 2007).

## 2.2 Proposed framework for multi-step recruitment

For each of the selection steps of the recruitment process for probability-based online panels, there are factors of the survey design that influence the decision to participate. The conceptual framework of survey participation is extended to the multi-step recruitment process. Figure 2.2 summarizes the assumed factors across three stages of the recruitment process.

### 2.2.1 Factors of the survey design

Stage 1 concerns the decision of the respondent to participation in the recruitment interview. It resembles the heuristic model of Groves & Couper (1998) that is described above for cross-sectional surveys. One factor of the survey design is the mode of administration. Depending on the recruitment protocol of the particular study, the mode of administration is constant for every respondent, or several modes are offered if a mixed-mode strategy is chosen. I will illustrate that the mode choice for the recruitment interview is a central element that differs from panel to panel (Tourangeau et al., 2013). The topic of the recruitment interview survey should be of general interest (Dillman et al., 2014) and not introduce topic specific bias (Groves, Presser, & Dipko, 2004; Tourangeau, Groves, Kennedy, & Yan, 2009).

The two crucial selection steps are stage 2, which encompasses the respondents' decision to consent to join the panel and stage 3, which includes the respondents' decision to participate in the surveys, given consent. These two steps are critical for the success of the panel (van der Laan, 2009). The survey design factors are identical for stage 2 and stage 3. Longer interviews are associated with higher burden (Bradburn, 1978; Dillman et al., 2014, p. 32). The implicit assumption is that, with increased length of an interview, the burden for participation increases and the participation rate decreases (Fricker et al., 2012). Thus, the recommendation is to keep the interview short. Bradburn (1978) identified the amount of experienced stress for the respondent as an additional factor that is related to burden. Hence, the interview should not contain questions that psychologically threaten the respondent (Tourangeau & Yan, 2007), for instance, questions that are defined as being sensitive (Tourangeau, Rips, & Rasinski, 2000).

An almost universal measure to increase response rates is the use of cash incentives (Singer, Van Hoewyk, Gebler, Raghunathan, & McGonagle, 1999). It is argued that incentives are an attribute that has a positive leverage for most respondents and thus can compensate for attributes with negative leverage (Groves et al., 2000). Varying incentives schemes are used during the recruitment process for probability-based online panels (Blom, Bosnjak, et al., 2016).

When designing a recruitment interview, the aim is to design it in such a way that, for the majority of respondents, the positive aspects of the overall survey experience outweigh the negative aspects.

Selection step three, the decision to participate in the online surveys, mark the mode switch from an interviewer-administered mode to the self-administered mode that is a central characteristic of the recruitment process of probability-based online panels. Lepkowski & Couper (2001) argue that the

decision to participate in a next wave is dependent on the survey experience of the interview of the previous wave. Studies on longitudinal surveys confirmed this assumption and showed that the overall experience of the previous wave is a good predictor for next-wave participation (Laurie, Smith, & Scott, 1999; Lepkowski & Couper, 2001; Olsen, 2005; Watson & Wooden, 2009). Also for panel recruitment, Bartsch (2012) showed that respondents that assessed the recruitment interview as a positive experience had a higher recruitment probability.

### 2.2.2 Interviewers

In addition to the survey design, Groves & Couper (1998) identify the interviewer as a second factor that is under the control of researchers. Depending on the setting, researchers can influence the recruitment of the interviewer and their general and specific training (Groves & McGonagle, 2001). Interviewers have several attributes that are advantageous or disadvantageous for gaining cooperation (Schaeffer, Dykema, & Maynard, 2010). Additionally, interviewers differ in their prior work-related experience. It is argued that more experienced interviewers are more successful in gaining cooperation (Groves & Couper, 1998; West & Blom, 2016). The third set of attributes is the expectations of the interviewer and their trust in their success (Blom, de Leeuw, & Hox, 2011; de Leeuw, 1999; Durrant, Groves, Staetsky, & Steele, 2010; Groves & Couper, 1998). Interviewers play a crucial role in gaining cooperation (Durrant et al., 2010) for the recruitment interview, as well as in convincing the respondents to consent to subsequent panel participation.

In the third selection step, the interviewers are no longer assumed to play a central role. Features of the web survey design (Callegaro, Lozar Manfreda, & Vehovar, 2015; Tourangeau et al., 2013), for instance, the announced length and the announced topic (Marcus, Bosnjak, Lindner, Pilischenko, & Schütz, 2007), as well as technical aspects, are assumed to influence the decision to participate.

### 2.2.3 Interaction between interviewer and target person

Finally, the interaction between target person and interviewer is assumed to play a crucial role. Groves and Couper (1998) identify two constructs that affect the respondents' decision to cooperate. The concepts are tailoring and maintaining interaction. Tailoring is a strategy of the interviewers to adapt their behavior and their arguments based on assumptions about the target person. Interviewers that applied tailoring strategies were shown to be more successful in gaining cooperation in personal interviews (Groves & McGonagle, 2001; Morton-Williams, 1993; Snijkers, Hox, & de Leeuw, 1999). Maintaining interaction describes the strategy of interviewers to decrease the probability of refusal by keeping the conversation going. Groves and Couper (1998) suggest that the longer the short conversation of the initial contact lasts the more cues does the interviewer get for applying tailoring strategies (Groves et al., 1992, pp. 478–479). While gaining cooperation, interviewers who are able to use verbal interaction techniques were shown to be most successful (Morton-Williams 1993). The techniques are aimed at creating a comfortable atmosphere and positive relation between interviewer

and respondent. The efforts of the interviewers can be subsumed under the vague concept of building rapport (Goudy & Potter, 1975). However, there is no consent on defining rapport, and there is a multitude of operationalization (Garbarski, Schaeffer, & Dykema, 2016; West & Blom, 2016). Establishing trust is one facet of rapport (Garbarski et al., 2016). The central task for the interviewer during the recruitment interview is establishing trust in the survey organization in order to maximize consent (Bartsch, 2012).

Regarding the decision to participate in an online survey, it is not the interviewer-respondent interaction that may have an influence, but the survey organization- respondent interaction that may influence the decision to participate in the survey (Dillman et al., 2014).

*Figure 2.2: Framework of survey participation adapted to a multi-step approach*

**Social environment**

| Stage 1: Recruitment interview | Stage 2: Consent to panel participation | Stage 3: Panel participation |

**Stage 1: Recruitment interview**

SURVEY DESIGN
- mode of administration
- topic
- respondent selection
- survey burden (announced length of interview)

INTERVIEWER
- socio-demographic characteristics
- experience
- expectations

TARGET PERSON - INTERVIEWER INTERACTION

Decision to participate or refuse participation in the **recruitment interview**

**Stage 2: Consent to panel participation**

SURVEY DESIGN
- topic
- survey enjoyment
- **length of interview**
- **sensitive questions**
- **incentives**

INTERVIEWER
- socio-demographic characteristics
- experience
- expectations

RESPONDENT - INTERVIEWER INTERACTION

Decision to give or reject consent to **panel participation**

**Stage 3: Panel participation**

SURVEY DESIGN
- survey experience
- recruitment interview
- visual design
- technical aspects
- invitation
- **incentives**

RESPONDENT – SURVEY ORGANISATION INTERACTION

Decision to participate or refuse participation in online surveys

In the next section, I provide a systematic overview of the available evidence on the recruitment of probability-bases online panels. First, I summarize the descriptive overviews of how the various panels were built, summarize correlational findings on the determinants of panel recruitment, and close with an overview of results that are based on experimental designs.

## 2.3 Evidence on probability-based online panel recruitment

The use of probability-based online panels as a data collection infrastructure has become popular for several reasons. For external researchers, data collection using a probability-based online panel is cost-effective because they save the fieldwork costs for personal or telephone interviews (Callegaro, Baker, et al., 2014; Lynn, 2013). Compared to more traditional methods of data collection, data are available very quickly (Scherpenzeel & Bethlehem, 2010), and researchers often have access to a large set of background variables which were collected in earlier waves (Lensvelt-Mulders et al., 2008).

The idea of collecting data by means of a probability-based self-administered panel without interviewers was first realized by the Dutch Telepanel in the mid-1980 (Saris, 1998). Throughout the three decades since then several panels that predominantly use the internet for data collection mode have emerged in the US and Europe (Bosnjak et al., 2016; Hays et al., 2015). All of these panels have in common that they are based on a probability-based sampling method that results in a known inclusion probability for each respondent (Couper, Kapteyn, Schonlau, & Winter, 2007). Being representative of the general population is the aim of those projects–just like traditional surveys that draw one sample each time. However, as Blom, Bosnjak, et al. (2016) point out appropriately "many roads can lead to a high-quality probability online and mixed-mode panel covering the general population." That means, there are several design decisions researchers are faced with when setting up a panel and that, in consequence, shape the recruitment process of the individual panel.

In the next section, I give a broad overview of the various recruitment protocols of existing probability-based online panels.

### 2.3.1 Taxonomy of probability-based online panels

To systematically summarize the multiple paths that lead to a scientific probability-based online panel, Table 2.1 provides a comprehensive overview of ongoing and visible panel projects and sketches the design features of the recruitment process. The list is organized by the founding year of the panel.

First, panels can be distinguished by the mode of recruitment, and, second, by the decision how to cope with the fact that not the whole population uses the internet (DiSogra & Callegaro, 2009). The decision about the recruitment mode depends on country-specific factors and budget constraints of the projects. Researchers chose between postal (Swedish Citizen Panel, Martinsson & Riedel 2015; Norwegian Citizen Panel, Hogestol & Skjervheim, 2013; Understanding America Study,(Understanding America Study, 2017)), telephone (SSRI Online Panel, Jónsdóttir, 2015), and

personal (German Internet Panel (GIP), Blom et al., 2015; GESIS Panel, Bosnjak et al.) recruitment. The modes might also be combined, either sequentially or simultaneously (LISS Panel, Scherpenzeel, Das, Kapteyn, & Soest, 2008; Gallup Panel, Rookey, Hanway, & Dillman, 2008; ELIPSS, Revilla et al., 2015; AmeriSpeak Panel, (Dennis, 2017)). A minority of the panels do not implement a specific recruitment interview but use established surveys of the general population for asking the respondents at the end about panel participation (e.g., American Life Panel, (RAND, 2017); Gallup Panel).

The second design decision on how to integrate non-internet users is related to the trade-off between two sources of survey error: coverage error and measurement error. The decrease of one of the error sources is accompanied by an increase in the other one (Blom, Bosnjak, et al., 2016). Panel providers that focus on a maximum of measurement equivalence provide all respondents with a device and internet access, nowadays with a tablet computer (e.g., ELIPSS, Revilla et al., 2015). This strategy allows researchers to assure the same visual appearance of the survey for all respondents (Callegaro, Baker, et al., 2014). It is assumed that no additional measurement error is introduced in that case. The more cost-effective alternative is to provide devices and access only to those respondents that do not use the internet (e.g., LISS Panel, GIP, American Life Panel). These measures were shown to increase the sample representativeness (Blom, Herzing, et al., 2016; Scherpenzeel & Bethlehem, 2010), even though the consent rate among previous non-internet users is lower compared to internet users (Blom, Herzing, et al., 2016; Leenheer & Scherpenzeel, 2013; Revilla et al., 2015)

Thus, coverage can be maximized by providing an alternative offline mode for the respondents who are not able or willing to participate online (Dillman, Reips, & Matzat, 2010). Panels that mainly focus on maximal coverage solve the problem of non-internet users, not by providing them with internet access, but by offering them a second mode of data collection that is more convenient for the respondents (Gallup Panel, GESIS Panel). This approach broadens the classification of probability-based online panels to mixed-mode panel infrastructures of the general population. A comprehensive overview of all potential paths of recruitment is given by DiSogra and Callegaro (2016, p. 31). Lately, panel providers in countries with very high internet penetration do not need to implement any measures, as, for example in Iceland (Jónsdóttir, 2015), Norway (Hogestol & Skjervheim, 2013) and Sweden (Martinsson & Riedel, 2015).

Besides the initial decisions about the recruitment mode and how to solve the problem of coverage, different panels reflect various further design decisions that shape the individual recruitment process (DiSogra & Callegaro, 2016).

The sampling procedure for the panels reflects national circumstances and traditions. In the US, samples are based on an RDD-frame or address-based sampling. In Europe, all samples are based on population registers. Depending on the recruitment mode, several measures are used additionally to maximize the return from the recruitment process. Response enhancing measures are, for instance, advance letters (Gallup Panel, LISS Panel), reminder letters, additional calls or emails, and incentives for participation in the recruitment interview and registration. The three panels that use postal

recruitment (Swedish and Norwegian Citizen Panel, AmeriSpeak Panel) send invitation postcards and several reminders. Among the four European panels that recruit via a personal interview, only the LISS Panel send out an advance letter, the other three supply project information through the interviewer.

The Gallup panel represents an exception as it does not provide any incentives to respondents (Rao et al., 2010). The majority of the European panels provide cash incentives for the recruitment interview and the registration. Incentives are either prepaid (ELLIPS, GIP as an experiment, LISS) or conditional on participation (GESIS Panel, GIP as an experiment). The Swedish (Martinsson & Riedel, 2015) and the Norwegian (Hogestol & Skjervheim, 2013) Citizen Panel, both of which use the postal recruitment approach, use lotteries that are contingent on registration as incentives (see Chapter 6 for a comprehensive overview of experimental evidence on incentives).

The majority of panels administer regular survey waves online. Five panels provide at least one alternative mode: the Gallup Panel and the American Trends Panel (Pew Research Center, 2017) use mail and telephone, the GESIS Panel uses the mail mode and the AmeriSpeak panel the telephone mode as an alternative. Besides the design factors, the table comprises key success indicators of the recruitment process. These are the recruitment rate and the registration rate. Unfortunately, indicators according to the AAPOR standard definition (American Association for Public Opinion Research, 2016) are not available for all of the panels. Due to differences in recruitment mode, sampling procedure, and other design characteristics, the response rates available are hardly comparable (Blom, Bosnjak, et al., 2016, p. 10). The panel registration rate (Callegaro & DiSogra, 2008) that is dependent on the response rate of the initial recruitment survey differs for the different modes and due to country-specific factors. Personal interviews reach higher response rates than telephone interviews and postal interviews (Blom, Bosnjak, et al., 2016, p. 8). Panel registration rates on the household level are reported by the LISS Panel (AAPOR RR 3: 48.7%) and the AmeriSpeak Panel (AAPOR RR 3: 48.7%). The rates for those panels that apply an interviewer-administered recruitment on the person level are lower: German Internet Panel (AAPOR RR 4: 18.1%), ELIPSS pilot study (AAPOR RR 3: 27.3%), and GESIS Panel (AAPOR RR 5: 25.1%). Finally, the cumulative response rate 1 that is reported for the Swedish Citizen Panel that applied a postal recruitment is between 5.8% and 14.2% depending on different experimental variations.

Since the primary goal of the expensive recruitment process is to maximize the number of respondents that become active panel members, experimental variation and validation of features of the recruitment process have hitherto been scarce. Design decisions, therefore, are often based on best practices (Tourangeau, 2004) – and not on empirical evidence. In the following sections, I give an overview of past findings of the recruitment for probability-based online panels. First, I summarize correlational findings, and, second, I summarize the evidence that is derived from experimental research. The overview of the previous research on the specific topic of the empirical chapters is summarized at the beginning of the respective chapter.

*Table 2.1: Probability-based online panels – Central characteristics*

| | Country | Start | Organization | Type of sample | Recruitment mode | Survey mode | Coping with coverage problem | Recruitment response rates |
|---|---|---|---|---|---|---|---|---|
| GfK KnowledgePanel[1] | USA | 1999 | GfK group, funded by Knowledge Networks | List-based RDD sampling, supplemented by address-based sampling | Telephone | Online | Provide all panelists with a device (Web TV) and access | Completion rate 65% |
| American Life Panel[2] | USA | 2003 | RAND | Recruitment from several sources, including representative samples of U.S. consumers RDD, address-based sampling | Telephone, Face-to-face | Online | Provide Non-Internet users with a device and access | Completion rate approx. 70% |
| Gallup Panel[3] | USA | 2004 | Gallup | List-based RDD sampling and address-based sampling | Postal mail, telephone | Online, Paper, Telephone | Including additional modes | Recruitment rate between 25.6% and 37.7% (26% AAPOR RR3) Registration rate between 6.7% and 14.5% |
| LISS Panel[4] | Netherlands | 2007 | University of Tilburg | Household sample from register | Telephone, face-to-face (offered sequentially) | Online | Provide non-internet users with a device and access | Recruitment rate (household): 73.2% (AAPOR RR 2) Panel registration rate (household): 48.3% (AAPOR RR 3) |
| Social Science Research Institute (SSRI) Online Panel[5] | Iceland | 2010 | University of Reykjavik | Sample of individuals from register | Telephone | Online | No measure due to high internet penetration | Consent rate 50.3% |

| | Country | Start | Organization | Type of sample | Recruitment mode | Survey mode | Coping with coverage problem | Recruitment response rates |
|---|---|---|---|---|---|---|---|---|
| German Internet Panel (GIP)[6] | Germany | 2012 | University of Mannheim | Sample of individuals within households from register | Face-to-face | Online | Provide non-internet users with a device and access | Recruitment rate (person): 52.1% (AAPOR RR2) Registration rate (person): 18.1% (AAPOR RR4) |
| ELIPSS panel[7] | France | 2012 | Science Po, Paris | Sample of Individuals from register | Postal mail, telephone, and face-to-face (offered sequentially) | Online | Provide all panelists with a tablet and access | Recruitment rate (person: 31.3% (AAPOR RR3) Registration rate (person): 27.3% (AAPOR RR3) |
| Swedish Citizen Panel[8] | Sweden | 2012 | University of Gothenburg | Sample of Individuals from register | Postal mail | Online | No measure due to high internet penetration | Recruitment rate (person) between 14.2% and 5.8% (Cumulative response rate 1) |
| GESIS Panel[9] | Germany | 2013 | GESIS - Leibniz-Institute for the Social Sciences | Sample of Individuals from register | Face-to-face | Online and Paper | Including the paper mode | Recruitment rate (person): 38.6% (AAPOR RR5) Registration rate (person): 25.1% (AAPOR RR5) |
| Norwegian Citizen Panel[10] | Norway | 2013 | University of Bergen | Sample of Individuals from register | Postal mail | Online | No measure due to high internet penetration | na |
| AmeriSpeak Panel[11] | USA | 2014 | NORC - University of Chicago | Household sample from NORC National Frame (an area probability sample frame constructed by NORC) | Mail, telephone, Email; Face-to-face nonresponse follow-up | Online, Telephone | Including telephone mode | Panel registration rate (household): 36.6% (AAPOR RR 3) |

|  | Country | Start | Organization | Type of sample | Recruitment mode | Survey mode | Coping with coverage problem | Recruitment response rates |
|---|---|---|---|---|---|---|---|---|
| Understanding America Study[12] | USA | 2014 | University of Southern California | Household sample based on address-based sampling | Mail survey | Online | Provide Non-Internet users with a device and access | na |
| American Trends Panel[13] | USA | 2014 | Pew Research Center | Individuals living in households (RDD sample) from the Political Polarization and Typology Survey | Including telephone and mail mode | Online and Paper | Mixed-mode strategy | Panel registration rate: 43% |

*Note.* Web sources of the panels: [1] http://www.gfk.com/products-a-z/us/knowledgepanel-united-states/ [2] https://alpdata.rand.org
[3] http://www.gallup.com/174158/gallup-panel-methodology.aspx [4] https://www.lissdata.nl/lissdata/Home [5] http://fel.hi.is/online_panel
[6] http://reforms.uni-mannheim.de/internet_panel/home/ [7] http://quanti.dime-shs.sciences-po.fr/en/pilot-study/ [8] http://lore.gu.se/surveys/citizen
[9] http://www.gesis.org/en/gesis-panel/gesis-panel-home/ [10] http://www.uib.no/en/citizen [11] http://amerispeak.norc.org/about-amerispeak/
[12] https://uasdata.usc.edu/ [13] http://www.pewresearch.org/methodology/u-s-survey-research/american-trends-panel/ and http://www.pewresearch.org/2015/04/08/results-2/

### 2.3.2 Correlational findings

Several studies investigate the nonresponse steps of the recruitment process and describe which groups of persons participate and which do not (see for example Hoogendoorn & Daalmans, 2009, Knoef & Vos, 2008; Lee, 2006). Researchers mainly describe differences between respondents who consent to subsequent panel participation and those who do not. They mostly focus on demographic differences because demographic information is available for both respondents and nonrespondents, and in potential benchmark surveys as well (Callegaro, Villar, Yeager, & Krosnick, 2014). However, demographic characteristics are not seen as being the cause for giving consent but being an indicator for an underlying mechanism (Groves & Couper, 1998). Characteristics that are used beyond demographic variables are, for example, political interest (Martinsson & Riedel, 2015), vote intention (Martinsson & Riedel, 2015), and internet literacy (Bartsch, 2012).

Hoogendoorn & Daalmans (2009), for instance, examined the selectivity of the recruitment process for the CentERpanel, a probability-based panel that was recruited by phone in the Netherlands. The sample was based on households with listed telephone numbers. Respondents who did not use the internet were provided with the necessary equipment. The authors linked recruitment information to data from the population register. They show that households with older persons and with lower income had a lower participation probability on all selection steps (CATI response, consent to panel participation, selection, actual participation). They did not find any differences concerning ethnicity. The underrepresentation of older persons is almost universally reported for all panels under consideration. In the LISS panel, Knoef & Vos (2008) found the elderly, especially women, to be underrepresented. Furthermore, van der Laan (2009) shows that elder target persons were easier to contact and easier to persuade to participate in the recruitment interview. The elder respondents dropped out at the stage of giving consent to panel participation. For Germany Bandilla et al. (2009) also showed that in a follow-up study in addition to the personal interview of the German General Social Survey (ALLBUS) the younger respondents have a higher probability of giving their consent to further online participation and also for survey participation. The Swedish Citizen panel represents the only expectation, where elder respondents were overrepresented at the recruitment stage (Martinsson & Riedel, 2015). The result is, however, not replicated for the Norwegian Citizen Panel, that applied mail recruitment like the Swedish Citizen Panel did, under the same conditions of an internet penetration rate above 90% (Hogestol & Skjervheim, 2013). The overrepresentation of higher-educated people of all age groups is a second characteristic that is reported for the samples of the majority of panels (Bandilla et al., 2009; Bosnjak et al., 2017; Hogestol & Skjervheim, 2013; Knoef & Vos, 2008; Martinsson & Riedel, 2015). In the LISS panel, single-person households and immigrants from non-Western countries were underrepresented (Knoef & Vos, 2008), which is mainly attributable to problems during the contact phase (van der Laan, 2009).

Besides demographics, voters are overrepresented in contrast to non-voters in the LISS panel (Scherpenzeel & Bethlehem, 2010) and the Swedish Citizen Panel, as well as respondents that are interested in politics (Martinsson & Riedel, 2015).

The familiarity with the internet and computer usage seems to be a central predictor for willingness and participation. A lower participation probability of former non-internet users is reported by Blom et al. (2015), Hoogendoorn & Daalmans (2009), Leenheer & Scherpenzeel (2013) and Revilla et al. (2015) as well. For the GESIS Online Panel Pilot (GOPP) data that are the basis of this work Struminskaya et al. (2014) also found a higher recruitment probability for more frequent internet users (Struminskaya et al., 2014). In line with Bartsch (2012, p. 116), respondents that had earlier experience with online surveys had a higher recruitment probability as well (Struminskaya et al., 2014). Additionally, respondents that used the internet for several activities like shopping, online banking and listening to music had a higher probability of giving their consent (Bartsch, 2012).

Vehre (2011) revealed for the recruitment process of a probability-based mixed mode panel that those respondents that enjoyed the survey participation in general and those who attributed a value to surveys had a higher probability of consenting to panel participation. Vice versa, respondents that experienced surveys as being burdensome had a lower probability of consenting.

Furthermore, Bartsch (2012, p. 125) included income nonresponse in the recruitment interview as a proxy for the trust in the survey organization in the model. She reported a significant negative effect on the willingness to participate for respondents that refused to provide income information. She points out that it is impossible to assess whether asking about the income itself has a negative effect on the recruitment success or the fact that respondents who do not report their income have a lower probability of consenting in general without experimental variation.

When faced with the decision on how to design the recruitment process for a probability-based panel the summary at the beginning of the section defines the possible set of possibilities and the correlative results summarized in section 2.3.2 informs about which groups should be primarily addressed during the recruitment process. In the next section, I summarize studies that experimentally varied one or more features of the recruitment process for probability-based online panels.

### 2.3.3 Experimental findings

Experimental studies offer better leverage for identifying the effects of different recruitment modes. Studies have most frequently assessed the effect of the value and the timing of incentives (Blom et al., 2015; Rao et al., 2010; Scherpenzeel & Toepoel, 2012). Consistent with previous meta-analytic studies for different survey modes (Mercer, Caporaso, Cantor, & Townsend, 2015; Singer & Kulka, 2002), prepaid incentives are found to be more efficient in the context of panel recruitment as well (Krieger, 2016; Scherpenzeel & Toepoel, 2012). Monetary incentives were shown to have a positive effect on cooperation at different stages of the recruitment process. For the German Internet Panel, Krieger (2016) showed that including a five euro note in the first reminder letter significantly

increased the registration rate of the respondents that consent to panel participation in the recruitment interview. Furthermore, the incentive shortened the time needed for the registration process and proved to be cost-effective with regard to the total recruitment costs. Finally, different incentive conditions did not affect the sample composition.

Nonmonetary incentives, e.g., a lottery incentive that was included in the invitation and recruitment postcard had a positive effect on recruitment in the case of postal recruitment as well in the Swedish case (Martinsson & Riedel, 2015). A small piece of quality chocolate as a symbolic prepaid incentive did, however, not significantly increase the recruitment rate of a Danish panel (Hansen, 2007).

The effect of the mode of recruitment on the recruitment success was an additional feature that was experimentally tested (Hansen & Pedersen, 2011; Rao et al., 2010; Scherpenzeel & Toepoel, 2012). The results have to be interpreted in relation to the country-specific survey practices. For the Netherlands Scherpenzeel & Toepoel (2012) found that personal and telephone recruitment were equally effective in terms of recruitment success for households with listed phone numbers. Rao et al. (2010) revealed that, in the U.S. context, postal recruitment was superior in terms of response rates compared to CATI mode and also more cost effective. In contrast, Hansen & Pedersen (2011) found that the phone mode was more effective in comparison to postal and text message recruitment mode in Denmark.

Further studies investigated effects of the timing of the recruitment request, the effect of advance letters (Rao et al., 2010), and the amount of information given in the letter (Scherpenzeel & Toepoel, 2012). All of those had no effect on recruitment success in the LISS panel (Scherpenzeel & Toepoel, 2012). For the recruitment of the American Gallup Panel Rao et al. (2010) found a significant increase in the recruitment rates for respondents that received an advance letter and nonrespondents that were addressed by a telephone follow-up. Martinsson & Riedel (2015) found that sending a reminder postcard during recruitment for the Swedish Citizen Panel had a positive effect on the recruitment rate. Very few of the experimental studies focused on the resulting sample composition or the bias that was introduced by varying features of the recruitment process (Krieger, 2016; Rao et al., 2010). Krieger (2016) found no differences regarding the demographic composition between recruited respondents of the different incentive groups. Rao et al. (2010) also reported no differences between respondents that were recruited by either mail or phone mode. They revealed differences in the demographic composition between respondents that were recruited by mail or by phone. The share of minorities (non-whites and renters) was significantly higher when recruited by mail, compared to the phone mode.

Some survey experiments that varied the amount of incentives also controlled for the bias introduced. The assumption about the effect of different incentive amounts on sample composition is that incentives may motivate respondents that are underrepresented in general population surveys to participate. These are for instance minorities or low socio-economic status groups (Singer & Kulka, 2002) and respondents with a low interest in the topic of the survey (Groves et al., 2004). The question

about the resulting sample composition was investigated for interviewer-mediated personal surveys (Pforr et al., 2015; Singer & Kulka, 2002; Singer et al., 1999), telephone surveys (Cantor, O'Hare, & O'Connor, 2008; Laurie & Lynn, 2009; Lipps, 2010a; Singer, Hoewyk, & Maher, 2000), as well as for online surveys (Göritz, 2004a; Marcus, Bosnjak, Lindner, Pilischenko, & Schütz, 2007). In sum, the results are not conclusive. In a review, Singer & Kulka (2002) cited few experimental studies that showed incentives being able to bring groups into the sample in interviewer-mediated surveys which are usually underrepresented in surveys. For telephone surveys, incentive effects on sample composition were not shown to be systematic as well (e.g., Cantor et al., 2008; Lipps, 2010).

## 2.4   Research gap and research questions

The overview of the recruitment processes has revealed that existing probability-based online panels exhibit quite some variation. While some features of the survey design are influenced by country-specific survey traditions and constraints (e.g. the sampling strategy), others depend on decisions about the main aim of the project (e.g. the way to include non-internet users).

As the summary of the literature showed, however, there is only limited experimental evidence that would allow gauging the effects of different design choices. The main focus of experiments on panel recruitment has so far been on the recruitment mode, the amount and timing of incentive at the different nonresponse stages, and on the logistics around the recruitment interview, e.g. the use and timing of a reminder letter. Questions about the design of the recruitment interview were not yet experimentally addressed.

It is, therefore, fair to say, that design choices are rarely based on experimental data. Most of the projects, furthermore, do not have the capacity to perform a large-scale pretest to assess different features of the recruitment process experimentally, the LISS panel (Scherpenzeel & Toepoel, 2012) and the ELIPPS pilot (Cornilleau, Cousteaux, Legleye, & Razakamanana, 2015) being notable exceptions.

Up to now, very few studies have investigated the relationship between experimental variation and resulting sample composition in terms of introduced bias. The majority of studies that did so did not reveal major differences in the sample composition of the different treatment groups. The scarcity of evidence is rather remarkable, given that findings I presented in section 2.3.2 show that probability-based online panels experience bias: older and less educated target individuals have a lower probability of consenting to panel participation. Furthermore, familiarity with the internet and computer ownership are central predictors of subsequent online panel participation.

With this dissertation, I will contribute to the existing literature by testing selected features of the survey design of a recruitment interview experimentally. In addition, I investigate the role of the interviewer during the recruitment process. On the basis of the framework of survey participation (Groves & Couper, 1998), three factors of the survey design that are assumed to influence the probability of providing consent to panel participation were selected due to theoretical reasons. Two of

the selected factors of the survey design are assumed to influence the received burden and the overall survey experience: the length of the recruitment interview and the inclusion of a sensitive question about the household income. The third factor of the survey design, the amount of promised incentives for online survey participation, contributes to the perceived benefit of the survey.

The first set of research questions guide the analysis of the experiments (Chapters 4-6). They address three quality aspects of the recruitment process. The research questions specific to the particular study and the corresponding hypotheses are outlined in detail in the respective analyses chapters.

Research question 1:    Does the experimental variation of the three selected survey factors affect the recruitment probability?

I expect all of the three factors to influence the recruitment probability. All three are related to the decision of the respondent to consent to subsequent panel participation at the end of the telephone recruitment interview. In contrast, only the amount of promised incentive should additionally affect the subsequent decision to participate in the online survey after having given consent, as the promised incentive is directly related to the action of survey participation.

Research question 2:    Does the experimental variation of the amount of promised incentive affect the probability for online participation in a probability-based online panel?

Finally, the third research question focuses on the aspect of the resulting sample composition.

Research question 3:    Does an experimental variation of the selected survey design features introduce sample composition bias at the stage of recruitment and the stage of online participation?

The role of the interviewers and their ability to cope with the experimental variations is part of the analyses as well. In the final empirical Chapter 7, I examine the role of interviewers as an additional error source during the recruitment process. The amount of interviewer variance on recruitment is quantified, and factors on interviewer and respondent level are introduced to explain the variation in interviewer performance. Additionally, the effect of the two fieldwork agencies, one with the focus on academic, social research, and the other with the focus on market research is examined. Finally, I analyze whether different interviewers cope differently with the varying experimental treatments.

The central research questions that guide the analysis of the final chapter are as follows:

Research question 4:   To what extent can interviewer variance in the recruitment propensity be explained by characteristics of the respondents, characteristics of the interviewers, and by the interaction of the characteristics of the interviewer and respondent?

Finally, I analyze whether different interviewers cope differently with the varying experimental treatments. Research question 5 is formulated accordingly.

Research question 5:   Is the propensity of being recruited different across interviewers for respondents of the different experimental groups?

Figure 2.3 incorporates the factors into the schematic representation of the framework of survey participation (Groves & Couper, 1998).

*Figure 2.3: Framework of survey participation adapted to a multi-step approach*
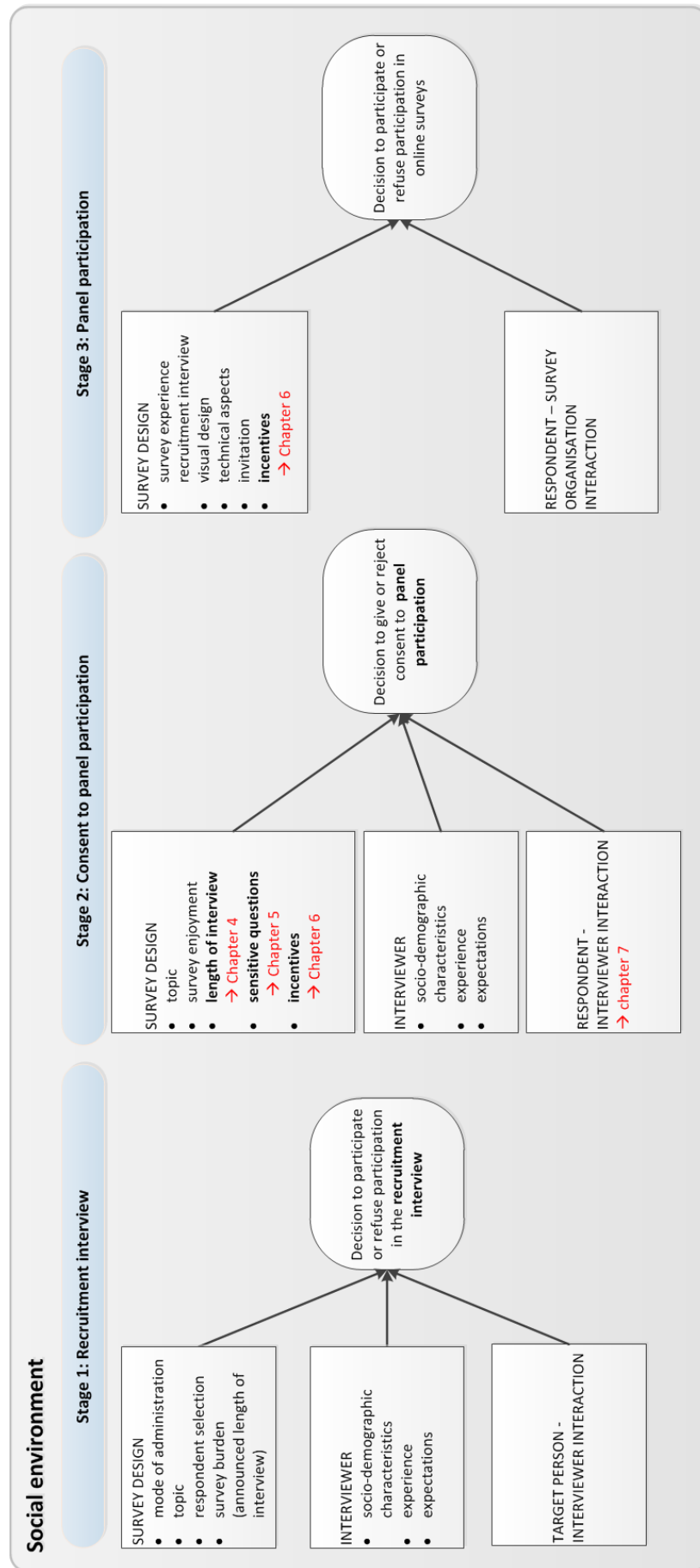
# 3. Data and method overview

In this chapter, I provide an overview on the recruitment process of the GESIS Online Panel Pilot as the central data source of the dissertation. In addition, I describe the procedure of dealing with missing values in section 3.2. The analyses of the three experiments have the identical structure and follow the logic of the multi-step recruitment process. For this reason, I define the central dependent variables in section 3.3. Finally, I briefly sketch the four studies of the dissertation in section 3.4 to provide the reader with an overview.

## 3.1 The GESIS Online Panel Pilot

### 3.1.1 Project overview

The GESIS Online Panel Pilot (GOPP) was a methodological project with the aim of developing best practices for the multi-step recruitment and maintenance of a probability-based online panel in Germany. The pilot project started in 2010 and ended in 2012 and was hosted by GESIS - Leibniz Institute of the Social Sciences.

Target frame are German-speaking, internet-using adults in Germany, aged 18 years and older.

Respondents were recruited offline into the panel via telephone using a probability sample that was based on an RDD sampling frame (Gabler, Häder, Lehnhoff, & Mardian, 2012). To take into account, the growing mobile-only population and the decreasing landline coverage, cell phone numbers were included in the sample. The dual frame approach was used, in which two samples were drawn independently from landline and cell phone frames under the assumption of independence of the frames (Gabler et al., 2012). Finally, it was aimed to have approximately 50% of eligible numbers from each frame. Within the landline frame, the last birthday method was used to ensure a random selection of the target person within a household. For cell phone numbers no further selection method was applied since cell phones are rarely shared in Germany (Gabler et al., 2012).

Respondents that did not use the internet were excluded and not provided with internet access due to budget constraints and the pilot nature of the project. At the end of the telephone interview, all internet-using respondents were asked to provide consent to subsequent online panel participation. Respondents that consented to panel participation and provided a valid email address were invited via email to eight monthly online surveys of durations between 10 and 15 minutes.

The project was executed in three sequential parts, each of those employed the strategies, which proved to be efficient during the previous part or corrected for deficiencies of the previous part. The placement of the internet screener question, for instance, was changed from the very beginning of the interview to the second place because the share of Internet users was assumed to be too low in study

part 1. The replacement successfully increased the proportion of Internet users[2]. The interviews were conducted by two data collection agencies in centralized CATI studios. The first one was an academic fieldwork agency, characterized by an interviewer pool that was specialized on scientific interviews within the social sciences and with a high amount of working experience (study part 1 and 2). The second one was a fieldwork agency with interviewers that had experience mainly in the field of market research interviews (study part 3).

Recruitment for the first part, preceded by a short technical pretest, started in February 2011 and continued until April 2011. The second and the third part were conducted partly parallel in summer 2011 by two fieldwork agencies. Up to 15 call attempts were made for a single number, in the second and third part of the recruitment it was reduced to 10 call attempts for numbers where no contact was made with a household or target person. In the case of personal contact, call attempts were retained at the minimum of 15.

The multi-step recruitment approach and the potential nonresponse steps in the GOPP are illustrated in Figure 3.1. Nonresponse at step 1 from sample to the interview can occur either because respondents cannot be contacted or because they are not willing or able to participate in the telephone interview (Dillman, Eltinge, Groves, & Little, 2002). Because only internet users were allowed to participate in the panel, non-internet users were screened out before the interview started. Those respondents that were interviewed could consent to participate in subsequent online surveys and provide their email address or refuse to participate in the panel (selection step 2). The respondents that gave their consent could participate in at least one online survey and become an active panel member (selection step 3). The main reason for nonparticipation at this step was nonresponse, only a few email addresses were invalid and could not be corrected even when respondents were contacted by phone again. In the following analysis, the focus is mainly on the effect of the experimental variations on the outcome at selection stage 2 and 3.

---

[2] In study part 1 the question about internet use was part of the screening procedure before the interview started. The analysis showed that as a consequence of this procedure the share of internet users was too low compared to external benchmarks. The low share gave hints that respondents used the screening question as a polite opt-out option. For this reason the position of the question about the internet usage was changed after a general question in study 2 and 3. This strategy seemed to have solved the problem resulting in a more realistic share of internet users (study part 2: 73%; study part 3: 77%).

*Figure 3.1: Selection process of the GESIS Online Panel Pilot (GOPP)*
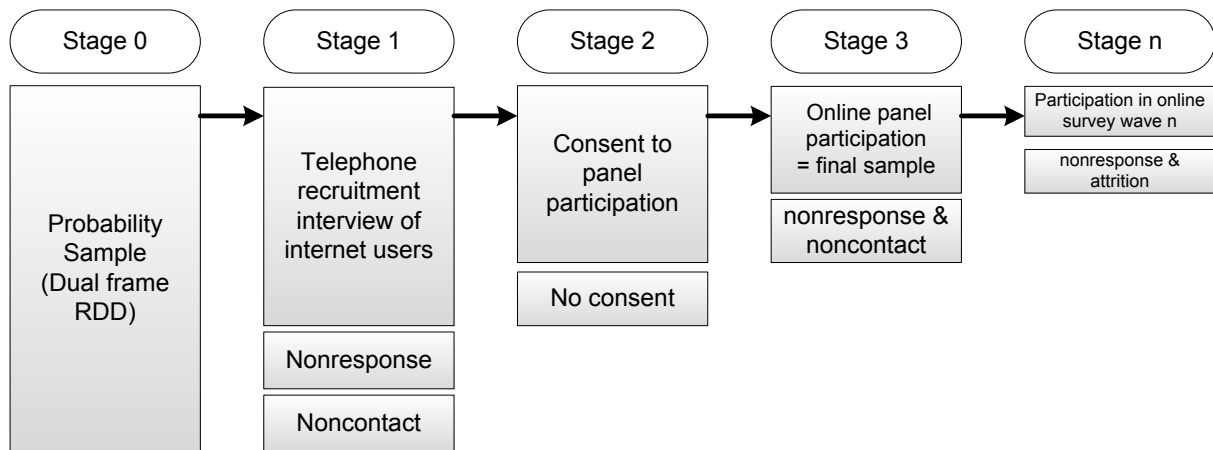


The GESIS Online Panel Pilot is a longitudinal panel in the sense that it collects data from the same sample units on multiple occasions over time (Lynn, 2009). Every monthly questionnaire had a leading topic (e.g., work or family) that was announced in the online invitation and included some demographical questions in order not to burden the respondent with all factual questions in a single survey. A small proportion of repeated measures were collected over the course of the panel (e.g., survey evaluation questions). Those measures were mainly included for methodological reasons rather than to study social change. For a detailed description of the online surveys see (Struminskaya et al., 2014; Struminskaya, 2016).

The online surveys for all three study parts were conceptualized, implemented and managed by the project team at GESIS – Leibniz Institute for the Social Sciences.

### 3.1.2 Telephone recruitment interview

Overall 3422 interviews with respondents that use the internet were conducted, resulting in an overall AAPOR Response Rate 3 (American Association for Public Opinion Research, 2016) of 17.84%. Of those who completed the recruitment interview about 53% consented to subsequent panel participation. Overall 47% of those that consented participated in at least one online survey. Since no profile survey was conducted (DiSogra & Callegaro, 2016) to collect key demographic variables, respondents that participated at least in one online survey were defined as being active panel members. Table 3.1 gives a brief overview of the key numbers separated by studies. The AAPOR Response Rate 3 for the telephone recruitment interview was significantly lower for the market research agency. The recruitment rate is defined as the proportion of respondents that gave consent to panel participation in relation to the number of completed telephone interviews. It was about ten percentage points higher in study part 1, compared to study part 2 and 3. The higher recruitment rate is most likely a result of the higher incentive amounts that were promised in study part 1 compared to the other study parts. The analysis of the experiment on incentives is part of Chapter 6. The online participation rate is defined as

the share of respondents that participated in at least one online survey in relation to the respondents that provided their consent to subsequent panel participation. It was also highest in study part 1 with about 20%-points difference compared to study part 2 and almost 30%-points difference compared to study part 3. The difference in the performance between the two agencies is part of the subsequent analyses. Chapter 7 focuses on the effect of the interplay of the interviewers and the experimentally varied survey features on the outcome variables.

*Table 3.1: Overview of the GESIS Online Panel Pilot project*

|  | Study part 1 | Study part 2 | Study part 3 | Overall |
|---|---|---|---|---|
| Fieldwork start | February 2011 | June 2011 | July 2011 | – |
| Fieldwork agency | Academic, social research agency | Academic, social research agency | Commercial market research agency | – |
| Announced panel duration | 8 months | n/a | n/a | |
| Number of completed telephone interviews with Internet users | 701 | 786 | 1935 | – |
| AAPOR Response Rate 3 before Internet usage screener | 23% | 26% | 14% | 18% |
| Recruitment rate | 61.63% (432) | 52.29% (411) | 52.51% (1016) | 3422 |
| Online participation rate (based on consent) | 75.92% (331) | 55.88% (233) | 46.54% (477) | 1859 |

The telephone recruitment interview survey was announced as a general one on "Everyday life and internet usage in Germany." The questionnaire consisted of questions about private internet use, online survey experience, attitudinal questions, questions on free time activities, questions about landline and mobile phone access, as well as basic demographic questions. Figure 3.2 provides an overview of the components and structure of the interview. The complete questionnaire can be found in the Appendix Chapter 3.

*Figure 3.2: Overview of the GOPP recruitment interview[3]*
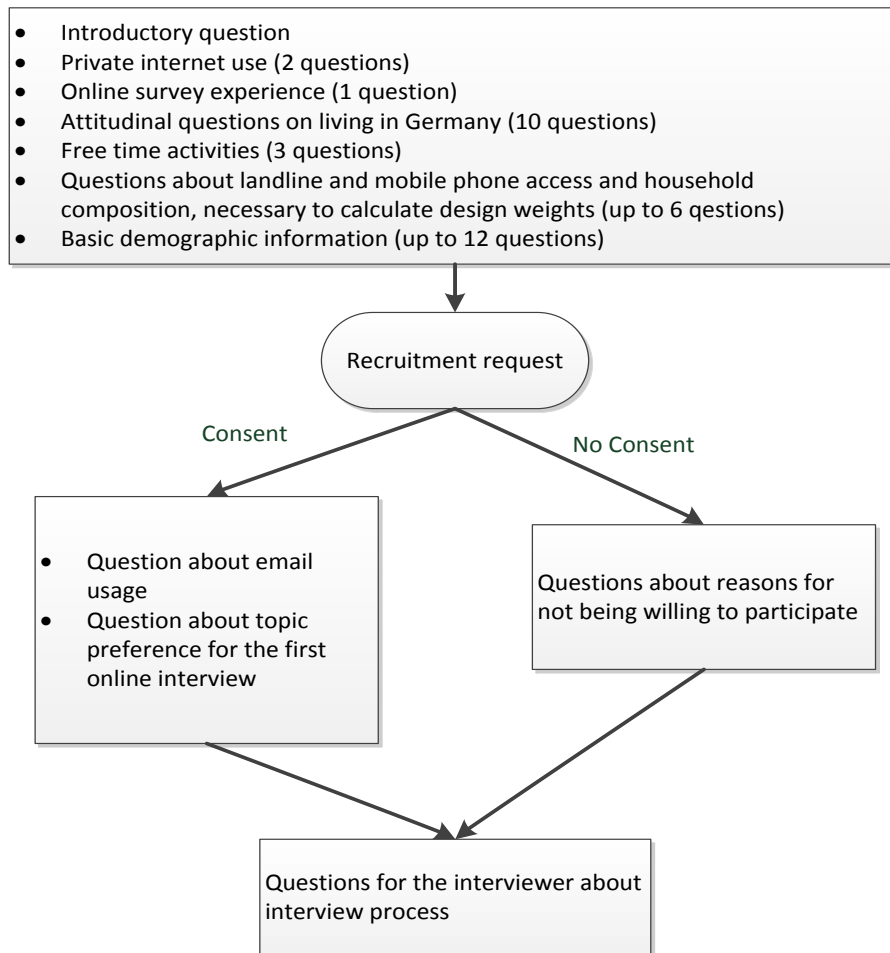


The recruitment interview was designed with the premise to keep the burden for respondents low, include a broad set of questions on different topics which are assumed to be of general interest, and that are furthermore simple to answer and not sensitive. Besides being entertaining for the respondent, a second aim of the recruitment interview was to establish trust in the survey organization.

The structure above shows that the panel was announced at the end of the standardized part of the interview. Hence, the decision to participate in the telephone interview was made independently of the online participation request by the respondents. The recruitment request was placed at the point where the interview would typically end. The interviewers were advised to shortly describe the panel, explaining the procedure of inviting to the following online surveys and asking for subsequent online panel participation as well as the respondents' email addresses. After this short standardized part, the interviewer was expected to continue in a conversational style (Schaeffer & Dykema, 2011). The interviewers were asked to provide as much further information and as many arguments as the respondents needed. Furthermore, they were instructed always to mention the amount of promised

---

[3] Note. In study part 1 the question about internet usage was part of the screening procedure before the interview started. As this procedure results in underestimating the share of internet users in the general population, the procedure was changed in the subsequent study parts.

incentive for a single online survey that was varied experimentally. Screenshots of the CATI-instrument with the German text of the panel request can be found in the appendix of this chapter. The respondents that agreed to participate in the panel and provided a valid email address were invited via email to eight monthly online surveys of durations between 10 and 15 minutes.

The nature of a pilot study gave the unique opportunity to implement various experiments during the recruitment process as well as during the panel operation (Struminskaya et al., 2014; Struminskaya, 2014, 2016). Factors that were experimentally varied during the recruitment interview are the length of the recruitment interview, the inclusion of a sensitive question in the interview, and the amount of promised incentive for online survey participation. The experiments are described in more detail in section 3.4.

In addition to these experimentally varied features of the survey, interviewers had the opportunity to make the recruitment request without conducting the recruitment interview beforehand under certain conditions. Prerequisite of using this opportunity was that a target person refused to participate in a telephone survey in general, multiple appointments failed before, or the target person seemed to be near an interview break-off. Interviewers solely asked about internet use, shortly described the project and tried to obtain panel consent and an email address. No demographic information was collected during the interview, expect for the respondent's sex, assessed by the interviewer. From those telephone interview respondents who further participated in the online panel, demographic information and information needed for the calculation of design weights were collected in the second online survey.

As the dissertation is based on the GESIS Online Panel Pilot as one single data source, I describe the strategy for dealing with missing information in the subsequent section, and define the central dependent variables of the analyses.

## 3.2 Dealing with missing data

The total number of cases that is available for the analyses is 3422 respondents. 1487 of the interviews were conducted by the academic, social research agency, and 1935 by the market research agency. A relevant share of the eligible cases contains missing values on one or more of the included covariates.

Missing data are problematic for several reasons. First, analyses based on samples that contain missing data may lead to biased estimates and, second, the reduced sample size reduces the power of the analysis (Little & Rubin, 2002). Little & Rubin (2002) classify data as being, 1) missing completely at random (MCAR), 2) missing at random (MAR), and, 3) missing not at random (MNAR).

There are several approaches to deal with missing data in the analysis. Applying fullcase analysis by casewise deletion of all cases that have missing values on at least one of the included variables is a commonly used approach (Heeringa, West, & Berglund, 2010). Only in the case of MCAR, the resulting estimates are not biased. However, the sample size and resulting power are reduced anyways.

Little & Rubin (2002, p. 43 ff) argue that the strategy of casewise deletion is inferior in comparison to a multiple imputation strategy, that generally assumes data to be at least MAR. With the recent developments in statistical software packages multiple imputation as a statistical technique to deal with missing data has become popular (White, Royston, & Wood, 2011). The most straightforward approach of multiple imputation is to replace missing data based on a statistical model conditional on observed data (van Buuren, 2012).

Rubin (1987) defines the three steps of multiple imputation as follows:

1. Imputing missing data conditioned on the observed values and generate multiply imputed data sets.

2. Analyzing the imputed data sets separately.

3. Pooling the estimates from the imputed data sets following Rubin's rules

Due to the experimental design, respondents that were assigned experimentally to the short interview group or were recruited without interview have a higher share of missing values. In these cases the data cannot be expected as being MAR. This peculiarity of the experimental design is taken into account in the strategy of dealing with missing data. I chose a two-step approach. First, I use the fact that some information that was missing from the recruitment interview was fielded in a later online survey. Hence, I combine the observed information from the recruitment interview and the online surveys. Relying solely on the so called strategy of logical imputation would not be sufficient because the information from later waves is only available for respondents that were recruited and not for the respondents that did not consent to panel participation. Hence, imputing the missing values that are left is inevitable.

After the combination of all available information from the surveys, the share of respondents that do not have any missing information on any of the included variables is 89% of respondents that were assigned to the regular interview group. 92% of the respondents that were assigned to the short recruitment interview group have valid information of all variables that were available. In the next step, I applied a multiple imputation model with chained equations in order to estimate plausible values for the missing data. Multiple imputation by chained equation is a technique to generating imputations based on a set of imputation models, that is adequate for each variable type with missing values (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001).

The imputation routines were conducted separately for the three versions of the interview (regular, short, recruitment without interview) due to the varying amount of information that is available for the respondents. Variables that were missing due to experimental design were not imputed. All variables and cases that will be used during the data analysis were included in the imputation model (White, Royston, & Wood, 2011). Concurrently, the inclusive strategy of including all available covariates in the imputation model ensures the MAR assumption to be more plausible (Schafer & Olsen, 1998).

Since hypothesis testing strongly depends on stable p-values, standard errors must be estimated as accurate as possible. Hence, I chose a quite high number of 30 imputations that adequately covers the

rule of thumb: the number of imputations should correspond to the percentage of incomplete cases (Bodner, 2008).

The analyses of Chapters 4 to 6 are based on the multiply imputed data sets. All estimates will be combined using Rubin's rules (Rubin, 1987). The analysis of Chapter 7 is not based on the multiply imputed data set because the variables that were included on the respondents level (sex) showed a share of missing values that was neglectable.

## 3.3 Defining central dependent variables

The structure of the analyses of the survey experiments in Chapter 4 to Chapter 6 follows the logic of the multi-step recruitment process. Hence, the same set of outcome measures is used throughout the subsequent analyses. The overall aim is to assess the quality of the recruitment process depending on the experimental variation of the survey features. As quality indicators, I use three response indicators (Wagner, 2012). First, the recruitment rate for panel participation that is associated with stage 2 of the recruitment process for the GOPP as illustrated in Figure 3.1 on page 35, and, second, the online survey participation rate that is associated with the third and fourth stage. Third, the cumulative participation rate summarizes the recruitment rate and online participation rate. As response rates are not necessarily correlated with data quality (Groves & Peytcheva, 2008; Wagner, 2012), I use the selection bias that is introduced by the experimental variation as an additional measure of sample quality.

### 3.3.1 Panel recruitment rate

The panel recruitment rate indicates the success of the recruitment interview. The binary variable of panel recruitment encompasses all respondents that finished the telephone recruitment interview and were asked to provide consent to online panel participation. Respondents who did not reach the end of the interview and were not asked the recruitment question are excluded from the analyses. Respondents are defined as recruited if they consented to participate in subsequent online surveys and provided an email address. Respondents are defined as not recruited if they rejected the request on subsequent panel participation or did not provide a valid email address.

### 3.3.2 Online survey participation rate

The online survey participation rate indicates the success of the subsequent step after recruitment.
All respondents that were recruited for the online panel according to the above given definition represent the base for the measures of online participation. The first measure for online participation indicates whether a respondent participated in at least one out of the eight online surveys. As the project did not include a profile survey (Callegaro & DiSogra, 2008), it does not matter in which of the eight online surveys the respondent participated. Respondents that participated in at least one online survey were defined as being active panel members. The distribution of the total number of online

surveys a respondent participated is presented in Figure 3.3. It shows a bimodal distribution with peaks at both ends of the scale. In selected analyses, the participation in all eight surveys of the study will be contrasted against no participation and participation in 1 to 7 online surveys.

*Figure 3.3: Distribution of online survey participation*
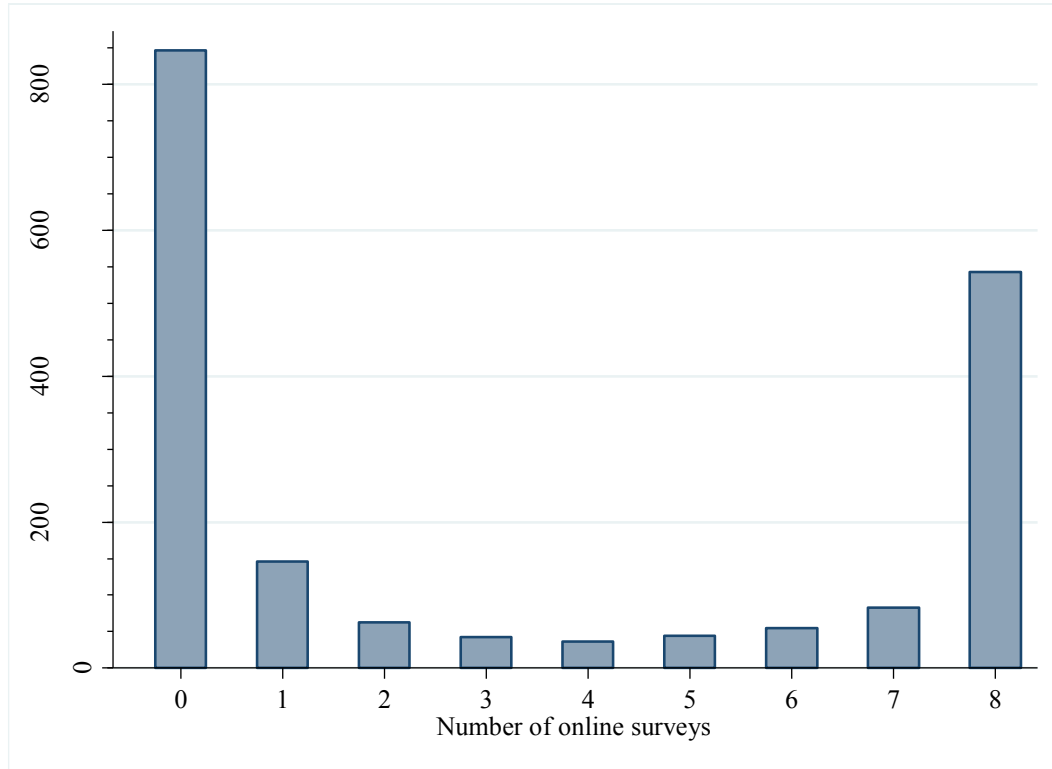


### 3.3.3 Cumulative participation rate

The third response indicator cumulative participation rate summarizes the success of the two previous steps panel recruitment and online participation. It is defined as the share of respondents that participated in at least one online survey (stage 3) out of the respondents that were asked the recruitment question (stage 1).

### 3.3.4 Sample composition

To assess whether the different experimental conditions affect the sample composition of the resulting sample, I compare the composition of respondents across the various treatment groups. As a measure of selection bias, I calculate the pairwise differences in the share of the respondents across experimental conditions at the stage of recruitment and online participation. Similar strategies were chosen for example by Rao et al., (2010) and Pforr et al. (2015).

As the focus of the work is to assess the effect on the sample composition across experimental groups, I do not compare the sample composition with an external benchmark survey. Comparison of the sample composition of the GOPP can be found in Struminskaya et al. (2014).

In the subsequent section, the four studies of the dissertation are briefly described. The aim is to give the reader a preview of what can be expected in the empirical part of the dissertation. Each of the studies is described in more detail in the respective chapter.

## 3.4 Objectives of the empirical studies

The objective of this dissertation is twofold. First, it examines the effect of selected factors of the survey design on the outcome of the recruitment process, and second, it examines the interviewers' role in the recruitment process and in conducting the experiments. In order to answer the question of how experimental variations in the recruitment process of a probability-based online panel in Germany affect 1) the respondents' recruitment probability, 2) the online survey participation probability, and, 3) the sample composition of the final sample, three survey experiments were administered within the recruitment interview. Survey experiments are defined as a systematical variation of an aspect of the survey across subjects to assess the effect of the variation on a measured outcome (Nock & Guterbock, 2010, p. 838). The respondents have to be assigned randomly to the treatment and control group (Tourangeau, 2004). Factors that were experimentally varied are the length of the recruitment interview, the inclusion of a sensitive question in the interview, and the amount of promised incentive for online survey participation.

The outcome variables defined above will be under consideration in the subsequent analyses: the recruitment probability, the online participation probability, and the selection bias introduced by experimental variation.

### 3.4.1 Study 1: Length of the recruitment interview

The duration of the recruitment interview is the first factor of the survey design that was experimentally varied in the recruitment interview. The study tested two versions of the telephone recruitment interview: one with the duration of 3 minutes (short) and one with 10 minutes of duration (regular).

For the shorter version, the recruitment interview was reduced to an initial question about living in Germany, basic demographic information, and the panel participation request. The actual measured length of the interviews matched very well the announced length: The mean length of the regular survey was 9 minutes and 29 seconds with a standard deviation of 3 minutes and ten seconds. The mean length of the short interview was 3 minutes and 55 seconds with a standard deviation of 1 minute and 53 seconds. 90% of the interviews were randomly assigned to the regular interview condition, 10% of the interviews were assigned to the short interview condition. The randomization and group assignment took place before the call. The decision to assign only a minor share of the interviews to the shorter version was made due to pragmatic reasons. As the aim of the GESIS Online Panel Pilot project was to test the recruitment as well as the maintenance of a probability-based online

panel, it would not have been affordable to lose the information collected in the recruitment interview of regular length from half of the sample. The experiment was conducted by both fieldwork agencies and implemented in study part 2 and part 3.

In Chapter 4 the following research questions are addressed:

Research question 4.1:   How does the actual length of the telephone recruitment interview affect the recruitment probability?

Research question 4.2:   Does the differential length of the initial telephone recruitment interview affect the composition of the recruited sample?

Research question 4.3:   How does the interview length of the telephone recruitment interview affect the online participation?

Research question 4.4:   Does the length of the initial telephone recruitment interview have an effect on the sample composition of the sample of respondents that participate online?

### 3.4.2   Study 2: Sensitive question

In the second study, the effect of the second factor of survey design, the inclusion of a sensitive question, on the recruitment probability and the sample composition is investigated. During the process of designing a recruitment interview, researchers are often concerned about including questions that are perceived as being sensitive by the respondents. The split-half experiment tests the effect of including a question about the household income versus not including it in the recruitment interview.

Respondents of the treatment group were asked about their monthly household income. Four bracketed answer categories were provided. Respondents of the control group were not asked about their income. The telephone numbers were randomly assigned to the experimental group before the call. The interviewers were kept blind to the experimental group beforehand. The experiment was conducted only in the regular interviews of full length in all three study parts by both fieldwork agencies.

In Chapter 5 the following research questions are addressed:

Research question 5.1:   How does asking about the income versus not asking about the income in the telephone recruitment interview affect the recruitment probability?

Research question 5.2:   Does asking about the income in the initial recruitment interview have an effect on sample composition of the recruited sample?

### 3.4.3 Study 3: Incentive amount

Incentives represent a response-enhancing measure that is also used at the various stages of the multi-step recruitment process. The experiment tested the effect of a promised incentive, on the recruitment probability, the online participation probability, and the sample composition. This study investigates the potential of promised incentives for online participation to affect nonresponse at several stages of the multi-step recruitment process.

Telephone interview respondents did not receive any incentive for participation in the recruitment interview. The experiment focused on the incentive that was promised for online survey participation. The amount was mentioned by the interviewers along with the information about the online panel during the process of recruitment.

Overall, four incentive groups were tested (0 vs. 2 vs. 5 vs. 10 Euro). Additionally, a bonus of 20 Euro was introduced for respondents that participated in all eight online surveys. Due to the pilot character of the project and a step-wise approach, it was not possible to test all combinations of the factors "amount of incentive" and "additional bonus" under the given constraints. Hence, the experiment was implemented in a factorial design where the factors were not fully crossed:

- Condition 1: 10 Euros + 20 Euros bonus for finishing all eight online surveys
- Condition 2: 5 Euros + 20 Euros bonus for finishing all eight online surveys
- Condition 3: 5 Euros
- Condition 4: 2 Euros
- Condition 5: no incentive

Conditions one and two were tested in study part 1. Conditions three, four and five were tested in study part 2 and study part 3.

The incentive groups were randomly assigned to the telephone numbers before dialing. The interviewers were kept blind to the incentive group until they reached the point in the interview where they had to mention the incentive amount.

The interviewers were advised to mention the respective amount for panel participation and explain the redemption alternatives on demand. To redeem the incentive respondents were offered several possibilities. They could choose between the options of bank transfer if they provided their bank account data, could opt for an Amazon voucher send via email, and had the possibility to donate to several charity organizations, e.g., World Wide Fund for Nature (WWF) or Doctors without Frontiers[4]. In Chapter 6 the following research questions are addressed:

---

[4] The option of donating the incentive was added in study part 2 and 3 as several respondents had asked for the option in the first study part. Respondents that were recruited in study part 1 were given the possibility to donate their accumulated incentives later on as well.

Research question 6.1:    How does the amount of promised incentive for online survey participation affect the recruitment probability and the probability of online survey participation?

Research question 6.2:    Does adding a bonus for loyal respondents versus not adding a bonus on top of the promised incentive have a positive effect on the recruitment probability and the probability of online survey participation compared to the incentive conditions without a bonus?

Research question 6.3:    Do the different amounts of promised incentive for online survey participation affect the composition of the resulting sample?

Table 3.2 summarizes the information on the field work settings for the three experiments.

*Table 3.2: Overview of the features of the three recruitment study parts of the GOPP*

| Feature | Study part 1 | Study part 2 | Study part 3 |
|---|---|---|---|
| Fieldwork start | February 2011 | June 2011 | July 2011 |
| Fieldwork agency | Academic, social research agency | Academic, social research agency | Commercial market research agency |
| Length of recruitment interview | - regular version | - regular version (90%)<br>- shorter version (10%) | - regular version (90%)<br>- shorter version (10%) |
| Experiment on household income question | - household income question asked in 50% of cases | - household income question asked in 50% of cases | - household income question asked in 50% of cases |
| Incentive experiment | - 10 EUR plus 20 EUR bonus for completion of all 8 questionnaires<br>- 5 EUR plus 20 EUR | - 5 EUR<br>- 2 EUR<br>- 0 EUR | - 5 EUR<br>- 2 EUR<br>- 0 EUR |

*Note.* Overview of the features of three recruitment study parts slightly adapted from Struminskaya, B., Kaczmirek, L., Schaurer, I., & Bandilla, W. (2014). Assessing representativeness of a probability-based online panel in Germany. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), Online Panel Research. A Data Quality Perspective (pp. 61–85). New York: Wiley. p. 64.

### 3.4.4    Study 4: Interviewer effects on panel recruitment

The final analysis Chapter 7 focuses on the interviewers as the second aspect that is assumed to influence the recruitment success. The chapter is divided into two separate analysis parts. The aim of the first part is to quantify and explain interviewer variance on the propensity of recruiting respondents. The second part of the analysis focused on the interviewers of the market research agency and their ability to conduct the experiments in a standardized way.

The panel recruitment part of the interview is assumed to be prone to interviewer effects as it includes elements of a conversational setting (Schaeffer & Dykema, 2011). The interviewers were advised to switch to a conversational style after having read the standardized panel introduction. They were also instructed to mention the amount of promised incentive for subsequent panel participation. However, it is assumed that it depends on the experience of the interviewers how they use the measures they were given at hand to convince the respondents to give their consent.

In Chapter 7 the following research questions are examined.

Research question 7.1:     Does the propensity of being recruited differ across interviewers?

Research question 7.2:     To what extent can the interviewer variance in the recruitment propensity be explained by characteristics of the respondents?

Research question 7.3:     To what extent can the interviewer variance in the recruitment propensity be explained by characteristics of the interviewers?

Research question 7.4:     To what extent can the interviewer variance in the recruitment propensity be explained by a match of the sex of interviewer and respondent?

Research question 7.5:     Does the performance of the interviewers of the two fieldwork agencies differ and can potential differences be explained by characteristics of the interviewers?

Research question 7.6:     Does the propensity of being recruited vary across experimental groups, depending on interviewers?

# 4. Study 1: Effect of the length of the recruitment interview on recruitment probability, online participation, and sample composition

## 4.1 Introduction and overview

The first factor of the recruitment interview design that is investigated in this dissertation is the length of the recruitment interview. The overall aim of this chapter is to test whether the length of the recruitment interview has an effect on the recruitment probability, the online participation probability, and the sample composition.

The length of an interview is mentioned by Bradburn (1978) as one of the central survey features that are related to survey burden, in addition to respondent effort and respondent stress. In Chapter 2 the length of the questionnaire is identified as one feature of the survey that is assumed to influence the probability of providing consent to subsequent panel participation.

Dillman, Smyth, & Christian (2014, p. 32) also emphasize that, independently of other factors, the length of a survey represents a large share of the costs of survey participation for the respondents. In the process of questionnaire design, the researchers always face the challenge of finding the optimal length for a questionnaire which does not put too much burden on the respondents and contains all the content that is necessary. The implicit assumption is that with increased length of an interview, the burden for participation increases and the participation rate decreases (Fricker et al., 2012). The concept of questionnaire length is broad (Lynn, 2014) and depends on the survey mode under consideration. First of all, researchers have to distinguish between the announced or assumed length and the actual length of an interview. The first concept is related to the information given to the respondent before the interview starts by the interviewer or in the advance letter in case of self-administered surveys. The second concept is related to the actual time that it takes to finish the interview or the questionnaire and is based on the experience of the respondents. In contrast to the objectively measured length of a survey, the subjectively perceived length of the survey may be influenced by additional factors like, for instance, the interest in the topic (Dillman et al., 2014; Fricker et al., 2012; Marcus et al., 2007).

Depending on the mode of administration, the length of the questionnaire is visible for the respondents beforehand. In the case of paper questionnaires, for instance, the size of the questionnaire is obvious for the respondents (Dillman et al., 2014, p. 353) and can be quantified in terms of pages or the number of questions (Yammarion, Skinner, & Childers, 1991). In contrast, in the case of computer-assisted personal interviews respondents have to rely on the information provided by the interviewer beforehand. For telephone interviews, the announced interview length is one of the central pieces of information in addition to the survey topic that interviewers have at hand in the short time interval that is available to persuade a respondent to participate in the survey (de Leeuw & Hox, 2004). The definition of "short" and "long" is also subjective to the specific survey and survey mode (Holbrook, Green, & Krosnick, 2003).

The discussion of the effect of questionnaire completion time is of particular importance in the context of longitudinal surveys because the previous interview experience was identified to influence the participation probability in the next panel wave (Lepkowski & Couper, 2001; Watson & Wooden, 2009). In the context of the multi-step recruitment process for probability-based online access panels, the experienced burden in terms of questionnaire length is one central feature that is assumed to influence the decision on providing consent to and participating in the online panel surveys.

In the next section, I summarize the pertinent past knowledge on the effect of questionnaire length on subsequent survey participation. On the basis of the literature, I derive the specific research questions addressed, and the hypotheses tested. This part is followed by the description of the data and the experiment. In the analyses, I focus on the effect of the experimental split on the two outcome variables − panel recruitment and panel participation−, as well as on the effect of the experimental variation of recruitment interview length on the sample composition. In the last section, I discuss the results and the implications for survey design.

## 4.2   Research on interview length

The research on questionnaire length can be mainly distinguished by the focus on, 1) the announced or assumed length and, 2) on the actual length of an interview. The focus of this chapter is to the effect of the actual length of the recruitment interview on the propensity for giving consent to subsequent panel participation. To give a broad overview of the topic, the subsequent summary of research on questionnaire length includes studies on the effect on announced length as well.

### 4.2.1   Announced and assumed length

The amount and clarity of evidence on the effect of the announced survey length vary between the different survey modes. For mail surveys, a meta-analysis by Edwards et al. (2002) indicates that the questionnaire length is negatively correlated with response rates. Older meta-analyses that did not include experimental studies came to contradicting results (Heberlein & Baumgartner, 1978; Yammarion et al., 1991). The results of the more recent meta-analysis by Edwards can be judged to be of higher quality because he exclusively included experimental studies. Compared to other survey modes, the relation between questionnaire length and response rate is well documented for mail surveys (Bogen, 1996).

For web surveys, two meta-analyses (Cook, Heath, & Thompson, 2000; Sheehan, 2001) did not reveal a significant relationship between questionnaire length and response rates. Both studies only included non-experimental studies and did not control whether respondents were informed about the length in advance or not. In several experimental studies, a negative effect of announced length on cooperation was shown for web surveys (Crawford, Couper, & Lamias, 2001; Galesic & Bosnjak, 2009; Marcus et al., 2007). The strength of the effect differs from a 4%-points difference in the response rate between an interview that was announced as lasting 20 minutes compared to 8 to 10 minutes (Crawford et al.,

2001), to 7%-points for the contrast between 30 and 10 minutes announced length (Galesic & Bosnjak, 2009), and finally a difference of 22%-points for the contrast between 10 – 20 minutes and 30 – 60 minutes stated interview time (Marcus et al., 2007).

Over two decades ago, Bogen (1996) stated that for interviewer-administered surveys only a small number of experimental studies are available and the negative relationship between announced length and response rate is more likely to be of anecdotal evidence (Bogen, 1996).

In a telephone survey, for example, Collins, Sykes, Wilson, & Blackshaw (1988) showed experimentally a 5%-point increase in the response rate of an interview that was announced as lasting twenty minutes compared to an interview that was announced as lasting forty minutes. The results of Hansen (2007) are in line with this results. He shows that the response rate for the telephone interview that was announced as lasting twenty minutes was significantly lower than an interview that was announced as lasting fifteen minutes. Roberts, Eva, Allum, & Lynn (2010) report about methodological research in the context of the European Social Survey where the announced questionnaire length of a telephone study was varied between 60 minutes, 45 minutes, and 30 minutes. The results show different patterns of the four participating countries in terms of response rate differences. For Hungary and Switzerland, the response rate of the shortest version was significantly higher compared to the two longer versions; the response rate for Germany was significantly lower for the longest version compared to the two other versions.

In the analysis of a series of experiments that tested the effect of various telephone survey introductions on response, de Leeuw & Hox (2004) included the length of the telephone interview as a control variable. The authors revealed that the cooperation rate for longer announced surveys was lower. The results were based on observational studies without experimental variation. Thus, it is not possible to make any statement about causality and the direction of the effect. The authors discuss the alternative explanation that the negative effect of interview length may be mediated by the interviewers' expectation and behavior. The role of interviewers is part of the analyses in Chapter 7 of this dissertation.


## 4.2.2 Actual length and participation in a subsequent survey

An indicator for survey burden that is based on the interview experience is the actual length of the survey. The discussion of the effect of questionnaire completion time is of particular importance in the context of longitudinal surveys or any other survey that asks for subsequent participation, like probability-based online access panels that apply a step-wise recruitment. Apodaca, Lea, & Edwards (1998) emphasize two components of respondent burden in the longitudinal context: first, the immediate burden, and, second, the burden of future interviews, anticipated by the respondents. The actual length of the recruitment interview may be used by the respondents as a proxy to anticipate the burden of the future interviews as it is widely recognized from panel surveys that previous interview experience influences the participation probability in the next panel wave (Watson & Wooden, 2009).

In contrast to the above-mentioned negative relation between announced length and survey participation rates, empirical studies show a positive relation between the actual length of the interview and participation in the next wave (Branden, Gritz, & Pergamit, 1995; Hill & Willis, 2001; Zabel, 1998). Also in the context of recruitment for an online survey Couper et al. (2007) show that the length of the recruitment interview and the willingness to participate in a subsequent online survey are positively related. All of the cited studies did not experimentally vary the questionnaire length. Watson & Wooden (2009) argue that the interview length is not exogenous to the interview process and displays rather the enjoyment and engagement of respondents that are more willing to invest time in an interview, than the received burden of participation. To disentangle the effect of enjoyment and burden, the questionnaire length has to be varied experimentally.

To the best of my knowledge, only three experimental studies examine the issue of questionnaire length and next-wave participation for interviewer-administered surveys. Sharp & Frankel (1983) compare the rate of consent to participate in a second interview one year after the initial personal interview for two versions of the interview that differed in their length. Respondents of the longer interview group (75 minutes) had a 14%-percentage points lower consent rate than the group of respondents of the shorter interview group (25 minutes). However, they did not find a significant difference in the response rate of the second wave between the willing respondents of the long and short interview groups. In line with this finding, Lynn (2014) did not find a significant difference in the second wave participation rate for an interview that took on average 31 minutes and one that took about 26 minutes. This finding holds for the request to complete a self-administered questionnaire, as well as for an interview request that was similar to the initial interview.

For a telephone survey, Fricker et al. (2012) report a small, but significant lower attrition rate between wave 2 and wave 3 for respondents of an interview that took on average 7 minutes less time (29 minutes vs. 22 minutes). Furthermore, the shorter interview was rated as being less burdensome by the respondents.

Finally, the only study that controls for the effect of the length of the recruitment interview for a mixed-mode access panel is the German pilot study conducted by researchers of the University Bremen (Engel, Bartsch, Schnabel, & Vehre, 2012, p. 74). The authors report a higher share of respondents that provided consent to panel participation at the end of a 20 minutes recruitment interview (74%) compared to the respondents that passed the shorter recruitment interview that lasted only 10 minutes (67%). However, the respondents were not assigned randomly to the groups, but the two interviews were conducted independently in separate random samples. The authors discuss the possible explanation that the content of the longer interview was more interesting for the respondents and gave them the opportunity to answer meta-questions about the interview experience. This discussion points to a central problem even if the questionnaire length is varied experimentally; it is not possible to completely disentangle the effect of questionnaire length and the content of the survey (Galesic & Bosnjak, 2009). One possible solution is to drop questions across all parts of the

questionnaire. A strategy that was chosen for example by Marcus et al. (2007), in order to "leave content substantively unchanged" (p. 376) in the shorter version.

Overall, the mixed results do not support the general assumption that longer questionnaires have a negative effect on the willingness to give consent to participation in subsequent survey waves and the actual participation.

Based on the results summarized above, it can be assumed that the association between the length of the recruitment interview and subsequent cooperation is not linear. In line with this (McCarthy, 2006, p. 97) argues that the association of burden and length is a "commonly held belief in survey research" that has not been extensively tested. A possible explanation for the mixed results may be that the received burden in terms of interview time of an interview does not increase linearly with each additional minute of interview time. Moreover, there may be a threshold that additionally may vary from respondent to respondent. In the literature, interviews that exceed one hour are defined as being burdensome for interviewer-administered surveys (Mercer et al., 2015; Singer et al., 1999). However, the threshold of one hour is not derived from empirical data. In conclusion, the insufficient state of research that has been stated by Bogen (1996) has not improved much until now. Furthermore, the duration of the interviews in the summarized experimental studies altogether exceeds the length of ten minutes. In consequence, the effect of very short interviews, where the social interaction between interviewer and respondent is limited to an introduction and few questions, has not been tested so far. The aspect that plays a crucial role with very short interviews is whether the short interaction between interviewer and respondent is sufficient to establish trust. As trust is seen as one base for fostering survey participation (Dillman et al., 2014), it is fundamental to test the effect of a very short interview on recruitment success.

## 4.3   Research questions and hypotheses

The basic research question aims at testing the effect of two versions of the recruitment interview. In the first study of this dissertation, I experimentally test the effect of two versions of the telephone recruitment interview that are both short in comparison to the above-presented studies. The first version of the interview was defined as the "regular interview" and was announced as lasting ten minutes; the short version was reduced to five questions and was announced as lasting three minutes. I examine whether the two versions of the recruitment interview have an effect on the overall recruitment probability, the online participation probability, and the sample composition. The first research question is formulated accordingly.

Research question 4.1:      How does the actual length of the telephone recruitment interview affect the recruitment probability?

As the findings on the effect of actual questionnaire length on subsequent survey participation are not conclusive, I formulate two competing hypotheses in relation to Research question 4.1. The assumed

direction of the effect depends on the underlying mechanism – burden or trust - that is relevant for the respondent's decision. In the literature, it is assumed that a shorter questionnaire is less burdensome for the respondents (Bradburn, 1978). Based on the assumption that less interview time decreases the respondent burden and that the subjective respondent burden is the factor that drives the decision to participate, I derive the following hypothesis.

Hypothesis 4.1:     Respondents of the short interview group have a higher recruitment probability compared to the respondents of the regular interview group.

The empirical evidence, and especially the results from Engel et al. (2012), give hints that the received burden is not sufficient to explain the decision of subsequent survey participation. Respondents may judge not only the amount of time they had to invest but also how they enjoyed answering the question and basing their decision furthermore on their overall judgment of the interview situation. It is assumed that the interviewers in the context of telephone surveys need a minimum of time to establish a trustworthy relationship with the respondent (de Leeuw & Hox, 2004), that enforces subsequent survey participation. In the process of establishing trust, the respondent-interview interaction plays a crucial role (Groves & Couper, 1998). Due to reciprocity and social exchange theory (Dillman et al., 2014), it is assumed that respondents tend to accept a request for further panel participation if they had a positive interaction with the interviewer before. On the basis of these considerations, I derive the following competing hypothesis.

Hypothesis 4.2:     The short interview time of the telephone interview leads to a lower recruitment probability compared to the regular interview.

The aim of the recruitment process is not only to maximize response at several stages but also to aim for a sample that is not biased because of selective nonresponse. Regarding the sample composition, it has to be examined whether the two different treatments result in different sample compositions at the stage of recruited respondents.

The study on sample composition is of exploratory nature, and there are no strong theoretical considerations that allow for prediction of the direction of a possible effect.

Research question 4.2:     Does the differential length of the initial telephone recruitment interview affect the composition of the recruited sample?

Considering the lack of strong theoretical arguments for either a positive or a negative effect, I expect to find no effect of interview length on actual panel participation if controlling for being recruited.

Research question 4.3:     How does the interview length of the telephone recruitment interview affect the online participation?

Finally, I examine the effect of the length of recruitment interview on the sample composition at the online participation stage.

Research question 4.4:    Does the length of the initial telephone recruitment interview have an effect on the sample composition of the sample of respondents that participate online?

## 4.4    Method

### 4.4.1    Experimental design

To test the effect of a very short interview on recruitment success, a survey experiment that varies the length of the telephone recruitment interview was implemented. Two versions of the recruitment interview were randomly assigned to the telephone number. For the shorter version, the recruitment interview was reduced to an introductory question on living in Germany, basic demographic information (sex, age, school education, employment status), and the panel participation request. The interview length was assigned to the telephone number before the call, and the interviewers announced the length of the telephone interview in the introduction of the survey. The regular survey was announced as being of ten minutes duration, the shorter version as being of three minutes duration. The two versions of the questionnaire can be found in the appendix of Chapter 3.

The actual measured length of the interviews matched the announced length very well. The total length of the interview is calculated as the duration of the interview from the moment the interviewer started to read out the first question until the interview and the panel request has been finished. The mean length of the regular survey was 9 minutes and 29 seconds with a standard deviation of 3 minutes and ten seconds. The mean length of the short interview was 3 minutes and 55 seconds with a standard deviation of 1 minute and 53 seconds. The aim was to assign 90% of the interviews randomly to the regular interview condition and to assign 10% of the interviews to the short interview condition. The decision to assign only a minor share of the interviews to the shorter version was made due to pragmatic reasons. As the aim of the GESIS Online Panel Pilot project was to test the recruitment as well as the maintenance of a probability-based online panel, it would not have been affordable to lose the information collected in the recruitment interview of regular length from half of the sample. The experiment was implemented in study part 2 and part 3.

### 4.4.2    Data

Only data from study part 3 that was carried out by the market research fieldwork agency are included in the subsequent analysis. In study part 2 an error occurred in the implementation of the experiment and all but seven interviews that were supposed to be short interviews were conducted as recruitment without an interview. Table 4.1 gives an overview of the number of interviews and the distribution of the different type of interviews, including the option of recruitment without an interview. Of the 1935

interviews conducted in study part 3, 1707 (88.22%) interviews were randomly assigned to the regular interview group and 228 (11.78%) were assigned to the short interview group. In both groups, the interviewers were given the opportunity for asking the recruitment question without previous interview under predefined circumstances. A prerequisite of using this opportunity was that a target person refused to participate in telephone surveys in general, multiple appointments had failed before, and/ or the interviewer had the impression that the target person seemed to be near an interview break-off. Interviewers solely asked about internet use, described the project briefly, and attempted to obtain an email address. No demographic information was collected during the interview. From those telephone interview respondents who further participated in the online panel, demographic information and information needed for the calculation of design weights were collected in the second online survey.

In the group of the regular length interview, about 17% of the recruitment requests were made without prior interview. In contrast, in the short interview group, the interviewers choose to skip the questions only in 1.6% of the interviews.

*Table 4.1: Fieldwork settings type of recruitment interview*

|  | Interview | | Recruitment without interview | | Total | |
| --- | --- | --- | --- | --- | --- | --- |
|  | n | % | n | % | n | % |
| Regular interview (10 minutes) | 1372 | 70.90 | 335 | 17.31 | 1707 | 88.22 |
| Short interview (3 minutes) | 197 | 10.18 | 31 | 1.60 | 228 | 11.78 |
| Total | 1569 | 81.08 | 366 | 18.91 | 1935 | 100.00 |

In the subsequent analyses, the 1569 cases that completed the interview, whether the short or the regular version, are included.

The analyses are based on 30 multiply imputed datasets; the estimates are pooled following Rubin's rules (Rubin, 1987). In all analyses, the clustering of the data within 64 interviewers is taken into account, by calculating robust standard errors.

The frequencies for the two dependent variables of the analyses are shown in Table 4.2. Overall less than half of the 1569 respondents that finished the recruitment interview provided consent to panel participation. Of those, who provided consent, overall 53.47% participated in at least one online survey.

*Table 4.2: Frequencies of dependent variables for the analysis sample*

|  | % | n |
| --- | --- | --- |
| Recruitment | 45.57 | 1569 |
| Online participation | 53.57 | 715 |

## 4.5 Results

In the following analyses, I examine the effect of the two versions of the telephone recruitment interview, the regular interview with ten minutes length, and the short interview with three minutes length. The central independent variable of the analyses is the binary variable that indicates which version – short or regular – of the recruitment interview was conducted. Following the stages of the multistep recruitment process, the outcome variables of interest are the probability of being recruited and the online participation probability. On both stages, I examine the effect of experimental variation on the resulting sample composition.

In a first step, I test whether the randomization process was successful. The sample composition of the two experimental groups is compared on key demographic information. Figure 4.1 shows the selection steps with the respective number of cases that are included in the subsequent analysis. Overall, 1569 respondents finished the recruitment interview (1372 regular interviews, 197 short interviews), 715 respondents provided consent to subsequent panel participation (639 regular interviews, 76 short interviews), and finally, 383 respondents participated in at least one online survey (343 regular interviews, 40 short interviews).

*Figure 4.1: Schematic recruitment process and number of cases that are part of the analysis of the experiment on interview length*
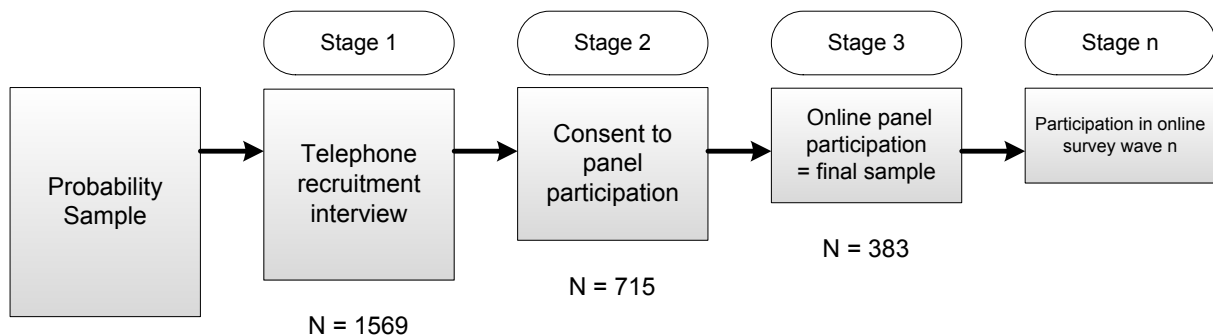


### 4.5.1 Test for selectivity on the telephone interview stage

The length of the interview is one of the central pieces of information that interviewers provide in the short interaction between interviewer and target person while trying to gain cooperation (de Leeuw & Hox, 2004). The above-cited studies revealed that the announced length of an interview affects the decision to participate. As the interview length was assigned to the telephone number before the call and the interviewers announced the length of the telephone interview in the introduction of the survey, selectivity may be introduced at the stage of the telephone interview. For the GOPP, the information about the group assignment of the telephone interview nonrespondents is not available. Hence, it is not possible to quantify the effect of the announced length on the telephone interview participation rate.

To test for selectivity at the stage of the telephone interview (stage 1), the sample composition of the two experimental groups is compared among the respondents that participated in the recruitment interview. The information that is available for respondents of the short and regular interview group is limited because of the limited set of questions that were asked in the short version.

The comparison of the distribution of the two groups shows that the share of highly educated respondents is significantly higher in the regular interview group (Table 4.3). Half of the respondents in the regular interview group were highly educated. In contrast, only 39% of the short interview respondents were in the highest education group. On the other hand, 16% of the respondents in the regular interview group had only basic school education and 21% in the short interview group. These differences give some indication that the very short telephone interview is a possible measure to attract respondents with a lower level of education, a group that is often underrepresented in surveys, and motivate them to participate in interviews for the social sciences.

*Table 4.3: Comparison of sample composition for regular and short interview on the interview stage*

|  | Regular interview (10 minutes) | Short interview (3 minutes) |  |
|---|---|---|---|
|  | % | % | $\chi^2$ |
| % male | 51.90 | 51.27 | 0.027 |
| % in paid work | 75.21 | 73.94 | 0.168 |
| % positive survey attitude | 92.49 | 89.85 | 1.668 |
| % mobile | 44.46 | 43.15 | 0.120 |
| education |  |  | 9.277* |
| *% low education* | 16.15 | 21.51 |  |
| *% middle education* | 33.71 | 39.65 |  |
| *% high education* | 50.13 | 38.85 |  |
|  |  |  | t |
| age (mean) | 42.59 | 43.94 | 1.18 |

*Note.* Comparison of groups with Chi²-test for the categorical variables and t-test for age (Coefficients based on multiply imputed data with M=30, applying Rubin's rules); + p < 0.10, * p < 0.05, ** p < 0.01

In addition, a logistic regression analysis is conducted with the type of interview as binary dependent variable and the characteristics sex, age, education, employment status, positive survey attitude, and kind of telephone sample (mobile vs. landline) as independent variables (see Table 4.4). The multivariate analysis mirrors the results of the bivariate analysis. The two groups only differ with regard to the share of highly educated respondents.

*Table 4.4: Sample composition at interview stage:*
*Logistic regression on experimental variable*
*type of interview*
*(0 = regular interview; 1 = short interview)*

|  | Odds ratio |
|---|---|
| Male | 0.974 |
| Age | 1.004 |
| Education *(Ref. low)* |  |
|    *middle* | 0.904 |
|    *high* | 0.595* |
| In paid work | 0.966 |
| Positive survey attitude | 0.753 |
| Mobile sample | 0.973 |
| N | 1569 |

*Note.* Based on multiply imputed data with M=30, applying
Rubin's rules; 64 clusters in interviewer
$+ p < 0.10$, $* p < 0.05$, $** p < 0.01$

### 4.5.2 Panel recruitment probability

First of all, the two competing hypotheses concerning the panel recruitment probability formulated in section 4.3 are tested. I hypothesize that there is either a positive effect of the shorter interview because of the lower respondent burden or a negative effect because of the shorter time available to establish trust. Next, the effect of the two interview groups on the resulting sample among the recruited respondents is examined to test for selectivity that is introduced by the questionnaire length. The selection step that is the focus of this analysis is marked in orange in Figure 4.2.

*Figure 4.2: Selection step panel recruitment*
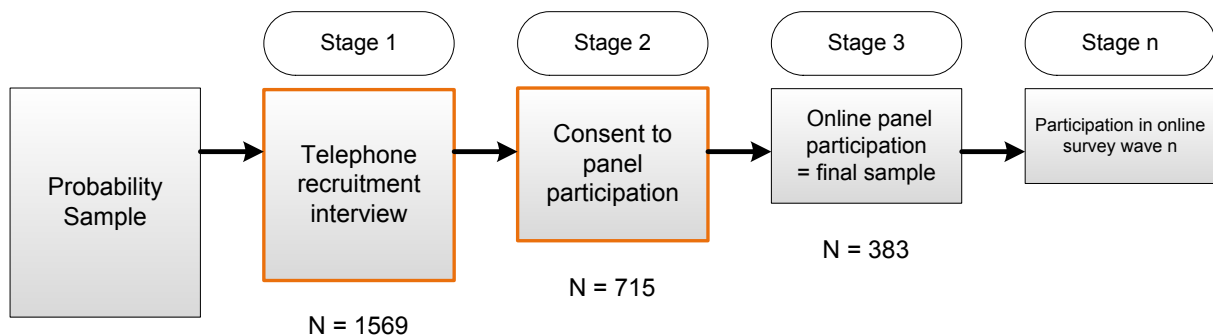


#### 4.5.2.1 *Descriptive results*

In Table 4.5 the bivariate relation between the type of interview and the outcome variable recruitment for all respondents is shown. In the regular interview group, 46.57% of the respondents that were asked the panel question provided consent to subsequent panel participation. In the short interview

group, the share of respondents that provided consent is eight percentage points lower (38.58%). The chi²-test indicates a significant difference between the recruitment rates for the two experimental groups.

*Table 4.5: Bivariate relation between length of interview and recruitment*

|  | % recruited | 95% CI | n recruited | n total |
|---|---|---|---|---|
| Regular interview (10 minutes) | 46.57 | [43.93; 49.21] | 639 | 1372 |
| Short interview (3 minutes) | 38.58 | [31.76; 45.39] | 76 | 197 |
| Total | 45.57 |  | 715 | 1569 |

Pearson chi² (1) = 4.4401 Pr = 0.035

### 4.5.2.2 *Multivariate results*

The bivariate results indicate a positive effect for the regular interview in comparison to the short interview at the stage of recruitment (stage 2). To control for possible selectivity at the previous stage of participation in the telephone survey (stage 1) in the multivariate analysis, available demographic information is included in the models. The analysis of the sample composition at the stage of the telephone interview in section 4.5.1 revealed differences concerning the education within the two experimental groups.

A logistic regression with panel recruitment as the dependent variable was conducted. As control variables, the available demographic information is included: sex, age, education, and employment status of the respondent. Additionally, a dummy variable on positive survey attitude of the respondent, judged by the interviewer, and the information about the experimental incentive group of the respondent is included. To account for the existing differences between the experimental groups, an additional interaction term between education and the experimental group is included in the second step.

The logistic regression on recruitment including the treatment variable and additional control variables shows a significant negative effect of the short interview in the multivariate analysis as well (see Table 4.6). However, the negative effect on panel recruitment for the short interview is no longer significant when controlling for demographic differences between the two groups, by the inclusion of the interaction of experimental group and education. None of the interaction effects of education and type of interview is significant. The calculation of the predicted probabilities (Long & Freese, 2014) for being recruited for the short and regular interview group, based on the second logistic model, results in a probability of 38.58 (CI: 30.48 - 46.68) for the short interview respondents and a probability of 46.57 (CI: 42.28 - 50.87) for the regular interview respondents. The small number of cases (197) in the short interview group leads to large confidence intervals.

It is assumed that the negative effect of the short interview that is present in the bivariate analysis is mainly attributable to differences that occurred on the previous selection step. It may be the case that

the shorter announced interview is more likely to attract respondents with lower education than with higher education.

I find a higher recruitment probability for men, respondents that were easy to persuade to participate in the telephone interview and respondents that were promised higher incentive for online participation. With increasing age of the respondents, the probability of being recruited decreases.

Table 4.6: Logistic regression predicting recruitment

|  | Model 1.1 | Model 1.2 |
|---|---|---|
|  | Odds ratio | Odds ratio |
| Short interview | 0.748[*] | 0.592 |
| Male | 1.254[*] | 1.259[*] |
| Age | 0.982[**] | 0.982[**] |
| Education *(Ref.: low)* | 1.000 |  |
| *middle* | 1.189 | 1.208 |
| *high* | 1.266 | 1.172 |
| Short interview*education |  |  |
| *short*middle* | - | 0.875 |
| *short*high* | - | 1.972 |
| In paid work | 0.797[+] | 0.794[+] |
| Positive survey attitude | 2.070[*] | 2.105[*] |
| Incentive |  |  |
| *2 Euros* | 1.305[+] | 1.319[+] |
| *5 Euro* | 1.341[*] | 1.349[*] |
| N | 1569 | 1569 |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules
64 clusters in interviewer
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

### 4.5.2.3    *Sample composition at the recruitment stage*

To test for selectivity effects that are induced by the experimental treatment, the sample composition of the two experimental groups at the recruitment stage is compared. The information that is available is limited because in the short interview very little information about the respondents was collected. As for the telephone interview stage, the two groups are compared on all available characteristics: sex, age, education, employment status, positive attitude towards surveys, and kind of telephone sample (mobile vs. landline). The distributions of the characteristics are shown in Table 4.7.

To compare the two samples, a logistic regression analysis with the type of interview as the dependent variable is conducted additionally (see Table 4.8). None of the coefficients are statistically significant. In conclusion, based on the considered characteristics no differences between the two experimental groups are found at the recruitment stage.

*Table 4.7: Comparison of sample composition for regular & short interview at the recruitment stage*

|  | Regular interview (10 minutes) | Short interview (3 minutes) |  |
|---|---|---|---|
|  |  |  | χ² |
| % male | 54.30 | 53.95 | 0.003 |
| % in paid work | 73.33 | 75.92 | 0.284 |
| % positive survey attitude | 94.99 | 96.05 | 0.164 |
| % mobile | 48.36 | 42.11 | 1.064 |
| education |  |  | 0.226 |
| % low educated | 14.80 | 16.71 |  |
| % middle education | 34.18 | 33.77 |  |
| % high education | 51.02 | 49.52 |  |
|  |  |  | t |
| age (mean) | 40.416 | 41.368 | 0.54 |
| N | 639 | 76 |  |

*Note.* Comparison of groups with Chi²-test for the categorical variables and t-test for age (Coefficients based on multiply imputed data with M=30, applying Rubin's rules)
+ p < 0.10, * p < 0.05, ** p < 0.01

*Table 4.8: Sample composition recruitment: Logistic regression on experimental variable type of interview*
*(0 = regular interview; 1 = short interview)*

|  | Odds ratio |
|---|---|
| Male | 0.966 |
| Age | 1.003 |
| Education *(Ref.: low)* |  |
| *middle* | 0.836 |
| *high* | 0.820 |
| In paid work | 1.169 |
| Positive survey attitude | 1.307 |
| Mobile sample | 0.782 |
| N | 715 |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules; 64 clusters in interviewer
+ p < 0.10, * p < 0.05, ** p < 0.01

### 4.5.3    Online participation probability

The third and the fourth research question that is addressed in the subsequent analyses focus on the selection step 3, the decision to participate in an online survey after having agreed to do so. In the following analysis, I compare the online participation rates between the respondents of the short interview group that provided consent to subsequent panel participation and the respondents of the

regular interview group that provided consent to subsequent panel participation (Research question 4.3). Like in the analysis for the recruitment stage, finally, I examine whether the two groups of respondents that participate in at least one online survey differ in their sample composition (Research question 4.4). Figure 4.3 shows the selection step that is considered in this analysis marked in orange.

*Figure 4.3: Selection step panel online participation*
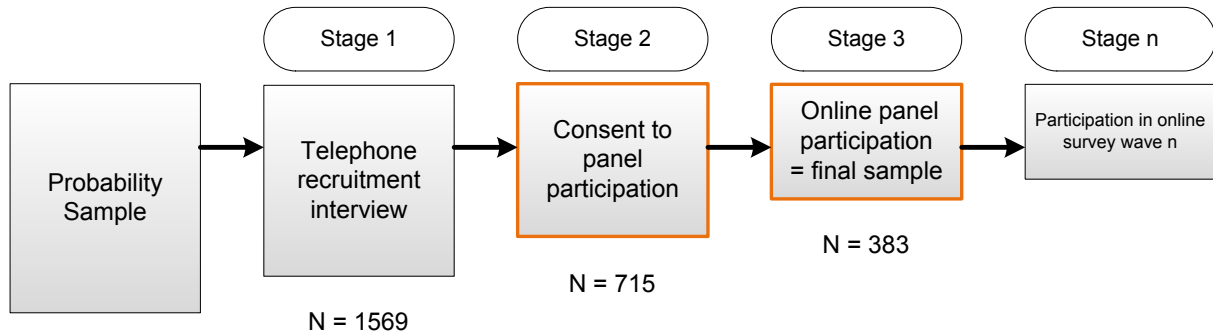


### 4.5.3.1 *Descriptive results*

In contrast to the recruitment stage, no difference in the participation rates of those that consented between the two interview groups is found. The bivariate relation between the length of the interview and actual panel participation is shown in Table 4.9.

*Table 4.9: Bivariate relation between length of interview and online participation*

|  | % participated | 95% CI | n participated | n total |
|---|---|---|---|---|
| Regular interview (10 minutes) | 53.68% | [49.81; 57.55] | 343 | 639 |
| Short interview (3 minutes) | 52.63% | [41.33; 63.93] | 40 | 76 |
| Total | 53.57% |  | 383 | 715 |

*Note.* Pearson chi² (1) = 0.0299 Pr = 0.863

The binary operationalization of online participation only takes into account whether a respondent ever participated in at least one of the eight online surveys or not. It does not reflect the overall number of online surveys. To get a more detailed view, a second threefold outcome variable was generated that differentiates whether a respondent did not participate online at all, participated in one to seven surveys, and finally in all eight surveys. However, the experimental variation of the interview length also does not have an effect on the frequency of online participation either. The detailed results are shown in Table A 1 in the appendix.

### 4.5.3.2 *Multivariate results*

The logistic regression that controls for several characteristics of the respondent confirms the bivariate result. The length of recruitment interview does not have any effect on actual panel participation among those respondents that provided consent to panel participation. Overall, the highly educated

respondents have a higher probability of participation compared to respondents that have a basic and medium level of education. Furthermore, at the stage of online participation, age has a positive effect.

*Table 4.10: Logistic regression on online participation*

|  | Odds ratio |
|---|---|
| Short interview | 0.926 |
| Male | 0.775 |
| Age | 1.026** |
| Education *(Ref.: low)* |  |
| *middle* | 1.314 |
| *high* | 2.586** |
| In paid work | 1.086 |
| Positive survey attitude | 1.520 |
| Incentive |  |
| *2 Euro* | 1.149 |
| *5 Euro* | 1.247 |
| N | 715 |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules; 64 clusters in interviewer
$+ p < 0.10, * p < 0.05, ** p < 0.01$

### 4.5.4 Cumulative participation probability

Finally, to evaluate the overall recruitment success, I calculated the cumulative recruitment probability for the respondents of the two experimental groups. The cumulative participation rate is defined as the share of respondents that participated in at least one online survey (stage 3) from the respondents that were asked the recruitment question (stage 1) (see Figure 4.4).

*Figure 4.4: Selection step cumulative participation*
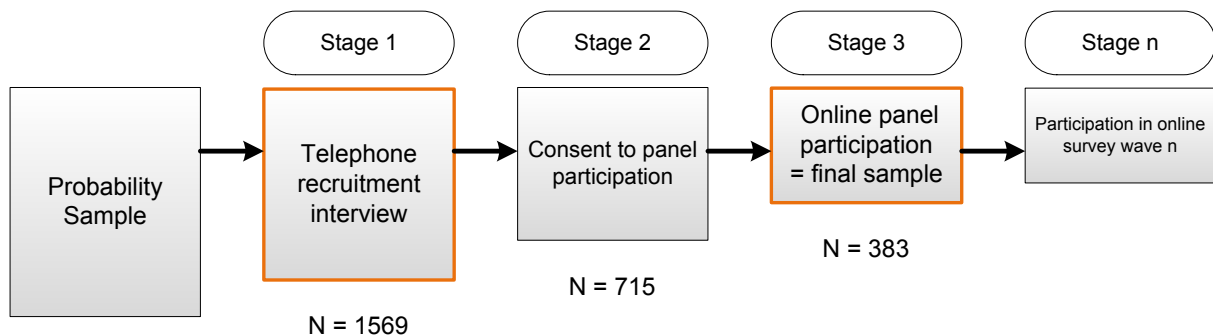


#### 4.5.4.1 *Descriptive results*

Overall, the cumulative participation probability is slightly higher for the regular interview group (25%) in comparison to the short interview group (20.30%).

*Table 4.11: Bivariate relation between length of interview and overall participation*

|  | % participated | 95% CI | n participated | n total |
|---|---|---|---|---|
| Regular interview (10 minutes) | 25.00 | [22.71; 27.29] | 343 | 1372 |
| Short interview (3 minutes) | 20.30 | [14.67; 25.94] | 40 | 197 |
| Total | 24.41 |  | 383 | 1569 |

Pearson chi² (1) = 2.0583 Pr = 0.151

### 4.5.4.2    *Multivariate results*

The multivariate analysis for the overall participation confirms the diminishing of the effect of the type of interview when controlling for differences in sample composition (see Table 4.12).

*Table 4.12: Logistic regression on overall participation*

|  | Odds ratio |
|---|---|
| Short interview | 0.853 |
| Male | 0.996 |
| Age | 1.003 |
| Education *(Ref.: low)* |  |
|     *middle* | $1.532^{+}$ |
|     *high* | $2.198^{**}$ |
| Short interview*edu. |  |
|     *Short\*medium edu* | 0.471 |
|     *Short\*high edu* | 1.473 |
| In paid work | 0.934 |
| Positive survey attitude | $2.292^{*}$ |
| Incentive |  |
|     *2 Euros* | $1.360^{*}$ |
|     *5 Euro* | $1.444^{*}$ |
| N | 1569 |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules; 64 clusters in interviewer
$+ p < 0.10$, $* p < 0.05$, $** p < 0.01$

### 4.5.4.3    *Sample composition at the stage of online participation*

Like at the recruitment stage, the sample composition of the online participants of the two interview groups is compared for the resulting sample at the online stage as well. The distribution of the variables is shown in in Table 4.13 below.

In addition, a logistic regression analysis is conducted with the experimental variable "type of interview" as the dependent variable and the characteristics sex, age, education, employment status, easiness to persuade the respondent to participate in the telephone interview, and kind of telephone sample (mobile vs. landline) as independent variables (see Table 4.14). None of the predictor variables

are statistically significant. Hence, based on these characteristics, no differences between the two experimental groups are found in the final sample.

*Table 4.13: Comparison of sample composition for regular and short interview at the online participation stage*

|  | Regular Interview (10 minutes) | Short Interview (3 minutes) | $\chi^2$ |
|---|---|---|---|
| % male | 51.90 | 47.50 | 0.277 |
| % in paid work | 75.22 | 70.00 | 0.516 |
| % positive survey attitude | 96.21 | 92.50 | 1.232 |
| % mobile | 43.73 | 32.50 | 1.849 |
| education |  |  | 2.188 |
| % low education | 10.31 | 15.42 |  |
| % middle education | 30.54 | 20.83 |  |
| % high education | 59.15 | 63.75 |  |
|  |  |  | t |
| age (mean) | 42.584 | 44.975 | 1.01 |
| N | 343 | 40 |  |

*Note.* Comparison of groups with chi²-test for categorical variables and t-test for age (Coefficients based on M=30, applying Rubin's rules)
+ p < 0.10, * p < 0.05, ** p < 0.01

*Table 4.14: Sample composition online: Logistic regression on experimental variable type of interview*
*(0 = regular interview; 1 = short interview)*

|  | Odds ratio |
|---|---|
| Male | 0.784 |
| Age | 1.008 |
| Education *(Ref.: low)* |  |
| *middle* | 0.458 |
| *high* | 0.692 |
| In paid work | 0.853 |
| Positive survey attitude | 0.499 |
| Mobile sample | 0.651 |
| N | 383 |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules; 64 clusters in interviewer
+ p < 0.10, * p < 0.05, ** p < 0.01

In the final section, I briefly summarize the results and discuss implications for future survey design, as well as avenues for future research.

## 4.6    Summary and discussion

The study adds to the existing literature by experimentally testing two comparably short versions of a telephone recruitment interview for a probability-based online panel that was hitherto not tested.

In the analysis, I investigated the effect of a very short version, which was announced as lasting three minutes, in comparison to a ten-minute interview at the recruitment probability and resulting sample composition. For the short version, the regular interview was reduced to an icebreaker question about living in Germany, basic demographic questions, and finally, the recruitment request. The measured mean length of the interviews matched the announced length remarkably well.

For Research question 4.1 on the effect of the interview length on panel recruitment, I formulated two competing hypotheses: Hypothesis 4.1 stated that the shorter interview time and lower respondent burden leads to a higher recruitment rate compared to the longer interview. Hypothesis 4.2 states the short interview leads to a lower recruitment rate compared to the longer interview because the short interaction is not sufficient to establish trust.

The analysis revealed that at the stage of the recruitment, the longer version is more successful in terms of recruitment with an eight percentage-point increase in the recruitment rate. However, the effect diminishes when controlling for differences in sample composition at the stage of the interview. There is some indication that at the previous nonresponse step, the announced length affects the sample composition of the respondents that participate in the telephone interview. The respondents of the short interview group were significantly less well educated compared to the respondents of the regular interview group. Hypothesis 4.1 that states the shorter interview time and lower respondent burden lead to a higher recruitment rate was not supported by the data and has to be rejected. Concurrently, the competing Hypothesis 4.2 has to be rejected. The difference in the recruitment probability between the two interview groups is not large enough to result in a significant difference based on the limited sample size. To derive statistically sound evidence on the superiority of the 10 minutes interview, the experiment has to be repeated with a larger group of respondents in the short interview group.

Next, the effect of the experimental manipulation on the sample composition of the recruited sample was evaluated (Research question 4.2). The analysis shows no differences in sample composition among the experimental groups based on a limited set of characteristics. The varying length of the telephone recruitment interview does not introduce selection bias at the stage of the recruited sample. Furthermore, the effect of the experimental variation of the interview length was evaluated at the selection stage of online participation. Once the participants were recruited, the length of the recruitment interview had no effect on the decision to participation in the online survey (Research question 4.3).

Finally, I evaluated the effect of the experimental manipulation on the sample composition of the online sample. As for the recruitment stage, no differences in the sample composition of the two groups were revealed (Research question 4.4).

The study suffers from some shortcomings that open up paths for future research. First of all, only two versions of the recruitment interview were tested. As discussed in related studies on questionnaire length (Hansen, 2007; Lynn, 2014) the effect may depend on the actual length, as well as the absolute and relative difference of the interview length between groups. It is reasonable that for longer recruitment interviews, above a specific threshold, the effect on recruitment will be negative. It is reasonable that the relation between interview length and panel recruitment probability is u-shaped. If the interview is too short to establish trust, it has a negative effect on participation, and if it is too long and becomes burdensome, the participation rate will be negatively affected as well. To test for this tradeoff between burden and trust and to find the optimal length, additional longer versions of the recruitment interview would have to be included. In addition, the assumed mechanism of trust and burden were not measured in the study and could not be tested in a strict sense. To test the concepts, in a next step, the received burden of the two versions of the interview and the trust in the credibility of the fieldwork agency have to be collected for the target persons of both experimental groups.

Surveys in the social sciences suffer from a bias in education, with an overrepresentation of highly educated respondents. The educational bias is enhanced by the additional selection steps of probability-based online panels (Bandilla et al., 2009; Blom, Herzing, et al., 2016; Bosnjak et al., 2017). The fact that less educated respondents are attracted by the very short announced survey length is a relevant finding that should be further investigated.

Finally, the combined effect of the announced and the actual length of the overall recruitment rate and the resulting sample composition was not possible to empirically test with the data from the analysis. In a future study, the previous selection stage that examines the decision to participate in the telephone recruitment interview needs to be included.

# 5. Study 2: Effect of a sensitive question in the recruitment interview on the recruitment probability and sample composition

## 5.1 Introduction and overview

The second study of this dissertation investigates the effect of asking about the income information as a sensitive question on the overall recruitment probability of the respondents. In addition to the length of the recruitment interview, the inclusion of sensitive questions in the interview is another aspect that is assumed to be related to survey burden (Bradburn, 1978). Following Tourangeau et al. (2000) three types of sensitive questions can be distinguished: first, a question that touches a "taboo" topic, where asking about it can be seen as an invasion of privacy (e.g., income, sexual behavior); second, a question that involves a threat of disclosure where answering the question honestly may have negative consequences for the respondent (e.g., drug consumption, tax fraud); and third, questions about behavior or attitudes where well-defined social norms exist about what is socially desirable (e.g., voting). The question about income is an example of the first type, as asking about the income information is often interpreted as an invasion of privacy by the respondents (Tourangeau, Groves, & Redline, 2010).

The inclusion of questions about income in a recruitment interview contradicts to the aim of the recruitment interview that should be designed with a low burden and maximized survey enjoyment.

On the other hand, the income information is part of the standard demographic information in social science surveys. And survey researchers are interested in quality indicators such as nonresponse bias and being able to compare the composition of the sample with external benchmarks and evaluate the quality of the recruitment process.

The contradicting aims lead to the discussion if it is advisable to include questions in a recruitment interview that are known to be sensitive and prone to item nonresponse, such as, for instance, the income question. It can be argued, that it not only negatively affects the item nonresponse rate but as well has a negative effect on the resulting recruitment probability. In the context of panel surveys, income nonresponse is known as an indicator of next wave unit nonresponse (Scherpenzeel, Bergmann, Friedel, & Bristle, 2016; Schräpler, 2004; Uhrig, 2008) To the best of my knowledge, only one study examines the connection between income item nonresponse and the consent to participate in an online panel (Bartsch, 2012).

The overall aim of the second study of this dissertation was to test whether asking about income information in the recruitment interview for a probability-based online panel has a negative effect on recruitment. A split-half experiment was conducted. Half of the telephone numbers were randomly assigned to the experimental group where the question about household income was asked and the other half to a control group without any question about income, everything else being equal.

In the remainder of the chapter research on income as a sensitive question and its effect on nonresponse is summarized, and research questions and accompanied hypotheses are formulated. This section is followed by the analysis and discussion of the results.

## 5.2    Research on income as a sensitive question and nonresponse

The question about the income is one of the survey questions known to be sensitive as it is interpreted as an invasion of privacy by the respondent (Tourangeau et al., 2000). Respondents may feel that their income is a topic they do not talk about with a stranger. The income question is one of the questions that show high levels of item nonresponse between 20% up to 40% (Juster & Smith, 1997; Moore, Stinson, & Welniak, 1999).[5]

Reasons for item nonresponse can be identified at several stages during the question-answer process (Tourangeau et al., 2000). While faced with a survey question respondents have to understand the question, recall the relevant information, use the information to make a judgment, format the judgment to fit it into the answer categories, and report an edited answer (Strack & Martin, 1987; Tourangeau, 1984). Income item nonresponse can either occur because a respondent cannot recall the information because he or she actually does not know the income, a fact that may be more plausible for the household income than the individual income. The second possible reason is that the respondent is not willing to report the information because the information is judged as being too confidential. For this reason, Schräpler (2004) argues that in analyses of income item nonresponse "do not know"-answers and refusals to income questions should be distinguished.

The interrelation between item and unit nonresponse is not well investigated (Yan & Curtin, 2010). In the traditional view, Groves & Couper (1998) describe item and unit nonresponse as two independent processes with different causes. In contrast, Loosveldt, Pickery, & Billiet (2002) state that unit and item nonresponse have common characteristics. Both occur during the interviewer-respondent interaction and are a negative reaction of the respondent to a request of the interviewer. Additional studies also rely on the idea of a response continuum (Loosveldt et al., 2002; Schräpler, 2004; Yan & Curtin, 2010) that is based on the idea that respondents can be placed on a continuum depending on their propensity to participate in a survey and answer survey questions (Yan & Curtin, 2010). The continuum goes from a zero propensity to respond to the survey request and survey question, over a low and high relative propensity to participate in the survey and answer the questions. Any change in a survey may change the response propensity of a respondent and move his or her position on the response continuum. The inclusion of a sensitive question in a recruitment interview could represent one change that may move respondents from the "participation" side of the continuum to the "nonresponse" side. In their study, Yan & Curtin, 2010 found support for the response continuum

---

[5] In the German General Survey 2012 (ALLBUS) about 14% of the respondents refused to provide the household income information (own calculation, based on ALLBUS codebook).

model and illustrated that respondents that had a lower response propensity to the survey had higher item nonresponse rates. Other studies also revealed that initial reluctand respondents that were converted to survey participation (Mason, Lesser, & Traugott, 2002; Stinchcombe, Jones, & Sheatsley, 1981) and respondents that were rated as being reluctand by the interviewers (Campanelli, Sturgis, & Moon, 1996; Couper, 1997) had higher item missing rates.

For panel studies, the connection between income item nonresponse and next wave nonresponse was given attention. Income item nonresponse was shown to be predictive of subsequent wave nonresponse in several German panel studies (Müller & Castiglioni, 2015; Scherpenzeel et al., 2016; Schräpler, 2004), as well as international studies like, for instance, the Belgian Election Study (Loosveldt et al., 2002) and the British Household Panel Study (Uhrig, 2008). Similarly, Sakshaug & Kreuter (2011) found a lower probability of giving consent to link survey data with administrative data for respondents that were reluctant to answer the income question in the current and previous wave of a panel. While the evidence on income question item nonresponse as an indicator for next wave unit nonresponse is well documented, the underlying mechanism is not clear (Yan & Curtin, 2010).

To the best of my knowledge, only one study examines the effect of income unit nonresponse on the willingness to participate in an online panel. In line with the evidence from longitudinal studies, Bartsch (2012, p. 135) reports a significant negative effect on the willingness to participate for respondents that refused to provide the income information. She points out that experimental data are needed to assess whether asking about the income itself has a negative effect on the recruitment success or the income item nonresponse is an indicator for a respondent that has a lower recruitment probability, for example, due to privacy concerns.

In the latter case, the income nonresponse can be interpreted as an indicator of trust or confidentiality concerns. Hence, asking about the income should not affect the respondents' position on the response continuum and should not have a negative effect on the decision whether to participate in a subsequent panel wave or not. In the former case, where asking about a sensitive topic triggers confidentiality concerns, asking about the income could have negative effects on the participation probability. Experimental studies revealed that confidentiality concerns contribute to both, unit and item nonresponse (Dillman, Sinclair, & Clark, 1993; Singer, Von Thurn, & Miller, 1995). Consequently, income item nonresponse was used as an indicator for privacy concerns of respondents (Sakshaug, Couper, Ofstedal, & Weir, 2012).

Several studies on characteristics of income nonrespondents came to inconclusive results. In the German sample of the Study of Health, Age and Retirement (SHARE), for instance, income nonrespondents had a higher probability to still work, have a high income. Also, they were less likely to be the person that is financially responsible in the household. Schräpler (2004) separated income refusal, and income "do not know" answers in his analyses with the data from the German Socio-Economic Panel. He showed income nonrespondents refused were more likely to have a high occupational status. Also, respondents that chose the self-completion mode had a higher probability to

refuse the income question compared to respondents of the personal interview group. In contrast, respondents that indicated to not knowing their income were more likely to have a lower occupation status. The replication of the study for the British Household Panel Survey revealed different characteristics for refusals and "do not know" answers as well. However, the occupational status was no predictor for income refusal in the British study. Man had a higher probability to refuse.

## 5.3 Research questions and hypotheses

To gain insights into the effect of including sensitive questions in a recruitment interview on subsequent panel participation, a split-half experiment was conducted. The leading research questions of the study and the accompanied hypotheses were the following:

Research question 5.1: How does asking about the income versus not asking about the income in the telephone recruitment interview affect the recruitment probability?

Two directed hypotheses are formulated. It was shown that income item nonresponse is an indicator for predicting future survey behavior. Respondents that refused to provide the income information had a lower probability of giving consent to subsequent panel participation, giving consent to data linkage, and had a lower probability of next wave participation. On the basis of these considerations the first hypothesis is derived:

Hypothesis 5.1: Asking about the income in a recruitment interview versus not asking about the income negatively affects the recruitment probability.

In Figure 5.1, the assumed causal relation between the treatment groups and the outcome variable consent to panel participation and the directions of the effects are visualized.

*Figure 5.1: Visualization of hypothesis 5.1*
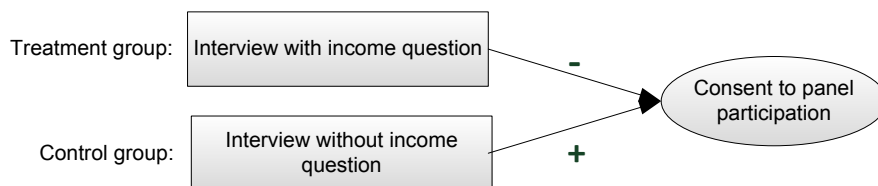


The income question is defined as a sensitive question in a sense, that it is perceived as an invasion of privacy by some respondents. Respondents that are concerned about their privacy and have doubts about the confidentiality of the information provided are assumed to have a higher probability of income item nonresponse. Respondents that are not concerned about the confidentiality of their data are assumed to have no concerns to provide their income information. Following this argumentation,

the income item nonresponse would be an indicator for a reluctant respondent, and the following specification of Hypothesis 5.1 is formulated:

Hypothesis 5.2: Asking about the income versus not asking about the income negatively affects the recruitment probability of the respondents that refused to provide the income information.

In Figure 5.2 the answering behavior of the respondents that were asked about the income as a moderator variable is included in the initial graph, and the assumed causal relations and directions of the effect are included.

*Figure 5.2: Visualization of hypothesis 5.2*
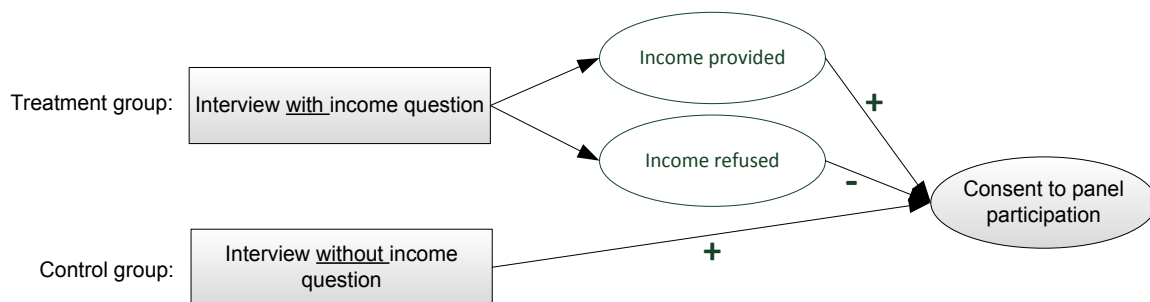


Being aware of the potential bias that is introduced by the survey design, the following research question regarding the sample composition is formulated:

Research question 5.2: Does asking about the income in the initial recruitment interview have an effect on sample composition of the recruited sample?

Earlier research about the characteristics of respondents that refused to provide the income information is inconclusive, depending on the kind of income information asked, the operationalization of nonresponse, and the country of the survey. Hence, no directed hypotheses will be derived.

The research on income item nonresponse in panel studies revealed higher rates of income item nonresponse among the first survey wave (Scherpenzeel et al., 2016; Schräpler, 2004) It is assumed that the effect of the income question mainly is to find at the recruitment stage. After respondents decided to give their consent to subsequent panel participation, independent of having refused to provide the income information, having provided the information, or not having been asked, no additional effect on the decision of actually participating online is assumed. For this reason, the focus of the analysis is at the recruitment stage.

## 5.4 Method

### 5.4.1 Experimental design

The split-half experiment was implemented in the interviews of regular length. Respondents of the treatment group were asked about the monthly net income of their household. They were provided four bracketed answer categories: Up to 1,500 Euro, between 1,500 und 2,500 Euro, between 2,500 und 4,000 Euro, and more than 4,000 Euro (for the exact German question wording, see Appendix Chapter 5). The respondents of the control group were not asked any income information, anything else being equal. The income question was part of the demographic information that was asked at the end of the questionnaire before the panel was introduced.

The experiment was not implemented in the shorter version of the interview where only basic demographic information was asked. It was conducted in all study parts and by both fieldwork agencies. The groups were assigned randomly before the call. However, the interviewer did not have any information in which of the two groups the interview is until the question popped up during the interview.

### 5.4.2 Data

Overall 2657 interviews were part of the experiment. 1336 (50.28%) respondents received the question about the household income, 1321 (49.72%) did not receive the question.

The frequencies for the two dependent variables of the analyses are shown in Table 5.1. Overall less than half of the 1569 respondents that finished the recruitment interview provided consent to panel participation. Of those, who provided consent, overall 53.47% participated in at least one online survey.

*Table 5.1: Frequencies of dependent variables for the analysis sample*

|                                     | %     | N    |
|-------------------------------------|-------|------|
| Recruitment                         | 51.04 | 2657 |
| Online participation, if recruited  | 61.14 | 1356 |

The distributions of the variables that were included as control variables in the multivariate models and used as indicators in the analyses of the sample composition are presented in Table 5.2. About 58% of the interviews were conducted on the landline, and 42% of the interviews were conducted via cell phone. 51% of the respondents were male. Almost half of the respondents were highly educated. 72% of the respondents were in paid work. As a proxy measure for a positive survey attitude, the interviewers' judgment on how easy it was to persuade the respondent to participate in the telephone interview is used. The vast majority of 89% of the respondents had a positive attitude towards surveys. About 20% of the telephone survey respondents had participated in at least one online survey in the

year before the interview. 86% of them used the internet at least several times a week. The mean age in the analysis sample was 43 years.

*Table 5.2: Frequencies of independent variables for the analysis sample at the interview stage*

|  | % | N |
|---|---|---|
| Treatment |  | 2657 |
| *income asked* | 49.72 |  |
| *income not asked* | 50.28 |  |
| Agency |  | 2657 |
| *social research agency* | 48.36 |  |
| *market research agency* | 51.64 |  |
| Sample |  | 2657 |
| *landline sample* | 57.96 |  |
| *mobile sample* | 42.04 |  |
| Male | 51.09 | 2657 |
| Education |  | 2657 |
| *low* | 16.02 |  |
| *middle* | 33.72 |  |
| *high* | 50.26 |  |
| In paid work | 71.95 | 2657 |
| Positive survey attitude | 89.04 | 2657 |
| Earlier online survey participation | 20.54 | 2657 |
| Internet usage (1=several times a week; 0=once a week or less) | 86.04 |  |
| Incentive group |  | 2657 |
| *0 Euro* | 27.71 |  |
| *2 Euro* | 26.08 |  |
| *5 Euro* | 24.28 |  |
| *5 Euro + bonus* | 12.15 |  |
| *10 Euro + bonus* | 11.78 |  |
|  | mean |  |
| Age | 43.26 | 2657 |

The analyses are based on 30 multiply imputed datasets; the estimates are pooled following Rubin's rules (Rubin, 1987). In all analyses, the clustering of the data within 88 interviewers is taken into account, by calculating robust standard errors.

## 5.5 Results

In the following analyses, the effect of the split-half experiment on asking about the income versus not asking in the recruitment interview is investigated. The exposure of the treatment asking about the income serves as the central independent variable of the analysis. It is coded zero for the control group that did not receive the income question, and one for the respondents of the treatment group that were asked about their household income.

1336 (50.28%) respondents were asked about their household income, 1321 (49.72%) did not receive the question. Within the treatment group 80.85% of the respondents provided the income information,

19.15% rejected to provide this information. Following Schräpler (2004), the reject category is divided into those who admit to not knowing the monthly net household income (7.04%, 93) and those who rejected to answer the question (12.11%, 160) (see Table 5.3).

The share of respondents that provided the income information is slightly higher for the market research agency (82.20% vs. 79.44%). However, the difference is not statistically significant (Chi²(1)=1.6149; Pr=0.204). The pattern of the two nonresponse reasons, however, differs between the two agencies (see Table 5.3). Respondents of the social research agency more often explicitly reject to answer the income question (15.30%) compared to not knowing the income (5.26%), whereas for respondents of the market research agency the share of the two categories is almost equal (9.05% refusal vs. 8.75% do not know). The Chi²-test indicates a significantly different pattern for the two agencies. Being aware of these differences, the potential differences between the two agencies is considered in the subsequent analyses.

*Table 5.3: Pattern of income (non)-response for the treatment group, by fieldwork agency*

|  | Income provided | Income refused | Income do not know | Total |
|---|---|---|---|---|
|  | % (n) | % (n) | % (n) | % (n) |
| Social research agency | 79.44 (514) | 15.30 (99) | 5.26 (34) | 100 (647) |
| Market research agency | 82.20 (554) | 9.05 (61) | 8.75 (59) | 100 (674) |
| Total | 80.85 (1068) | 12.11 (160) | 7.04 (93) | 100 (1321) |

*Note.* Pearson chi² (1) = 16.6987 Pr = 0.000

Following the stages of the multistep recruitment process, the central dependent variables for the analyses is the panel recruitment probability (see Figure 4.1). Panel recruitment is defined as the given consent of a respondent to participate in the subsequent online surveys by providing a valid email address. The bases are all respondents of the CATI recruitment interview which finished the interview and were asked the recruitment question. Overall, 51.04% (1356) of the respondents in the analysis consented to be invited to subsequent online surveys. Online participation is defined as having finished at least one out of the eight online surveys. The bases are all respondents that consented to participate in subsequent online surveys and provided an email address. Of the 1356 respondents that provided consent, 61.14% (829) at least finished one online survey.

In order to make sure that the random assignment of experimental and treatment group was successful the distribution of available demographic measures in the two interview groups among the respondents that participated in the telephone interview is compared. There are no significant differences between the two groups concerning the age distribution, the proportion of male respondents, the proportion of respondents that are working, the proportion of respondents that were interviewed on a mobile phone, and the proportion of respondents that were easy to persuade to participate in the telephone survey. Solely the share of respondents with middle education does differ between the two groups. Among the treatment group there are less middle educated respondents in comparison to the experimental group (see Table A 3 in the appendix).

*Figure 5.3: Schematic recruitment process and number of cases that are included in the analysis of the experiment on asking about the income*
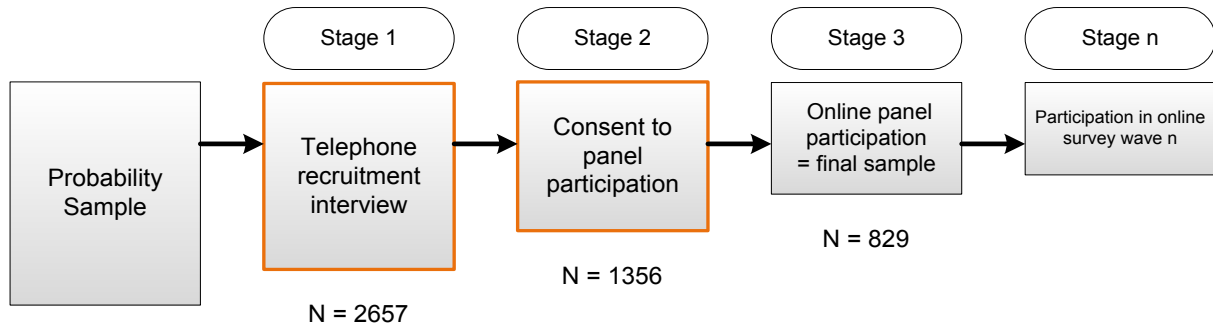


### 5.5.1 Descriptive results panel recruitment probability

First of all, in Table 4.5 the recruitment rates for the treatment and control group are listed. The differences in the groups are marginal, and it indicates no statistically significant difference in the recruitment rates for the respondents that were asked about their income and those that were not asked about it (Chi²(1)= 0.8926).

*Table 5.4: Bivariate relation between income experiment and recruitment*

|  | % recruited | 95% CI | n recruited | n total |
| --- | --- | --- | --- | --- |
| Income not asked | 51.95 | [49.27; 54.63] | 694 | 1336 |
| Income asked | 50.11 | [47.42; 52.81] | 662 | 1321 |
| Total | 51.04 |  | 1356 | 2657 |

*Note*. Pearson chi² (1) = 0.8926 Pr = 0. 345

The separation of the respondents of the treatment group that provided their income, and those that did not provide the income either because they stated to not knowing their income or refused to provide it, showed differences in the recruitment rates for the three groups. The recruitment rate for respondents of the treatment group that provided their income is on the same level (54.03%) as the rate for the respondents of the control group that was not asked about the income (51.95%). For both groups of income nonrespondents, the recruitment rates were lower. Consistent with the assumptions formulated above, respondents that refused to provide their income had the lowest recruitment rate of less than a third (29.38%). The recruitment rate of the respondents that stated to not knowing their income was 40.86% (see Table 5.5).

*Table 5.5: Bivariate relation for the treatment group of income experiment and recruitment*

|  | % recruited | Confidence Interval | n recruited | n total |
|---|---|---|---|---|
| Income provided | 54.03 | [51.04; 57.02] | 577 | 1068 |
| Income refusal | 29.38 | [22.30; 36.45] | 47 | 160 |
| Income do not know | 40.86 | [30.82; 50.91] | 38 | 93 |
| Total | 50.11 |  | 662 | 1321 |

*Note.* Pearson chi² (2) = 37.2510 Pr = 0.000

The pattern is not the same for the two agencies (see Table 5.6). Overall, the recruitment success was higher for the social research agency compared to the market research agency. This fact is mainly attributable to the higher incentive groups that were used in the interviews of study part 1 that was conducted by the social research agency (see Chapter 6). For both agencies, the recruitment rate for income refusers was below 30%. However, the respondents of the social research agency in the "do not know"-group had an almost equivalent recruitment rate compared to the respondents in the control group and those that provided their income. For the market research agency the recruitment rate for respondents that state to not knowing their income were almost as low as for the income refusers. However, the results have to be interpreted with caution, due to the very low numbers of the income nonresponse groups.

*Table 5.6: Bivariate relation for the treatment group of income experiment and recruitment, by agency*

|  | Social research agency | Market research agency |
|---|---|---|
|  | % recruited | % recruited |
| Income not asked | 58.31 (372/638) | 46.13 (322/698) |
| Income asked *Income provided* | 57.78 (297/514) | 50.54 (280/554) |
| *Income refusal* | 29.29 (29/99) | 29.51 (18/61) |
| *Income do not know* | 55.88 (19/34) | 32.20 (19/59) |
| Total | 53.32 (717/1285) | 47.03 (639/1372) |

### 5.5.2 Multivariate results panel recruitment probability

The bivariate results indicate no effect for the treatment of asking about the income. However, it was shown that for the specific group of respondents that refused to provide the income the probability of being recruited is more than 20 percentage points lower than for respondents that provideed the income information or were not asked about their income. Furthermore, the effects were not homogeneous across the two fieldwork agencies. To control for differences between the two agencies, a multivariate logistic model with panel recruitment as the dependent variable is conducted. It takes into account several characteristics of the respondents, the survey (incentive experiment, telephone sample), and controls for the agency (see Table 5.7). The treatment effect of asking about the income is not significant in Model 5.1. As the bivariate analysis indicated differences in the treatment effect for the two agencies, an interaction term between treatment and agency is additionally included in Model 5.2. It shows a negative effect of the treatment, as well as a negative effect of the market

research agency, and an overall positive effect of the interaction term. This means while the recruitment rate is overall lower for the market research agency, the effect of the treatment is smaller for the market research agency in comparison to the social research agency. The odds of being recruited are 0.8 times smaller for respondents that were asked about their income in comparison to respondents that were not asked about their income. The odds ratio as a measure of effect size indicates that the effect can be classified as a small one (Chen, Cohen, & Chen, 2010).

*Table 5.7: Logistic regression on recruitment*

|  | Model 5.1 | Model 5.2 |
|---|---|---|
|  | Odds ratio | Odds ratio |
| Treatment: Income asked | 0.919 | 0.804[*] |
| Agency (0= social research, 1= market research) | 0.796 | 0.703[*] |
| Treatment*agency |  | 1.288[+] |
| Male | 1.296[**] | 1.303[**] |
| Age | 0.982[**] | 0.982[**] |
| Education *(Ref.: low)* |  |  |
|   *middle* | 1.164 | 1.158 |
|   *high* | 1.414[**] | 1.406[*] |
| In paid work | 0.767[**] | 0.766[**] |
| Positive survey attitude | 2.833[**] | 2.825[**] |
| Incentive group |  |  |
|   *2 Euro* | 1.180 | 1.181 |
|   *5 Euro* | 1.325[*] | 1.321[*] |
|   *5 Euro + bonus* | 1.747[*] | 1.749[*] |
|   *10 Euro + bonus* | 2.078[**] | 2.086[**] |
| N | 2657 | 2657 |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules
Robust standard errors (88 clusters in interviewer)
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

To facilitate the interpretation of the results predicted probabilities of being recruited based on Model 5.2 are computed. The predictions include the variables of the model as observed and are separated for the experimental groups and the two agencies. For respondents of the control group that were interviewed by the interviewer of the social research agency the probability of being recruited was 58.31% (95%-CI: 53.77;62.84), whereas for the respondents of the control group it was five percentage points lower (53.32%; 95%-CI: 46.94;59.70). In contrast, respondents that were interviewed by interviewers of the market research agency the recruitment probability does not differ between the control group (46.13%; 95%-CI: 40.22;52.04) and the treatment group (47.03%; 95%-CI: 42.64;51.42).

The bivariate analysis indicated differentiated effects for the respondents of the treatment group, dependent on their answering behavior. Hence, in Model 5.3 the logistic model is calculated including the separated treatment outcome whether a respondent answered the income question, stated to not knowing the income, and refused to answer the question (see Table 5.8). It additionally includes an interaction term between the treatment variable and the fieldwork agency as the bivariate analysis

shows different patterns of the effect for the two agencies. The analysis indicates a significantly negative effect of the treatment in comparison to the control group only for the respondents of the treatment group that refused to provide the income information.

*Table 5.8: Logistic regression on recruitment with separated treatment information (0 = not recruited; 1 = recruited)*

|  | Model 5.3 |
| --- | --- |
|  | Odds ratio |
| Treatment (Ref: control group) |  |
| *Income provided* | 0.948 |
| *Income refused* | 0.324** |
| *Income do not know* | 0.743 |
| Agency (0= social research, 1=market research) | 0.703* |
| Treatment*agency (Ref: income refused*market res. agency) |  |
| *Income provided* | 1.257 |
| *Income refused* | 1.776 |
| *Income do not know* | 0.632 |
| Male | 1.315** |
| Age | 0.982** |
| Education *(Ref.: low)* |  |
| *middle* | 1.142 |
| *high* | 1.387* |
| In paid work | 0.768** |
| Positive survey attitude | 2.711** |
| Incentive |  |
| *2 Euro* | 1.181 |
| *5 Euro* | 1.323* |
| *5 Euro + bonus* | 1.756* |
| *10 Euro + bonus* | 2.069** |
| N | 2657 |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules; 88 clusters in interviewer
+ *p* < 0.10, * *p* < 0.05, ** *p* < 0.01

The predicted probabilities of being recruited are computed for Model 5.3 as well to facilitate interpretation (see Table 5.9). They are computed separately by experimental groups and fieldwork agency. The respondents of the social research agency that refused to provide their income information had a lower recruitment probability compared to the control group, respondents that answered the income question, and also respondents that stated to not knowing their income. In contrast, the respondents of the market research agency of both income nonresponse groups (refusal and do not know) have a lower probability of being recruited compared to respondents of the control group and

respondents that answered the income question. Overall the results have to be interpreted with caution because the numbers in the two refusal groups are low.

*Table 5.9: Predicted probability of being recruited by agency and separated treatment information*

|  | Social research agency (N=1285) | Market research agency (N=1372) |
|---|---|---|
| Income not asked | 0.5831 | 0.4613 |
|  | [0.5381,0.6280] | [0.4023,0.5203] |
| Income asked |  |  |
| *Income provided* | 0.5778 | 0.5054 |
|  | [0.5141,0.6416] | [0.4541,0.5568] |
| *Income refusal* | 0.2929 | 0.2951 |
|  | [0.1937,0.3922] | [0.1845,0.4057] |
| *Income do not know* | 0.5588 | 0.3220 |
|  | [0.4013,0.7164] | [0.2065,0.4376] |

*Note.* Based on multiply imputed data with M=30, applying Rubin's rules; 88 clusters in interviewer; 95% Confidence interval in parentheses

The results described above indicate that fieldwork agency and the related interviewers matter. Most of the interviewers of the academic, social research agency were very experienced ones and specialized in the field of social and political sciences. On the other hand, the focus of the daily work of interviewers of the market research agency was mainly in the area of market research. In Chapter 7 the interplay of the interviewer and experimental treatment is the focus of the analysis.

### 5.5.3   Sample composition at the recruitment stage

To test for selectivity effects that are induced by the experimental treatment, the sample composition of the two experimental groups at the recruitment stage is compared. The two groups are compared on several sets of characteristics that were integrated on an earlier analysis of the same data by Struminskaya (2014, Chapter 3) as predictors of panel participation. Variables that go beyond demographic measures and that are related to online panel participation are introduced to capture differences that are not mirrored by the demographic distribution (Lee, 2006; Schonlau, Soest, & Kapteyn, 2007). These are the frequency of the internet usage, the earlier online survey experience of the respondent and a positive survey attitude of the respondent. Online access panels are often designed as multi-topic surveys, with multiple research objectives. Thus, the relevance of sample composition measure depends on the substantive analytical objective. The two substantive variables life satisfaction and a general measure of trust are included. And finally, characteristics that are related to the survey design (fieldwork agency, type of telephone sample (mobile vs. landline)). The bivariate analysis indicates no significant differences between the recruited respondents of the two experimental groups (see Table 5.10)

*Table 5.10: Sample composition at the recruitment stage*

| | | Income asked | Income not asked | $\chi^2$ |
|---|---|---|---|---|
| Demographic information | Male | 53.17% | 53.75% | 0.0449 |
| | Age (mean) | 40.72 | 41.70 | t=-1.22 |
| | Education | | | 4. 3373 |
| | low | 14.91% | 12.45% | |
| | middle | 29.76% | 34.58% | |
| | high | 55.33% | 52.97% | |
| | In paid work | 68.72% | 71.49% | 1.2528 |
| | Migration background | 6.40% | 5.95% | 0. 1287 |
| | Married dummy | 47.81% | 45.97% | 0. 4675 |
| | Urban area dummy | 39.74% | 39.22% | 2.4098 |
| Characteristics related to (online) survey participation | Internet usage | 92.2% | 90.77% | 0. 9792 |
| | Earlier online survey participation | 30.25% | 25.73% | 3.5088[+] |
| | Positive survey attitude | 93.75% | 93.70% | 0.0124 |
| Substantial variables | Life satisfaction (mean) | 7.49 | 7.42 | t=0.79 |
| | Trust | | | 1.6496 |
| | no | 18.43% | 15.89% | |
| | yes | 21.78% | 23.54% | |
| | depends | 59.79% | 60.56% | |
| Survey design | Mobile | 46.52% | 44.24% | 0.7165 |
| | Agency | 47.89% | 46.40% | 0.3009 |
| | N | 662 | 694 | |

*Note.* Comparison of groups with Chi²-test for the categorical variables and t-test for age and life satisfaction
Coefficients based on multiply imputed data with M=30, applying Rubin's rules
+ p < 0.10, * p < 0.05, ** p < 0.01

In the multivariate analysis of the sample composition, the share of middle educated respondents differs significantly on the 5%-level between the two groups. There are fewer respondents with middle education among the recruited persons of the treatment group.

The respondents of the two groups do not differ on the included survey-specific characteristics, the substantial variables, as well as the variables of survey design. On the basis of the set of included characteristics, it can be concluded that asking about the income in comparison to not asking about the income does not introduce selectivity into the sample at the stage of recruitment.

The comparison of the respondents of the treatment group that provided their income information with those that refused to provide the information also does not reveal major differences. The only significant difference is related to the interviewer rated willingness to participate in the telephone survey, a proxy for general survey attitude: the income nonrespondents are rated as being less willing to participate in the telephone survey.

*Table 5.11: Sample composition recruitment stage: Logistic regression on*
*experimental variable income asked*
*(0 = income not asked; 1 = income asked)*

|  |  | Odds ratio |
|---|---|---|
| Demographic information | Male | 0.963 |
|  | Age | 0.993+ |
|  | Education *(ref.: low)* |  |
|  | *middle* | 0.678* |
|  | *high* | 0.797 |
|  | In paid work | 0.878 |
|  | Migration background | 1.000 |
|  | Married dummy | 1.227+ |
|  | Urban area dummy | 1.030 |
| Characteristics related to (online) survey participation | Internet usage | 1.140 |
|  | Earlier online survey participation | 1.222 |
|  | Positive survey attitude | 0.980 |
| Substantial variables | Life satisfaction | 1.012 |
|  | Trust (ref.: no) |  |
|  | *yes* | 1.191 |
|  | *depends* | 0.953 |
| Survey design | Mobile | 1.067 |
|  | Agency | 1.068 |
|  | N | 1356 |

*Note.* Coefficients based on multiply imputed data M=30, applying Rubin's rules; 84
clusters in interviewer
+ p < 0.10, * p < 0.05, ** p < 0.01

## 5.6 Summary and discussion

With a split-half experiment on asking about the income in a recruitment interview, the study gave the opportunity to obtain insights into the causal mechanism of income refusal and lower probability for subsequent survey participation. Two research questions lead the analysis: 1) whether the inclusion of the income question negatively affects the recruitment probability (Research question 5.1), and, 2) whether asking about the income versus not asking about it affects the sample composition of the resulting sample on the recruitment stage (Research question 5.2). Regarding Research question 5.1, two hypotheses were formulated. Overall, it was shown that the respondents of the treatment and experimental groups did not differ in their probability of being recruited. Hypothesis 5.1 that assumed an overall negative effect of asking about the income needs to be rejected. In Hypothesis 5.2 the assumption is formulated that asking about the income only negatively affects the recruitment probability of those who refused to provide the income information. Overall, about 20% of the treatment group did not provide the income information. The analyses showed that the group of respondents that refused to provide the income had a lower probability of consenting to join the panel

as well. For both fieldwork agencies, only a third of them provided consent in contrast to over 50% in the control group. It can be concluded that not the income question as such decreases the overall recruitment probability, but respondents that refuse to provide the income information have a lower probability of being recruited. Hence, income nonresponse can be used as an indicator for reluctant respondents. This knowledge can be applied to design a responsive recruitment strategy for reluctant respondents in future recruitment studies. As possible measures to increase the recruitment probability of respondents with privacy issues, the interviewer could stress confidentiality issues and the trustworthiness of the institution.

The composition of the recruited sample of the treatment and the control group were compared on several measures (Research question 5.2). The two groups did not systematically differ with regard to demographic characteristics, survey-specific characteristics, substantial variables, and characteristics of the survey design. Also, the small group of respondents that refused to provide the income information differed only regarding their overall survey attitude compared to the respondents in the experimental group that provided the income information. This finding underlines the interpretation of income nonrespondents as being the more reluctant respondents.

Even though the experimental data showed that the inclusion of one income question in a recruitment interview does not negatively affect the recruitment probability, the study suffers from some shortcomings. First, only one question about the household income was included, and second, it provided bracketed answer options. It needs to be further investigated if the manipulation was sufficiently powerful (Mutz, 2011). Bracketed information was shown to be an effective measure to reduce income item nonresponse. It aims to reduce the burden of answering income questions (Juster & Smith, 1997). Asking about the income by means of an open-ended question or more extensively was not tested.

The experiment was a first step in investigating the interrelation between topic-specific item nonresponse and next wave nonresponse. The successful establishment of probability-based online panels as vehicles of high-frequency data collection provides the opportunity for subsequent research. The connection between item and unit nonresponse in high-frequency panels needs to be tested. In this context the effect of additional types of sensitive questions, such as, for instance, questions about racial prejudice or deviant behavior on next wave nonresponse could be tested. Furthermore, not only sensitivity is an issue in the context of high-frequency data collection but also the effect of repeated topics. In access panels where external researchers submit their questionnaire modules, the panel provider has limited control of topics. In consequence, several consecutive waves can include questions that are perceived as being similar and redundant by the respondents. This challenge, that is specific to access panels, opens the opportunity for a new thread of research that investigates the interrelation of topic-specific nonresponse and unit nonresponse in the subsequent panel waves.

# 6. Study 3: Effect of incentives on recruitment probability, online participation, and sample composition

## 6.1 Introduction and overview

The third factor of the recruitment interview design is the amount of incentive that is promised and paid for online survey participation. In contrast to the two factors "recruitment interview length" and "sensitive question", that are assumed to contribute to the burden of the interview, the incentive is assumed to increase the perceived benefit of survey participation (Groves et al., 2000). Monetary incentives are a universal measure aimed at fostering cooperation in surveys (Singer et al., 1999). In the context of the multistep recruitment process for probability-based online panels incentives are one of the measures that are used in order to minimize nonresponse at several stages (Blom et al., 2015; Martinsson & Riedel, 2015; Rao et al., 2010; Scherpenzeel & Toepoel, 2012). In self-administered studies, incentives represent a large share of the survey costs, though the decision upon an incentive scheme in longitudinal survey needs to be well-considered (Laurie & Lynn, 2009). Nevertheless, only a few studies experimentally examined the effect of incentives on cooperation during the recruitment process of a probability-based online panel so far. They mainly focused on the effect of providing a promised versus a prepaid incentive for participation in the recruitment interview and panel registration (Blom et al., 2015; Martinsson & Riedel, 2015; Rao et al., 2010; Scherpenzeel & Toepoel, 2012). The effect of promising an incentive for subsequent survey participation during the recruitment interview was, to the best of my knowledge, hitherto not tested.

The experiment described in the present chapter examines, first, the effect of promising an incentive for subsequent online survey participation during the telephone recruitment interview on the recruitment probability and the probability of online survey participation. The amount of the promised incentives was experimentally varied between 0 Euros, 2 Euros, 5 Euros, 5 Euros with bonus of 20 Euros for participation in all eight online surveys, and 10 Euros with a bonus. Secondly, the effect of the varying incentive amounts on the resulting sample composition at the stage of recruitment and the stage of online participation is tested.

In the subsequent section, I briefly sketch the general function of incentives and the use of incentives among probability-based online panels. I summarize pertinent studies on the effects of incentives and elaborate on the addressed research questions and hypotheses. This part is followed by the description of the data and the survey experiment. In the analyses, I focus on the effect of varying amounts of promised incentives on the two outcome variables – panel recruitment and panel participation–, as well as on sample composition bias. In the last section, I discuss the results and the implications for survey design.

## 6.2  Research on incentives

### 6.2.1  General function of incentives

Incentives are a widely used measure to increase response rates of surveys (Singer et al. 1999). Distinctions are made between prepaid incentives and promised (contingent) incentives (Church, 1993) as well as monetary and non-monetary incentives (e.g., lotteries, chocolate, or result summaries). Prepaid incentives are paid in advance, independently of any effort or action by the potential respondent as a physical token of appreciation (Dillman et al., 2014). Promised incentives are offered in advance of a survey and paid for successful survey participation.

Several studies revealed that the response rate increases with higher amounts of incentives across different modes (Göritz & Neumann, 2016; Mercer et al., 2015; Singer et al., 1999; Singer & Ye, 2013).

Several studies showed that monetary incentives are more effective than non-monetary incentives in interviewer-administered surveys (Singer et al., 1999) as well as in mail surveys (Church, 1993). As a consequence, general population studies are using recently mainly monetary incentives. In the most recent meta-analysis, Mercer et al. (2015) exclusively included population studies that used monetary incentives and excluded those that used non-monetary incentive strategies.

Several meta-analyses came to the conclusion that prepaid incentives are more effective in increasing response rates in comparison to promised incentives. The increase in response rate was shown to be highest in mail surveys (Church, 1993; Edwards et al., 2002; Mercer et al., 2015) compared to other survey modes. Among mail surveys prepaid incentives performed significantly better compared to promised incentives in telephone surveys (Edwards et al., 2002), whereas Mercer et al. (2015) did not find a significant difference between prepaid and promised incentives in interviewer-administered, personal interviews. The usefulness of incentives in web surveys was proven by a meta-analysis as well (Göritz, 2006). The effect was smaller compared to the traditional survey modes. The majority of the online studies included in the meta-analysis used a lottery incentive as a promised incentive for survey participation. The use of lottery incentives, that were proven not to be effective in interviewer-mediated surveys (Singer & Ye, 2013), may be one possible explanation for the small incentive effect in web surveys in comparison to other survey modes (Göritz, 2006).

In the context of web surveys paying incentives in advance is difficult if respondents are invited via the web (Couper 2002). For this reason, the meta-analysis by Göritz (2006) included only studies that were using promised incentives. As one possible solution to deliver prepaid incentives in web surveys as well, the intermediary service PayPal was used. Several studies revealed that PayPal is not an adequate substitute for real money (Bosnjak & Tuten, 2003; Göritz & Neumann, 2016). The scientific online panels that are using incentives have adopted the approach of the respondents' account that is

common in market research online access panels[6]. A promised amount of money or bonus points that are transferable into money is credited to the respondent's account after questionnaire completion. The respondents are then able to redeem the money using bank transfer or vouchers for online shops (e.g., GIP[7], LISS; Blom, Bosnjak, et al. 2016). Whether respondents perceive the money on their virtual account as equivalent to real money was not tested so far.

In a review of incentive experiments in Germany, Pforr et al. (2015) showed that the incentive effects that were mainly found in the US and Great Britain are also applicable to the German survey environment.

### 6.2.2 Incentives during the recruitment process for probability-based online panels

During the multistep recruitment process for probability-based online panel surveys, incentives are used as one response enhancing measure. The incentive scheme of the recruitment process is often characterized by a combination of different incentives at different stages, to minimize the overall nonresponse. Table 6.1 gives a brief overview of the incentive schemes for selected European panels across several stages. Some of the incentive schemes were derived from experimental studies. In the field pretest for the initial sample of the LISS panel, for instance, the optimal amount and timing of incentive were tested experimentally. Scherpenzeel & Toepoel (2012) showed that consistent with previous meta-analytic studies (Mercer et al., 2015; Singer & Kulka, 2002) prepaid incentives were more effective than promised incentives in the context of panel recruitment as well. In contradiction to their assumptions, it did not make a difference whether the advance letter included 10, 20, or 50 Euros. As a consequence, the 10 Euros prepaid cash incentive was judged as being the most cost effective one. The pretest additionally revealed that the highest share of respondents dropped out between expressed panel consent and registration (step 3 in Figure 6.1) (Scherpenzeel & Toepoel, 2012). On the basis of the empirical evidence, it was decided to provide an additional promised incentive of ten Euros for online registration. The measure showed to be successful in decreasing the nonresponse rate at the registration stage (van der Laan, 2009).

Sending an advance letter that included an incentive was proven to be helpful in several other studies as well (e.g., Rao et al. 2010). For the German Internet Panel, Blom et al. (2015) showed that a 5 Euros prepaid incentive is more successful in terms of recruitment than a 10 Euros promised incentive that was announced in the advance letter. For the same study, Krieger (2016) additionally showed that including a 5 Euros note in the first reminder letter for online registration significantly increases the registration rate of the respondents that provided consent to panel participation in the recruitment interview. Furthermore, the incentive shortened the time needed for the registration process, and it was assessed as being cost effective with regard to the total recruitment costs. Nonmonetary incentives, for instance, a lottery incentive that was sent along with the postal invitation

---

[6] See for example the Mingle Respondi Panel
https://mingle.respondi.com/faq.php?chosen_site=3&SES=112d967df40518d8d441dcea2a1a821f&frmnd=faq
[7] http://www.gesellschaft-im-wandel.de/unser-dank/ihr-dankeschoen/

revealed to have a positive effect in the case of postal recruitment of the Swedish Citizen Panel (Martinsson & Riedel, 2015), as well as the Norwegian Citizen Panel (Hogestol & Skjervheim, 2013). In contrast, a small piece of quality chocolate as a symbolic prepaid incentive had no effect on the response rate (Hansen, 2007).

Several panels reward the survey participation in the regular waves with incentives as well. If mentioned during the recruitment process, it may also positively affect the decision to give consent to panel participation. The LISS panel, for instance, promises 15 Euros per hour survey participation, the German Internet Panel 5 Euros per survey. In both cases, the incentives are credited to the respondent's account that is accessible online. Respondents are offered several ways to redeem the incentive either by bank transfer, Amazon voucher, or charity donation (Blom et al., 2015). The GESIS panel includes a 5 Euros bill in the invitation letter of the regular panel waves (Bosnjak et al., 2017). In a quasi-experiment, it was revealed that switching from the prepaid incentive scheme to a postpaid incentive scheme between the first and second self-administered survey wave decreased the response rate substantively (Schaurer & Bosnjak, 2016). Hence, based on empirical evidence it was decided to use the prepaid incentive also for the regular panel waves. In contrast, the Swedish (Martinsson & Riedel, 2015) and the Norwegian Citizen Panels (Hogestol & Skjervheim, 2013) both do not pay their respondents for survey participation.

*Table 6.1: Incentive schemes of selected online panels across stages*

| | Recruitment interview | Registration/ first online interview | Regular panel wave | Additional incentives |
|---|---|---|---|---|
| **LISS Panel, Netherlands** | €10 prepaid in advance letter | €10 promised for registration | €15 per hour of interview time promised on respondents' account → bank transfer | No incentive |
| **German Internet Panel, Germany** | Experiment: €5 prepaid in advance letter vs. €10 promised for recruitment interview | Experiment: €5 prepaid incentive in first invitation reminder letter €5 promised for online welcome interview on respondents' account | €4 on respondents' account → bank transfer, amazon voucher, charity donation | Bonus of €5 promised for participation in all but one survey within a year; Bonus of €10 for participation in all surveys within a year |
| **ELIPSS pilot study, France** | Experiment: €10 prepaid incentive in advance letter vs. no incentive | No incentive | Tablet PC for the time of panel participation as incentive | No incentive |
| **Swedish Citizen Panel, Sweden** | *Postal recruitment, no interview* | Experiment: lottery incentive in postal invitation vs. no incentive Experiment: lottery incentive in reminder letter vs. no incentive | No incentive | No incentive |
| **GESIS Panel, Germany** | €10 promised for participation | €5 prepaid in invitation letter | €5 prepaid in invitation letter | No incentive |
| **Norwegian Citizen Panel, Norway** | *Postal recruitment, no interview* | Gift card for travel lottery in invitation letter, contingent on panel registration | No incentive | No incentive |

*Note.* Source: Adaption and extension of two tables from Blom et al. (2015)

The available evidence on the use of incentives in probability-based online panels was mainly based on personal or mail recruitment where addresses of the target persons were available. Rao et al. (2010) showed that sending a monetary incentive in the advance letter of a telephone recruitment interview for the Gallup Panel increased the cooperation rate as well. The incentive had no effect on the overall panel consent rate. In their study, Rao et al. (2010) did only include respondents with listed phone numbers and corresponding addresses. The authors admit that the exclusion of parts of the general population without listed address may have introduced bias. In cases where a random digital dialing sampling method (Gabler & Häder, 2002) is used as the base for the telephone recruitment, addresses are not available. Hence, sending a prepaid incentive to target persons is impossible. In consequence, optimal strategies for the use of promised incentives need to be examined.

### 6.2.3 Incentives and sample composition

Compared to a large number of experimental studies on the effect of incentives on response rates, only a few studies quantify the nonresponse bias that is introduced by the use of incentives (Singer & Ye, 2013, p. 113). Incentives are assumed to not only increase the response rate, but may also motivate respondents with different characteristics than those that would participate without the use of incentives (Ryu, Couper, & Marans, 2006; Singer et al., 2000). As an intended consequence of incentives, some authors mention the possible correction of nonresponse bias by attracting groups that are often underrepresented in general population surveys. These are, for instance, minorities or low socio-economic status groups (Singer & Kulka, 2002) and respondents with a low interest in the topic of the survey (Groves et al., 2004).

The majority of studies that focused on nonresponse bias that is introduced by the use of incentives did not find differences concerning the sample composition. In several reviews of experimental studies on the incentive effect on the resulting sample, Singer and colleagues (Singer & Kulka, 2002; Singer et al., 1999; Singer & Ye, 2013) reported mixed results. Singer et al. (1999) refer to three studies that revealed an incentive effect on nonresponse bias, while five studies did not find any effects, and one study came to mixed results. Singer & Kulka (2002) cited few experimental studies that showed incentives being able to bring groups into the sample in interviewer-mediated surveys that are often underrepresented in surveys.

For telephone surveys, Cantor et al. (2008) concluded that there are no major effects on sample composition introduced by incentives. In contradiction, Singer et al. (2000) reported for a random digital dial survey that lower educated respondents were more likely to be recruited if they received a $5 prepaid incentive as opposed to the group that did not receive any incentive. Laurie & Lynn (2009) revealed mixed results for the effect of incentives on panel retention bias. In a study on $5 versus $10 prepaid incentives in a telephone interview, Curtin, Singer, & Presser (2007) did not find any differences among sex, income, and ethnicity of the respondents. However, the sample was small, and the authors did not include a control group that received no incentive. Lipps (2010a), as well, reported no effect of varying incentive amounts on sample composition of the telephone interviews of the Swiss Household Panel.

In an overview of several incentive experiments in German general population face-to-face surveys, Pforr et al. (2015) also reported mixed results for the effect of incentives on sample composition. They did not find any differences in the mean age and gender ratio. They reported differences on some variables that are not consistent across the studies under consideration. Overall, the authors concluded that the incentive effect on sample composition is minor and does not follow a systematic pattern. The comparison of the sample composition between respondents of the German part of Survey of Health, Age, and Retirement (SHARE) that received a prepaid incentive and the control group that did not receive any incentive showed only significant age differences (Börsch-Supan, Krieger, & Schröder, 2013). The authors found that the proportion of younger respondents (between 50 and 55 years) came

closer to the sampling frame compared to the no-incentive group (Börsch-Supan et al., 2013). Furthermore, there was some indication that the mean income was lower among the incentive group. Very few studies examined the nonresponse bias introduced by incentives among web surveys. For instance, Göritz (2004b) and Marcus et al. (2007) did not find any differences among the respondents of different incentive groups with respect to sample composition. Marcus et al. (2007) revealed a positive effect for topic salience on the response rate, but no interaction with the amount of incentive. This result is in line with the results of Groves et al. (2004), that showed for face-to-face surveys that incentives did not compensate for low topic interest as it was assumed.

## 6.3    Research questions and hypotheses

On the basis of the above-outlined previous research, several factors can be varied when deciding upon an optimal incentive strategy for a given survey. First of all, researchers have to decide whether the incentive is paid unconditionally of survey participation in advance, or conditional on survey participation as a promised incentive. Next, researchers have to decide whether to use a monetary or a non-monetary incentive. Based on the previous two decisions the optimal amount of incentive per respondent and survey under given financial constraints has to be defined. Finally, in the context of repeated surveys, researchers have to decide whether to pay an extra bonus for loyal respondents or not.

However, not all of the four factors need to be experimentally tested to derive an optimal incentive strategy. The payment of prepaid incentives requires the availability of the respondents' address information, which is not available in a telephone sample based on an RDD-design. The empirical evidence for the use of monetary instead of non-monetary incentives was proven in various studies.

The study presented in this chapter contributes to the incentive literature in several ways: First, it tests the usability of promising incentives in telephone recruitment interviews for a probability-based online panel where no address of the respondents is available. Second, it examines the effect of promising an incentive for subsequent online survey participation on the recruitment probability and the panel participation rate. The amount of the promised incentive was experimentally varied between 0 Euros, 2 Euros, 5 Euros, 5 Euros with a bonus of 20 Euros for participation in all eight online surveys, and 10 Euros with a bonus. Third, it examines the effect of the varying incentive amounts on the resulting sample composition at the stage of recruitment, and the stage of online participation.

Overall, three research questions lead the analyses of the chapter:

Research question 6.1:    How does the amount of promised incentive for online survey participation affect the recruitment probability and the probability of online survey participation?

Research question 6.2: Does adding a bonus for loyal respondents versus not adding a bonus on top of the promised incentive have a positive effect on the recruitment probability and the probability of online survey participation compared to the incentive conditions without a bonus?

Research question 6.3: Do the different amounts of promised incentive for online survey participation affect the composition of the resulting sample?

The hypotheses are formulated separately for the two stages recruitment and online participation that represent the two subsequent stages of the recruitment process. The following set of hypotheses refers to the first part of the recruitment process, the consent to participate in subsequent surveys.
According to the above-summarized studies, the following directed hypothesis is formulated regarding Research question 6.1:

Hypothesis 6.1: The higher the amount of incentive that is promised for subsequent online survey participation, the higher the recruitment probability of a respondent.

Regarding Research question 6.2, I hypothesized that the additional bonus increases the overall expected value of the incentive. Thus, the promised incentive including a bonus are expected as being a higher incentive in comparison to the same amount without additional bonus and thus will increase the recruitment rate.

Hypothesis 6.2: The recruitment probability is higher for respondents of the incentive groups with an additional bonus for loyal respondents compared to respondents in the incentive groups without bonus.

According to the above-outlined studies, I assume that a higher amount of promised incentive attracts groups that tend to be underrepresented in surveys of the general population (Research question 6.3). Groups that tend to be underrepresented are, for instance, respondents with low education, respondents with a lower household income and respondents that are not in paid work. For the GESIS Online Panel Pilot data, Struminskaya et al. (2014) showed that man had a higher probability of giving consent to online participation than women. This result can partly be explained by higher internet affinity of man and a more frequent use of the internet at the time of data collection. Accordingly, I expect the increasing amount of incentive can partly offset the lower affinity towards online surveys of women. Finally, I assume higher incentives to compensate for reluctance towards surveys in general.

Hypothesis 6.3: With increasing amount of promised incentive the share of respondents that are typically underrepresented in surveys of the general population increases at the stage of recruitment.

The following set of hypotheses refers to the second part of the recruitment process, the participation in the online surveys after having expressed the willingness to do so.

As receiving an incentive is directly related to survey participation and the amount of incentive was mentioned in each email invitation, I assume to find an additional effect of the amount of promised incentive at the stage of survey participation as well.

According to the above-summarized studies, the following directed hypothesis is formulated in regarding Research question 6.1:

Hypothesis 6.4:    The higher the amount of incentive that is promised for subsequent online participation, the higher the online participation probability of a respondent.

Furthermore, I assume that the additional bonus increases the expected value of the incentive and increases the online participation probability as well.

Hypothesis 6.5:    The online participation probability is higher for respondents of the incentive groups with an additional bonus for loyal respondents compared to respondents in the incentive groups without bonus.

Also, I expect the promised bonus for loyal respondents motivates respondents to participate in each of the eight surveys. After having taken part at least once, they were able to log-in to their respondent's account and redeem the incentive. Hence, after having made a positive experience, the bonus may additionally foster the motivation to participate.

Hypothesis 6.6:    Adding a bonus for loyal respondents increases the share of panelists that participate in all eight surveys compared to respondents in the incentive groups without bonus .

In line with the argumentation at the stage of recruitment, I assume that a higher amount of incentive attracts groups that tend to be underrepresented in surveys of the general population at the stage of online participation as well.

Hypothesis 6.7:    With increasing amount of promised incentive the share of respondents that are typically underrepresented in surveys of the general population increases at the stage of panel participation.

Before analyzing the data and testing the hypotheses, in the next section, I describe the experimental design and the data used for the analyses.

## 6.4 Method

### 6.4.1 Experimental design

The optimal amount of incentive and the use of a bonus for loyal respondents are the focus of the experiment. Overall, four incentive groups were tested (0 vs. 2 vs. 5 vs. 10 Euros). Additionally, a bonus of 20 Euros was introduced for respondents who participated in all eight online surveys.

A fractional design with the following groups was implemented:

- Condition 1: 10 Euros + 20 Euros bonus for completing all eight online surveys
- Condition 2: 5 Euros + 20 Euros bonus for completing all eight online surveys
- Condition 3: 5 Euros
- Condition 4: 2 Euros
- Condition 5: no incentive

The incentive groups were randomly assigned to the telephone numbers before dialing. The amount of incentive for online participation was stated at the end of the recruitment interview after the interviewer had asked for consent to participate in subsequent online surveys and the respondent provided an email address. Interviewers were kept blind to the incentive group until they reached the end of the interview where they introduced the panel. They were instructed to describe the online panel and stated the amount of incentive respondents would receive as a result of participating in each online survey. In addition to the incentive amount, interviewers were instructed to describe the procedure of incentive delivery on request. In the invitation email for each online survey, the respondents were informed about the amount of promised incentive.

After each finished online survey, the amount of incentive was credited on respondents' online user account. To redeem the incentive respondents had to actively log-in and request either a voucher for an online trader (Amazon), donate the money to a list of organizations or provide information about their bank account.

The respondents of the initial telephone recruitment interview did not receive any incentive for participation.

To test the optimal combination in a given survey context a fully crossed experiment of all factors would be the optimum. However, under given constraints of time and budget, a fully crossed design was not applicable. Due to the pilot character of the study, a step-wise approach was applied.

Table 6.2 provides a brief overview of the relevant fieldwork settings and the number of respondents within each group for the incentive experiment. Both fieldwork agencies implemented the experiment. However, only the academic, social research agency conducted interviews with all five experimental groups.

*Table 6.2: Fieldwork settings incentive experiment*

| Incentive group | Social research agency | | Market research agency | |
| --- | --- | --- | --- | --- |
| | Study part 1 | Study part 2 | Study part3 | Row total (%) |
| 0€ | – | 210 | 541 | 751 (26.18) |
| 2€ | – | 242 | 517 | 759 (26.46) |
| 5€ | – | 208 | 511 | 719 (25.05) |
| 5€ + bonus | 324 | – | – | 324 (11.29) |
| 10€ + bonus | 316 | – | – | 316 (11.01) |
| Column total | 640 | 660 | 1569 | 2869 (100.00) |

## 6.4.2 Data

Overall 2869 respondents are part of the analysis. 1300 respondents were interviewed by interviewers of the social research agency, and 1569 by interviewers of the market research agency. The data include interviews of full length, as well as short interviews. 553 respondents that were asked the recruitment question without interview are excluded from the analyses because the only available information about them is their sex, that was assessed by the interviewers.

The analyses are based on 30 multiply imputed datasets; the estimates are pooled following Rubin's rules (Rubin, 1987). In all analyses, the clustering of the data within 88 interviewers is taken into account, by calculating robust standard errors.

Table 5.2 shows the frequencies for the three dependent variables of the analyses for the sample of analysis. Overall about half of the 2869 respondents that finished the recruitment interview, provided consent to subsequent online survey participation. Of the respondents that provided consent, overall 60.72% participated in at least one online survey. 34.05% of the recruited respondents participated in all eight online surveys.

*Table 6.3: Frequencies of dependent variables for the analysis sample*

| | % | n |
| --- | --- | --- |
| Recruitment | 50.05 | 2869 |
| Online participation, if recruited | 60.72 | 1463 |
| Participation in all 8 surveys | 34.05 | 1463 |

The distributions of the variables included as control variables in the multivariate models and as indicators in the analyses of the sample composition are presented in Table 6.11. The majority of interviews were conducted as regular ten minutes interviews. About 58% of the interviews were conducted on the landline, and 42% of the interviews were conducted via cell phone. The mean age in the sample is 43 years. 51% of the respondents were male. And almost half of the respondents were highly educated. 72% of the respondents were in paid work and a quarter of the respondents lived in a household with a low household income. The household income is available for only half of the sample because of the experimental split that was part of the analysis of Chapter 5. As a proxy measure for a positive survey attitude, the interviewers' judgment on how easy it was to persuade the

respondent to participate in the telephone interview is used. The vast majority of 88.94% of the respondents had a positive attitude towards surveys.

*Table 6.4: Frequencies of independent variables for the analysis sample at the interview stage*

|  | % | n |
|---|---|---|
| Incentive |  | 2869 |
| *2 Euros* | 26.18 |  |
| *5 Euros* | 26.46 |  |
| *5 Euros + bonus* | 25.06 |  |
| *10 Euros + bonus* | 11.29 |  |
| Type of interview |  | 2869 |
| *regular interview* | 92.89 |  |
| *short interview* | 7.11 |  |
| Sample |  | 2869 |
| *landline sample* | 57.96 |  |
| *mobile sample* | 42.04 |  |
| Age | 43.31 | 2869 |
| Male | 51.03 | 2869 |
| Education |  | 2869 |
| *low* | 16.40 |  |
| *middle* | 34.11 |  |
| *high* | 49.49 |  |
| In paid work | 72.06 | 2869 |
| Low household income | 23.00 | 1321 |
| Positive survey attitude | 88.94 | 2869 |

## 6.5 Results

In the following analyses, the amount of promised incentive for online survey participation serves as the central independent variable. Dependent variables are the three binary variables that are shown in Table 6.3 above: panel recruitment, online participation, and participation in all eight surveys.

For the analyses of the sample composition, the selection bias across experimental groups for several measures is calculated and compared.

The analysis of the incentive experiment follows the logic of the multi-step recruitment process (see Figure 6.1).

*Figure 6.1: Schematic recruitment process and number of cases that are included in the analysis*
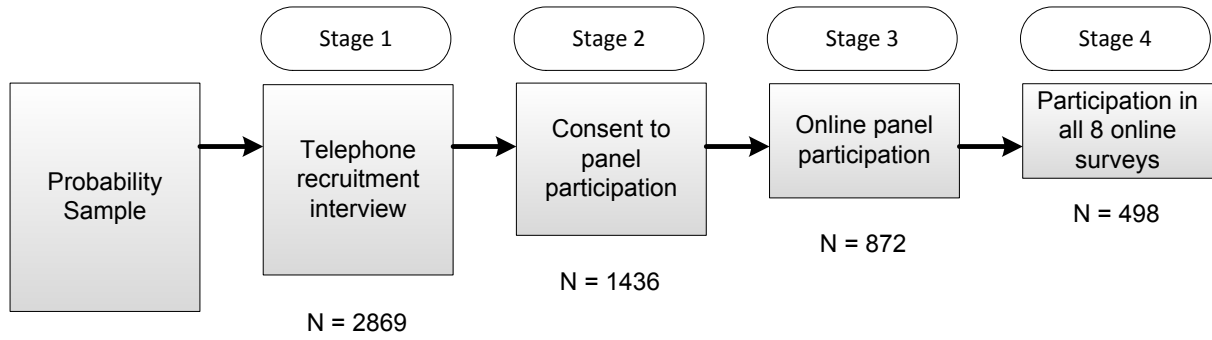


As a first step, the sample composition of the experimental groups is compared at the stage of the telephone interview to test the effectiveness of the randomization. The interviewers were not aware of the incentive group at the beginning of the interview. Hence, they could not mention the amount of promised incentive while gaining cooperation and I do not expect systematic differences between the experimental groups. Regarding the characteristics age, gender, education, being in paid work, low household income, positive survey attitude, and mobile sample no systematic differences were found. The exact numbers can be found in the appendix Table A 4.

### 6.5.1 Panel recruitment probability

In the following analysis, I investigate the effect of the five incentive groups on panel recruitment probability, that is the transition from stage 1 to stage 2 (see Figure 6.2). The hypotheses formulated in section 6.3 are tested.

*Figure 6.2: Selection step panel recruitment*
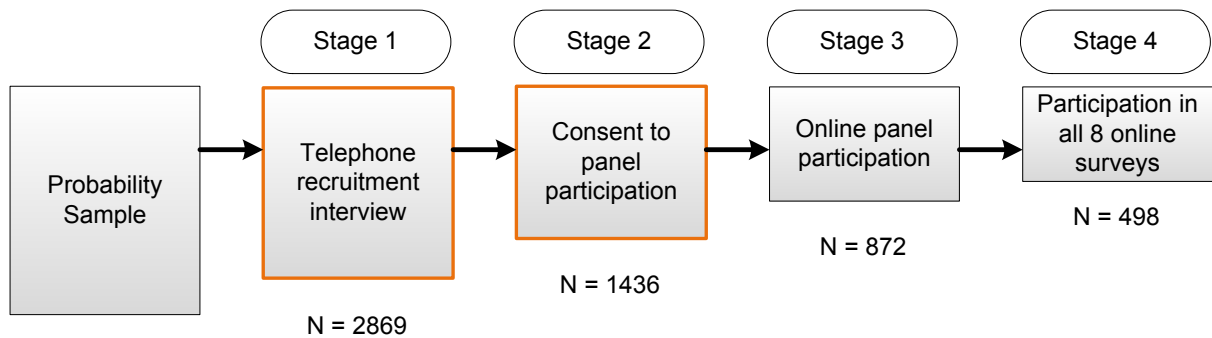


### 6.5.1.1 *Descriptive results*

Table 6.5 gives an overview of the bivariate relation between incentive group and the outcome variable recruitment. As a baseline, about 42% of the respondents of the control group that were not promised any incentive provided consent to subsequent online participation and provided an email address. The share of recruited respondents increased with a higher incentive amount. It increased to a

maximum of 62.66% in the group that was promised 10 Euros per survey plus bonus. The difference between the group that was promised 5 Euros with bonus and the 10 Euros group with bonus is marginal. In contrast, the difference between the five Euros groups with and without bonus reaches about 10-percentage points. The additional bonus results in a significant increase of the recruitment probability.

*Table 6.5: Bivariate relation between incentive group and recruitment*

|  | % recruited | 95% CI | n recruited | n total |
|---|---|---|---|---|
| 0 Euros | 42.34 | [45.54, 52.04] | 318 | 751 |
| 2 Euros | 47.69 | [49.77, 56.24] | 362 | 759 |
| 5 Euros | 50.49 | [52.33, 58.84] | 363 | 719 |
| 5 Euros + bonus | 60.19 | [55.17, 65.27] | 195 | 324 |
| 10 Euros + bonus | 62.66 | [57.98, 68.27] | 198 | 316 |
| Total | 50.05 |  | 1436 | 2869 |

*Note.* Pearson chi² (4) = 52. 9865 Pr = 0.000

### 6.5.1.2 *Multivariate results*

The bivariate results indicate a positive effect of the increasing amount of incentive at the stage of recruitment. To control for features of the survey and possible selectivity on the previous stage of participation in the telephone survey available additional information is included in the multivariate analysis. A logistic regression with panel recruitment as the dependent variable is conducted. The central independent variable is the experimental incentive group. As control variables, the fieldwork agency, sex, age, education, and employment status of the respondent is included. Additionally, a dummy variable on the respondent's survey attitude and the type of interview (short vs. regular) is included (see Table 6.6). The logistic regression model confirms the positive effect of the increasing incentive amount, also when controlling for several variables.

*Table 6.6: Logistic regression on recruitment*

|  | Odds ratio |
|---|---|
| Incentive |  |
| *2 Euros* | $1.238^{+}$ |
| *5 Euros* | $1.402^{**}$ |
| *5 Euros + bonus* | $1.829^{*}$ |
| *10 Euros + bonus* | $2.106^{**}$ |
| Agency *(Ref. social research agency)* | 0.805 |
| Male | $1.288^{**}$ |
| Age | $0.982^{**}$ |
| Education *(Ref.: low)* |  |
| *middle* | 1.154 |
| *high* | $1.461^{**}$ |
| In paid work | $0.789^{*}$ |
| Positive survey attitude | $2.878^{**}$ |
| Short interview | $0.724^{*}$ |
| N | 2869 |

*Note.* Coefficients based on M=30, applying Rubin's rules
88 clusters in interviewers
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

To foster the interpretation, the predicted probabilities of being recruited are calculated for each incentive group based on the logistic regression model above. The control variables of the logistic model are included as observed in the data. Figure 6.3 illustrates an increasing probability of being recruited with increasing incentive amount. However, the confidence intervals are overlapping. The differences are statistically significant for the control group compared to the 5 Euros plus bonus and the 10 Euros plus bonus groups.

*Figure 6.3: Predicted probability of being recruited by incentive group (n=2869)*
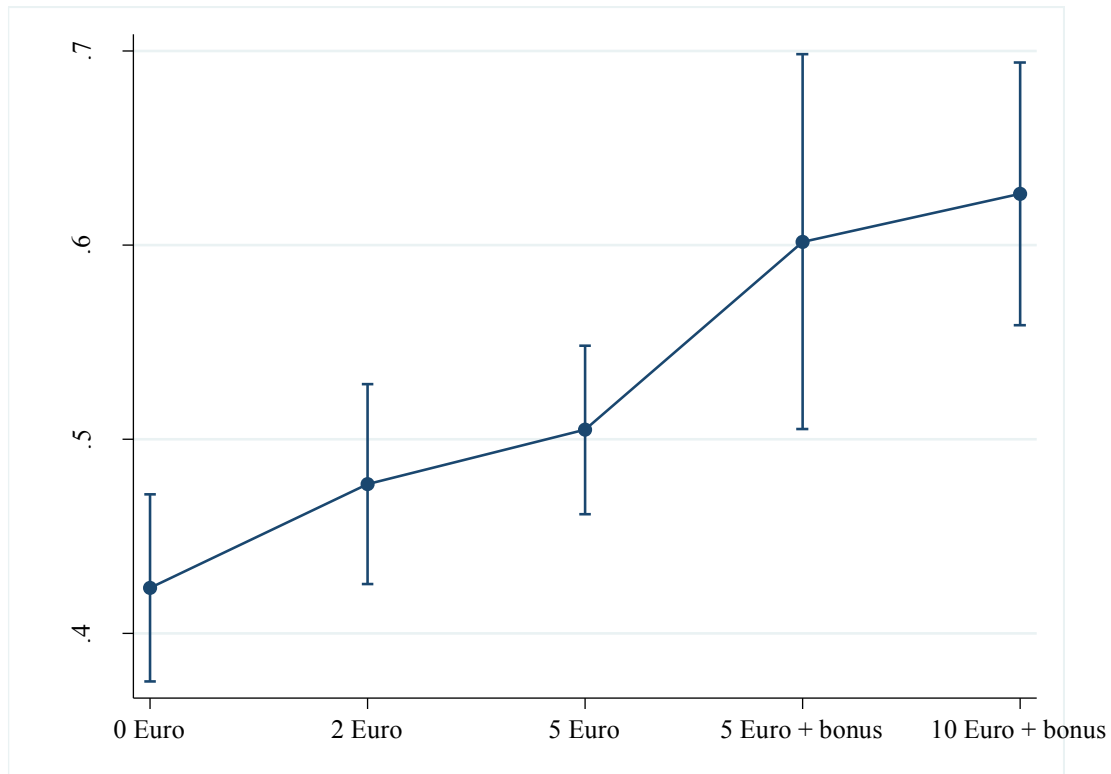


### 6.5.1.3 *Sample composition at the recruitment stage*

To test for selectivity effects that are induced by the experimental treatment, the sample composition across the five experimental groups at the recruitment stage is compared. It was hypothesized that with increasing amount of promised incentive the share of respondents that are typically underrepresented in survey increases (Hypothesis 6.3).

These are respondents with low education, respondents with a lower household income and respondents that are not in paid work, and respondents with a negative survey attitude. In online surveys, the share of women is lower. In addition to the characteristics that are related to the hypotheses formulated above, the mean age of the experimental group and the share of respondents that were interviewed on a cell phone are included as additional measures. All measures with the expectation of age are coded as dummy variables. For these measures, the pairwise differences in the share of the respondents over the experimental groups are calculated as a measure of selection bias. The proportion for each measure across experimental groups is shown in Table 6.7. In the last column, the contrasts that are statistically significant are indicated. In sum, only a minority of 8 out 60 of the contrasts are of statistical significance. The overview of the differences of each contrast can be found in *Table A 5* in the appendix.

*Table 6.7: Sample composition: Share of respondents for selected characteristics across incentive groups at the stage of recruitment*

| | 0 Euros | 2 Euros | 5 Euros | 5 Euros + bonus | 10 Euros + bonus | significant contrasts |
|---|---|---|---|---|---|---|
| | % | % | % | % | % | |
| Low education | 14.26 | 16.92 | 11.68 | 11.62 | 12.73 | * Δ 2 Euros - 5 Euros<br>+ Δ 2 Euros - 5 Euros + bonus |
| In paid work | 68.93 | 71.65 | 71.98 | 66.79 | 71.92 | - |
| Low hh-income | 28.49 | 24.23 | 27.14 | 19.96 | 20.52 | * Δ 0 Euros - 5 Euros + bonus<br>+ Δ 0 Euros - 10 Euros + bonus<br>+ Δ 5 Euros - 5 Euros + bonus |
| Male | 55.66 | 52.21 | 52.89 | 53.33 | 53.54 | - |
| Positive survey attitude | 93.53 | 94.28 | 93.43 | 91.28 | 96.28 | * Δ 5 Euros + bonus - 10 Euros + bonus |
| Mobile sample | 46.23 | 43.65 | 41.60 | 47.18 | 50.00 | + Δ 5 Euros - 10 Euros + bonus |
| Mean age | 40.78 | 41.04 | 40.30 | 42.97 | 42.37 | * Δ 5 Euros - 5 Euros + bonus |

*Note.* Comparison of groups with Chi²-test for the categorical variables and t-test for age (Coefficients based on multiply imputed data with M=30, applying Rubin's rules); n = 1436; n = 662 for low household income
+ p < 0.10, * p < 0.05, ** p < 0.01

Figure 6.4 illustrates the differences for each pairwise comparison of all measures that are related to the formulated hypotheses. The first bar, for instance, represents the difference between the share of respondents that have a low education in the control group and the 2 Euros group (14.26%-16.92%=-2.66). A negative value reflects a higher share in the second mentioned group. A positive value reflects a higher share of respondents in the first mentioned group. The statistically significant differences regarding the share of respondents with low education and the respondents living in a low household income were all positive; indicating a higher share of respondents with low education in the lower incentive groups. In contradiction to the above formulated assumption, none of the differences were statistically significant for the share of respondents that were in paid work. Also, the share of men was not affected by the incentive group. Concerning the survey attitude, the only significant difference is found between the 5 Euros with bonus group and the ten Euros with bonus group, indicating a higher share of respondents with a positive survey attitude in the higher incentive group. With regard to the characteristics under consideration, no systematic effect of the incentive group is revealed at the stage of recruitment. It can be concluded, that the higher incentives did not motivate respondents to participate that are generally underrepresented in surveys, as it was hypothesized.

*Figure 6.4: Pairwise differences for selected measures at the recruitment stage*
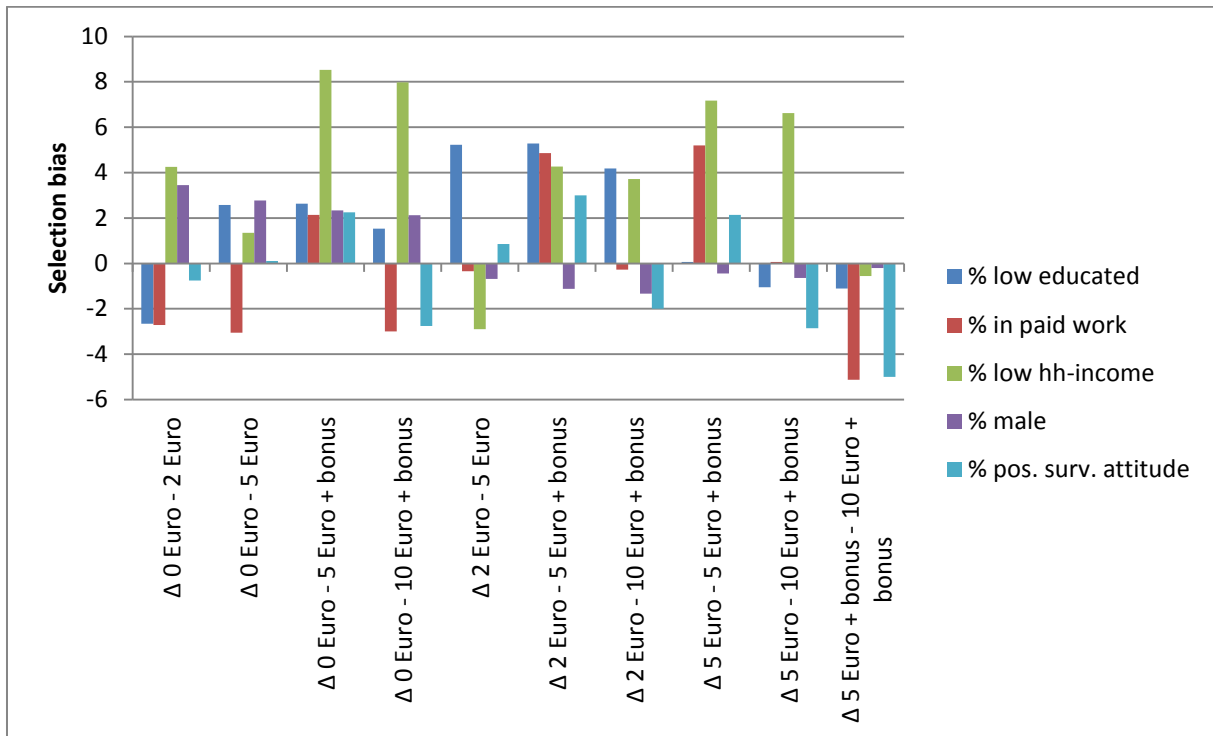


## 6.5.2 Online participation probability

In the following section the subsequent selection step, from giving consent (stage 2) to panel participation (stage 3) is investigated (see Figure 6.5). It is assumed, that the amount of promised incentive is connected with the decision of the respondent to participate, as the amount was also mentioned in each invitation email. The hypotheses formulated in section 6.3 are tested.

*Figure 6.5: Selection step online participation*
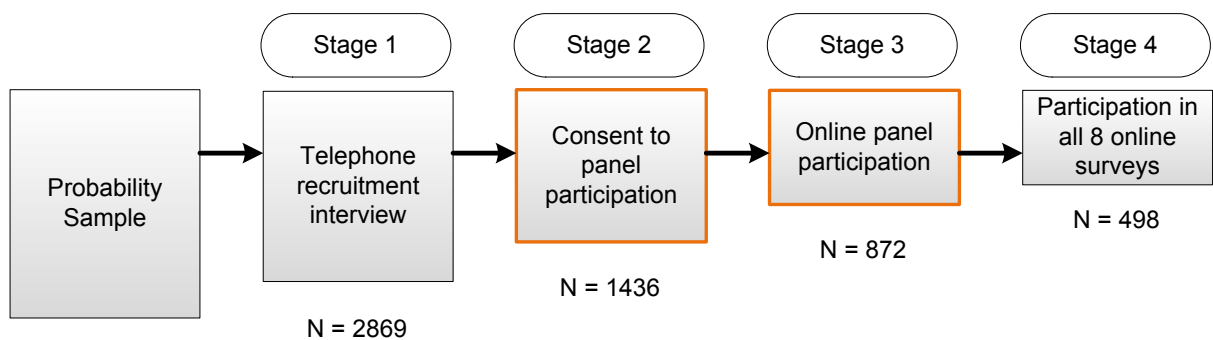


### 6.5.2.1 Descriptive results

Overall 1436 respondents that consented to panel participation were part of the subsequent analyses. Table 6.8 gives an overview of the bivariate relation between incentive group and the outcome variable online participation for all respondents that are included in the subsequent analyses. As a baseline, about 49% of the respondents of the control group that was not promised any incentive

participated in at least one online survey. The share of rcruited respondents that participated in the survey increased with higher incentive amount and went up to a maximum of 76% of the group that was promised 10 Euros per survey plus an additional bonus.

*Table 6.8: Bivariate relation between incentive group and online participation*

|  | % ever participated online (if recruited) | 95% CI | n online participation | n total |
|---|---|---|---|---|
| 0 Euros | 49.37 | [43.87, 54.87] | 157 | 318 |
| 2 Euros | 56.91 | [51.80, 62.01] | 206 | 362 |
| 5 Euros | 58.40 | [53.32, 63.48] | 212 | 363 |
| 5 Euros + bonus | 74.87 | [68.77, 80.98] | 146 | 195 |
| 10 Euros + bonus | 76.26 | [70.32, 82.20] | 151 | 198 |
| Total | 60.72 |  | 872 | 1436 |

*Note.* Pearson chi2(4) =56. 6285 Pr = 0.000

As the incentive is interrelated to the respondents' behavior, it is relevant to investigate the online survey behavior in more detail. After having completed one online survey, the respondents were able to see the credited incentive amount in their respondent's account. To get a more detailed view, a second threefold outcome variable was generated that differentiates whether a respondent did not participate online at all, participated in one to seven surveys, and finally in all eight surveys.

In Figure 6.6 and Table 6.9 the bivariate relation between the incentive group and the three-fold online participation variable is shown. It illustrates the effect of the additional bonus on panel loyalty of the respondents. In both incentive groups that included a bonus for participation in all eight online surveys more than half of the respondents participate in all eight online surveys. In contrast, the share of loyal respondents plummets for the three groups without additional bonus: 34% of the 5 Euros group, 28% of the 2 Euros group, and 17% in the control group without incentive participated in all eight online surveys. It also illustrates remarkable differences between the 2 Euros and the 5 Euros groups. In total, the share of respondents that ever participated in at least one online survey is almost equivalent in both groups. However, the higher amount of incentive has a positive effect on loyalty.

*Figure 6.6: Online participation by incentive groups (n = 1463)*
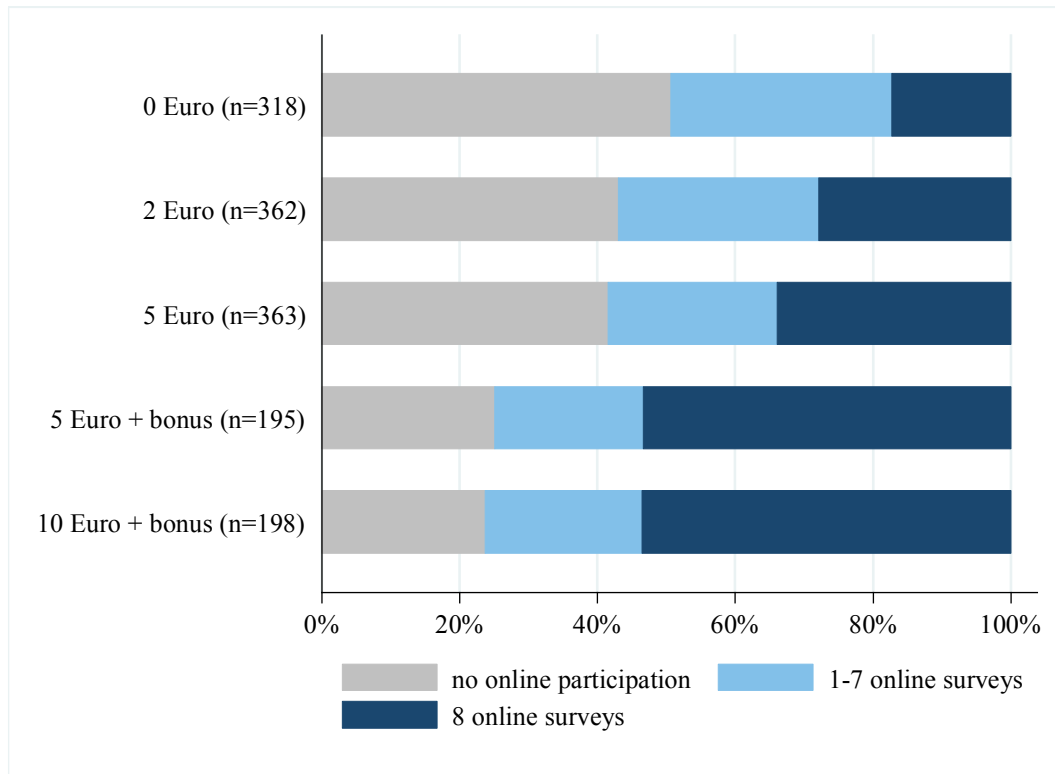


*Table 6.9: Bivariate relation between amount of incentive and online participation (three categories), based on recruited*

|  | No participation | Participation in 1-7 online surveys | Participation in all 8 online surveys | Total |
|---|---|---|---|---|
|  | % (n) | % (n) | % (n) | % (n) |
| 0 Euros | 50.63 (161) | 32.08 (102) | 17.30 (55) | 100.00 (318) |
| 2 Euros | 43.09 (156) | 29.01 (105) | 27.90 (101) | 100.00 (362) |
| 5 Euros | 41.60 (151) | 24.52 (89) | 33.88 (123) | 100.00 (363) |
| 5 Euros + bonus | 25.13 (49) | 21.54 (42) | 53.33 (104) | 100.00 (195) |
| 10 Euros + bonus | 23.74 (47) | 22.73 (45) | 53.54 (106) | 100.00 (198) |
| Total | 39.28 (564) | 26.67 (383) | 34.05 (489) | 100.00 (1436) |

*Note.* Pearson chi² (8) =115.9274 Pr = 0.000

### 6.5.2.2 *Multivariate results*

In Table 6.10 the results of the logistic regression on online participation for those respondents that provided consent is shown. When controlling for factors of respondent characteristics and survey specific factors, the effects of incentive on providing consent resembles those of the bivariate analysis. The odds for online participation increase with increasing amount of incentive. Promising an incentive of 2 Euros increases the odds of being recruited by a factor of 1.36, compared to promising no incentive. However, this effect is not statistically significant. Respondents who were promised an

incentive of 5 or 10 Euros with an additional bonus have more than 2 times higher odds of being recruited in comparison to those who were not promised any incentive.

*Table 6.10: Logistic regression on online participation*

|  | Odds ratio |
|---|---|
| Incentive | |
| *2 Euros* | 1.361+ |
| *5 Euros* | 1.472* |
| *5 Euros + bonus* | 2.816** |
| *10 Euros + bonus* | 2.905** |
| Agency *(Ref. internal)* | 0.869 |
| Male | 0.955 |
| Age | 1.018** |
| Education *(Ref.: low)* | |
| *middle* | 1.152 |
| *high* | 2.178** |
| In paid work | 1.028 |
| Positive survey attitude | 1.495+ |
| Short interview | 0.919 |
| N | 1436 |

*Note.* Coefficients based on M=30, applying Rubin's rules
64 clusters in interviewer; 84 clusters in interviewers
+ p < 0.10, * p < 0.05, ** p < 0.01

The predicted margins are computed separately for each group of incentive based on the logistic regression above, illustrate the predicted probabilities of online participation for the five incentive categories (see Figure 6.7). The results show that the increasing amount of promised incentive that was also announced in the email invitation, positively affected the participation probability. About half of the respondents of the control group that consented to panel participation participated in at least one online survey. The online participation probability increases with increasing incentive amount. However, the confidence intervals are overlapping for the control group, the 2 Euros group, and the 5 Euros group. The probability of online participation rised up to 75% for the two incentive groups that included an additional bonus.

*Figure 6.7: Predicted probability of online participation by incentive group (n=1436)*
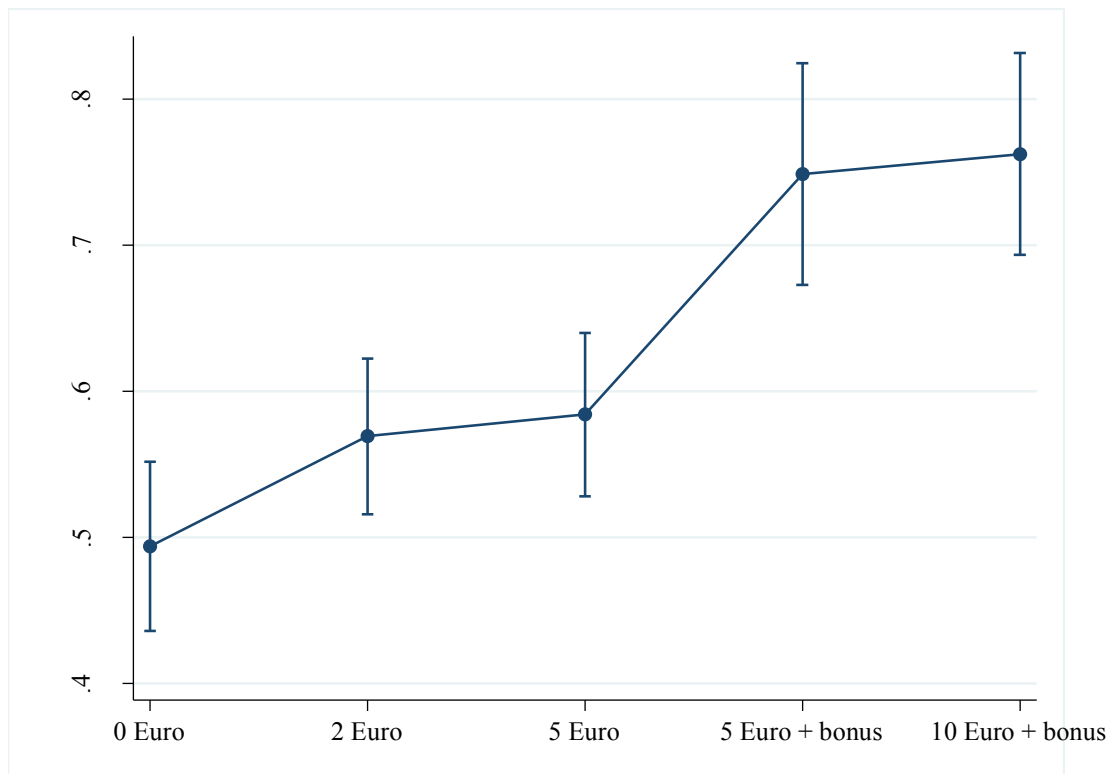


**Participation in all eight online surveys**

The bivariate analysis indicated a positive effect of the additional bonus on the respondents' loyalty. To test for the effect of the bonus on the loyalty, an additional logistic regression was conducted. A dummy variable that indicates whether a recruited respondent participated in all eight surveys versus zero to seven surveys served as independent variable (see Table 6.11). The model indicates large odds ratios for the two bonus groups. The odds of being recruited increased by a factor of 5.059 if respondents were promised an incentive of five Euros including an additional bonus on top versus not promising any incentive, and increased by a factor of 4.951 for the ten Euros plus bonus group.

*Table 6.11: Logistic regression on eight online surveys*

|  | Odds ratio |
|---|---|
| Incentive |  |
| *2 Euros* | 1.886** |
| *5 Euros* | 2.541** |
| *5 Euros + bonus* | 5.059** |
| *10 Euros + bonus* | 4.951** |
| Agency *(Ref. internal)* | 0.848 |
| Male | 0.975 |
| Age | 1.015** |
| Education *(Ref. low)* | 1.000 |
| *middle* | 1.390 |
| *high* | 2.372** |
| In paid work | 0.809 |
| Positive survey attitude | 1.660[+] |
| Short interview | 0.996 |
| N | 1436 |

*Note.* Coefficients based on M=30, applying Rubin's rules
64 clusters in interviewer; 78 clusters in interviewers
+ p < 0.10, * p < 0.05, ** p < 0.01

Figure 6.8 shows the predicted probabilities for participation in all eight online surveys. The baseline probability for the control group is 17.30%. It increased with increasing incentive amount. The increase in the probability of participation in all surveys sums up to 20%-points between the 5 Euros group without and with a bonus. The increase is not only statistically significant but also meaningful. There was no difference between both bonus groups.

*Figure 6.8: Predicted probability of participation in all eight surveys (n=1436)*
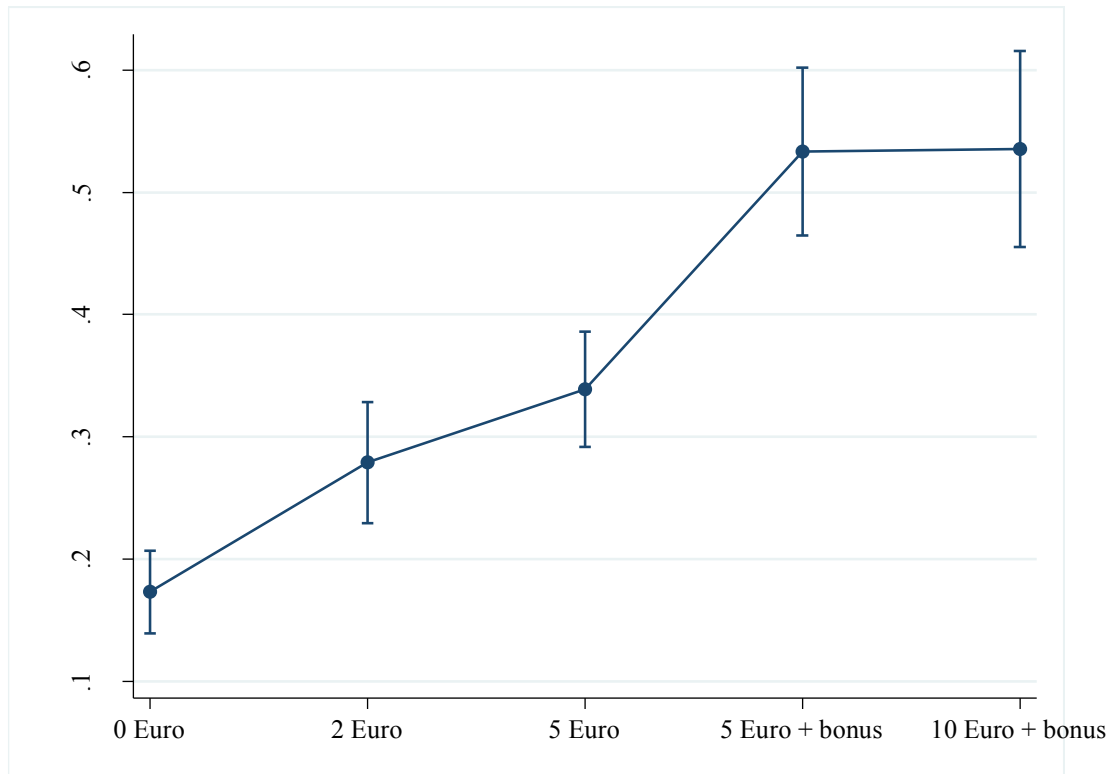


### 6.5.3 Cumulative participation probability

Finally, to evaluate the incentive effect on the resulting sample the cumulative participation probability is calculated. It combines the results of the two previous selection steps and focuses on the effect of the varying incentive amounts on the nonresponse that occurs between the telephone interview (stage 1) and the online survey participation (stage 3). Analogous to the analysis of the stage of recruitment (stage 2) above, the composition of the resulting sample is evaluated.

*Figure 6.9: Selection steps cumulative participation*
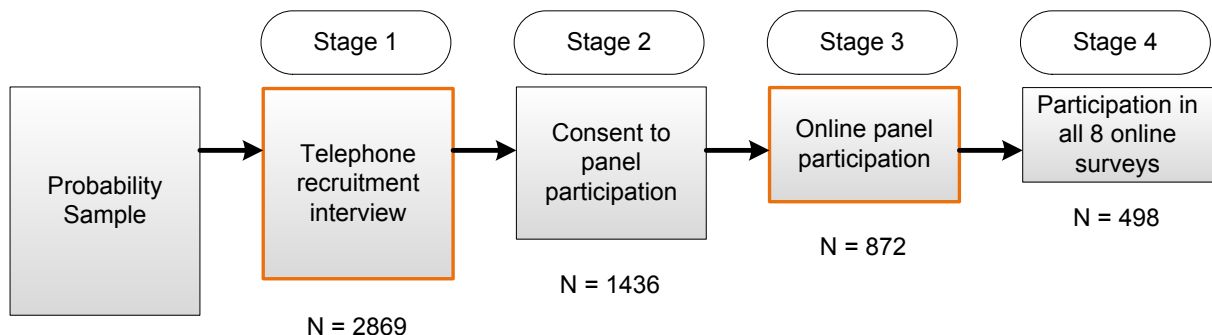


On the basis of the logistic regression model, analogous to the previous analyses, the probability of overall participation for respondents that participated in the telephone recruitment interview is shown in Figure 6.10. The logistic regression can be found in Table A 6 in the appendix. The baseline defined

by the control group is about 20% of the respondents of the recruitment interview finally participated in at least one online survey. The cumulative participation rate increased with increasing incentive amount. For the 2 Euros and the five Euros group the cumulative participation rate is almost identical around 30%. For the two highest incentive groups the increase in the cumulative response rate increases considerably by more than 25%-points compared to the control group. However, the ten Euros group does not perform better compared to the 5 Euros incentive group that included the additional bonus.

*Figure 6.10: Predicted probability of overall participation*
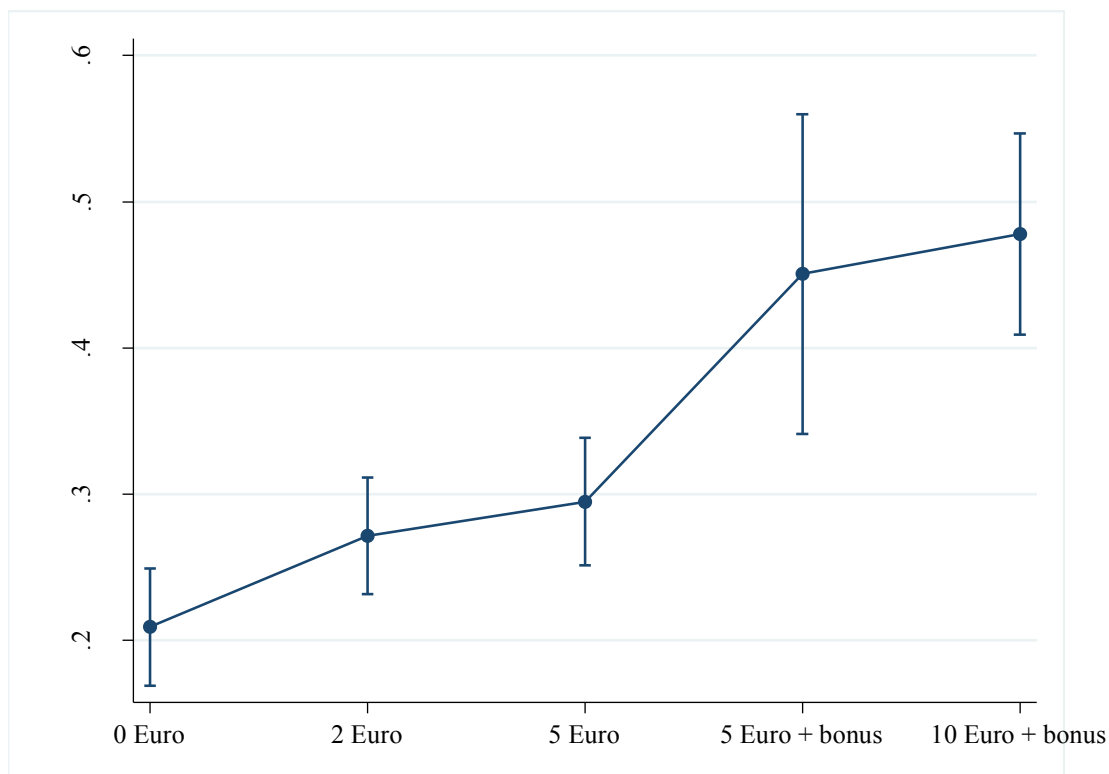


As a test of sensitivity, the probability of overall participation is computed separately for both agencies. Table 6.12 shows the overall participation probability for the social research agency, Table 6.13 the overall participation probability for the market research agency, across incentive groups. The probability of participation is equal for the respondents of the control group for both agencies. The increase in the overall participation is larger for respondents of the social research agency compared to the respondents of the market research agency with increasing amount of incentives. As the confidence intervals are overlapping the aggregated analysis of both agencies is justified. The analysis of potential interviewer effects is part of the following Chapter 7. The predicted probabilities for the recruitment and online participation separated for the agencies can be found in Table A 7 to Table A 10 in the appendix.

*Table 6.12: Predicted probability of overall participation by incentive group*
*for the social research agency (n= 1293)*

|                  | margins | 95% CI           |
|------------------|---------|------------------|
| 0 Euros          | 0.2136  | [0.1529, 0.2743] |
| 2 Euros          | 0.3096  | [0.2249, 0.3944] |
| 5 Euros          | 0.3558  | [0.2832, 0.4284] |
| 5 Euros + bonus  | 0.4506  | [0.3383, 0.5630] |
| 10 Euros + bonus | 0.4778  | [0.4099, 0.5458] |

*Table 6.13: Predicted probability of overall participation by incentive group*
*for the market research agency (n=1569)*

|         | margins | 95% CI           |
|---------|---------|------------------|
| 0 Euros | 0.2089  | [0.1585, 0.2593] |
| 2 Euros | 0.2553  | [0.2118, 0.2988] |
| 5 Euros | 0.2701  | [0.2165, 0.3237] |

### 6.5.3.1    *Sample composition in the resulting online sample*

To test for selectivity effects that are induced by the experimental treatment, the sample composition across the five experimental groups for the final online sample is compared. It was hypothesized that with increasing amount of promised incentive the share of respondents that are typically underrepresented in surveys of the social sciences increases. In addition to the characteristics that are related to the hypothesis formulated above, the mean age of the experimental group and the share of respondents that were interviewed on a cell phone are included as additional measures. The selection bias is calculated for each contrast. Table 6.14 shows the share of respondents for each measure across the incentive groups. In the last column, the significant contrasts are indicated. No differences across the incentive groups were found regarding the share of low educated respondents and respondents that were in paid work. The only significant contrast for the share of respondents that lived in households with low-income, points on the contrary direction as hypothesized. The share of man did as well not differ across the groups. There were some significant differences regarding the share of respondents with a positive survey attitude. The direction of deviation is inconclusive. Also, the sample differences regarding age and the share of respondents interviewed on the cell phone, are minor and not systematically.

*Table 6.14: Sample composition: Share of respondents for selected characteristics across incentive groups at the stage of online participation*

| | 0 Euros | 2 Euros | 5 Euros | 5 Euros + bonus | 10 Euros + bonus | significant contrasts |
|---|---|---|---|---|---|---|
| | % | % | % | % | % | |
| Low education | 9.64 | 12.85 | 10.63 | 11.85 | 11.32 | |
| In paid work | 68.15 | 72.33 | 72.81 | 66.44 | 72.19 | |
| Low hh-income | 26.82 | 20.80 | 20.16 | 18.49 | 21.75 | + Δ 0 Euros - 5 Euros + bonus |
| Male | 54.14 | 50.97 | 52.83 | 54.79 | 52.98 | |
| Positive survey attitude | 94.14 | 97.38 | 91.98 | 90.41 | 98.59 | * Δ 0 Euros - 10 Euros + bonus<br>* Δ 2 Euros - 5 Euros<br>* Δ 2 Euros - 5 Euros + bonus<br>* Δ 5 Euros - 10 Euros + bonus |
| Mobile sample | 42.68 | 39.81 | 37.74 | 45.89 | 50.33 | * Δ 2 Euros - 10 Euros + bonus<br>* Δ 5 Euros - 10 Euros + bonus |
| Mean age | 42.75 | 43.25 | 41.75 | 43.02 | 42.70 | + Δ 2 Euros - 10 Euros + bonus<br>+ Δ 5 Euros - 5 Euros + bonus |

*Note.* Comparison of groups with Chi²-test for the categorical variables and t-test for age; coefficients based on multiply imputed data with M=30, applying Rubin's rules; n = 872; n = 406 for low household income
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Analogous to the recruitment stage, Figure 6.11 illustrates the differences for each pairwise comparison of all measures that are related to the formulated hypotheses. The first bar, for instance, represents the difference between the share of respondents that have a low education in the control group and the 2 Euros group (9.64%-12.85%=-3.21). A negative value reflects a higher share in the second mentioned group. A positive value reflects a higher share of respondents in the first mentioned group. All differences can be found in Table A 11 the appendix.

*Figure 6.11: Pairwise differences for selected measures at the stage of online participation*
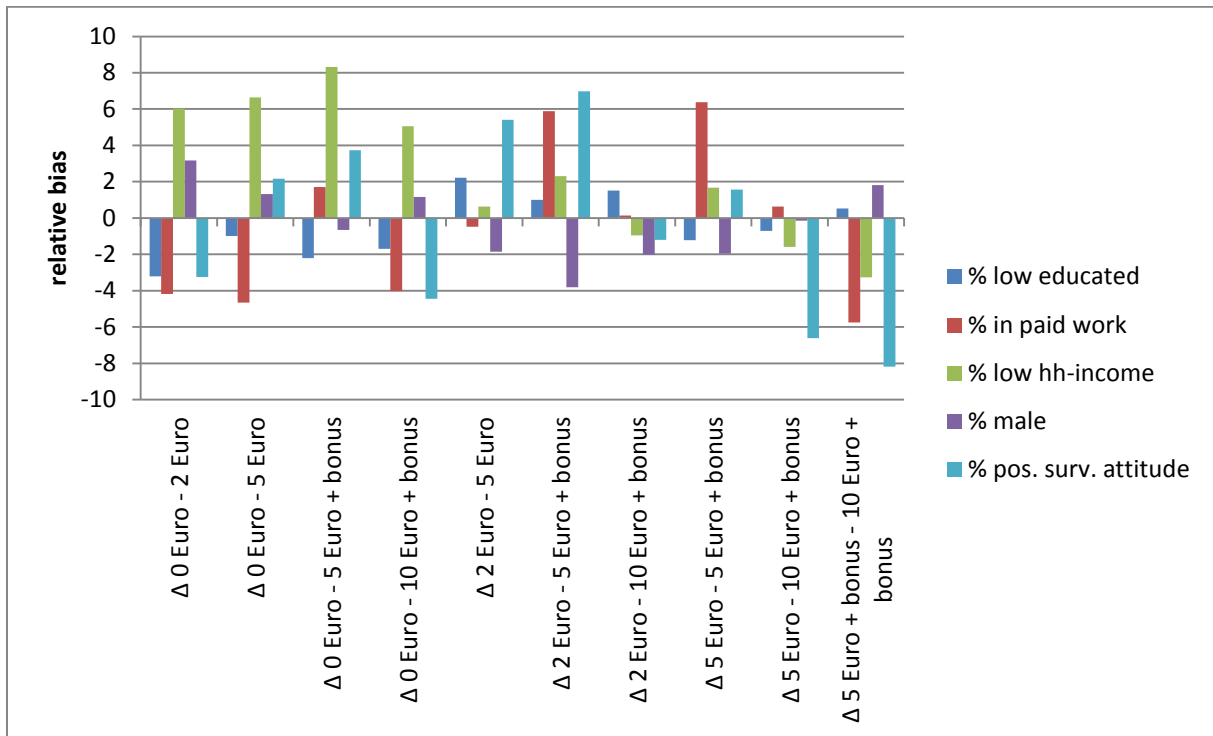


## 6.6 Summary and discussion

In this chapter, I was able to test several hypotheses that have implications for the design of incentive scheme for probability-based online panel recruitment and maintenance. In a longitudinal setting, the decision about a particular amount of incentive has long-lasting consequences. Nonetheless, the evidence of research on the long-term impact of the incentive scheme on respondents' cooperation is scarce (Laurie & Lynn, 2009). This chapter adds to the existing evidence in several ways.

First, it tests the usability of promising incentives in telephone recruitment interviews for a probability-based online panel where non-available postal addresses of the respondents impede the use of prepaid incentives. Second, it examines the effect of promising an incentive for subsequent online survey participation on the recruitment rate, the panel participation rate, and the cumulative response rate. The amount of the promised incentive was experimentally varied between 0 Euros, 2 Euros, 5 Euros, 5 Euros with a bonus of 20 Euros for participation in all eight online surveys, and 10 Euros with a bonus. Third, it examines the effect of the varying incentive amounts on the resulting sample composition, at the stage of recruitment, and the stage of online participation.

First of all, the analysis showed that promising an incentive for subsequent online participation affects the decision of respondents of consenting to join the panel during the telephone interview positively (Research question 6.1). Promising only 2 Euros is not sufficient to increase the recruitment probability compared to the control group. Promising at least 5 Euros for each online survey versus promising no incentive increases the probability of being recruited by about 8%-points. The finding is in contradiction to earlier research that showed promised incentive not to be effective in increasing response rates in cross-sectional surveys where respondents are approached once (Church, 1993;

Mercer et al., 2015). The respondents that have already committed themselves to participating in the telephone recruitment interview seem to trust in the survey organization. In sum, using a promised incentive to motivate respondents at a later stage of the multi-step recruitment process after already having conducted the recruitment interview was shown to be effective. Overall, the positive tendency of increasing response rate with increasing incentive gives some support for Hypothesis 6.1.

In Hypothesis 6.2, I formulated the assumption that the bonus increases the expected value of the incentive and increases the recruitment probability. On the recruitment stage, the strategy to add a bonus for loyal respondents that have finished all eight surveys cannot be unequivocally evaluated as being superior to not adding a bonus. Hence, this hypothesis has to be rejected.

At the stage of online participation, the positive incentive effect of the high incentive groups was even more pronounced (Research question 6.2). Being in one of the two high incentive groups increased the online participation probability by about 25%-points compared to the control group. Hypothesis 6.4 that states the online participation probability to increase with increasing amount of incentive is partly supported.

In Hypothesis 6.5, I formulated the assumption that the bonus increases the share of loyal respondents. The data give support for Hypothesis 6.5. The bonus additionally motivated respondents to stay loyal and participate in all eight online surveys. The positive effect of the bonus across stages results in about 45% cumulative participation rate for the respondents of the bonus groups compared to about 20% cumulative participation rate for the control group. The result of the control group provides a good baseline estimate of the recruitment success that can be expected with a similar design without any incentive.

Research question 6.3 dealt with the sample composition and potential bias that is introduced by the different amounts of incentive. In sum, the results regarding the sample composition are inconclusive. On both stages, no systematic effect of the incentive group on the sample was found. The assumption that respondents that are generally underrepresented in surveys may be motivated by a higher amount of incentive is not supported by the data. Additional variables that were considered here also did not indicate a differential nonresponse for different amounts of incentive. Thus, Hypothesis 6.3 and Hypothesis 6.7 have to be rejected.

The focus of the study presented here was to the effect of promising a prepaid incentive for subsequent online participation. It adds to the existing research on incentives by proving the effectiveness of a promised incentive during the multi-step recruitment process for a probability-based online panel. It provides knowledge about the design of the incentive scheme in a context in which the use of a prepaid incentive is not possible.

A promising path that is already being followed by some of the surveys that use a multi-step approach is to combine several incentive strategies in order to maximize the overall response under given budget constraints.

Due to the pilot nature of the study, the incentive experiment was implemented in a step-wise approach. This fact results in an experimental design that did not fully cross the two factors incentive amount and additional bonus. The bonus condition was implemented only together with the higher incentive groups and was only icorporated in interviews of the social research agency. This design decision resulted in several limitations. First, due to the fractional design, the effect of the fieldwork agency and the bonus conditions could not be disentangled completely. And second, there is no evidence on the effect of the bonus in combination with the lower incentive groups. It could not be tested whether the combination of a lower incentive and the high bonus would yield comparing results. The use of a bonus is an aspect that is worth following up. In the study, the bonus amount was not experimentally varied. The chosen amount of 20 euros represents a considerable share of potential incentive gain. Hence it is worth considering if a lower bonus is also sufficient to motivate respondents to stay loyal.

In addition, a more flexible use of a bonus is also worth considering. Probability-based online panels include cross-sectional as well as longitudinal studies. For longitudinal studies, minimizing unit nonresponse and achieving a measurement for all respondents in all waves is a major concern. Hence, promising respondents a bonus for participation in all waves that include a particular longitudinal study is a strategy that could be tested in a future study.

# 7. Study 4: Interviewer effects on panel recruitment

## 7.1 Introduction and overview

In addition to factors of the survey design, interviewers are assumed to contribute to the success of recruitment. They play a central role in the process of data collection in telephone surveys. They represent a crucial element in order to secure the successful administration of the survey and have to master several tasks (Japec, 2008). As interviewers differ in their ability to manage all these tasks, in the context of the Total Survey Error framework (Biemer, 2010b) they are also considered to be a potential source of non-sampling error (Groves et al., 1992; Groves & Couper, 1998; Groves, 1989). In their research synthesis on interviewer effects, West & Blom (2016) classify the interviewers' task and the potential corresponding survey errors: Interviewers might contribute to coverage error, nonresponse error, measurement error, as well as processing error. In the multistep recruitment process for a probability-based online panel, interviewers have to obtain additional consent to subsequent panel participation at the end of the interview and thus represent a potential source of nonresponse error.

In the previous chapters, the effect of the experimentally varied survey features, incentive amount, and length of recruitment interview on the recruitment probability were analyzed. The effects were assumed to be uniform across interviewers. However, little is known about how different interviewers cope with different experimentally varied features of the survey. The question is, for instance, whether the increase of the recruitment probability for respondents that were promised five Euros incentive compared to respondents that were promised no incentive is the same for all interviewers. To the best of my knowledge, there is no empirical study that tests whether effects of experimentally varied survey features on recruitment are uniform across interviewers or not. Although interviewers are mentioned as a possible error source while implementing survey experiments by Tourangeau (2004), existing studies mainly focus on how incentives affect the interviewers' motivation or expectation, not on inter-interviewer variance (Lynn, 2001; Singer et al., 2000).

The objective of this chapter is twofold. It first aims to quantify the magnitude of the interviewer variance on the panel recruitment probability and explain parts of the variance, and, second, identify possible interviewer effects on the recruitment probability for the experimental variations of the incentive amount and length of the recruitment interview.

In the following section, I summarize the pertinent past knowledge on interviewer effects on nonresponse and cooperation. Whenever possible, evidence for telephone surveys is in focus. This part is followed by the outline of the research questions and hypotheses, and the description of the data used. The analysis aims to quantify and explain interviewer variance on recruitment and investigate the role of the interviewers in conduction standardized survey experiments. In the last section, I discuss the results and sketch possible implications.

## 7.2 Interviewer effects on nonresponse and cooperation

While early studies on interviewer effects mainly tried to quantify the magnitude of interviewer variance on the outcome variables (Kish, 1962), a growing number of studies try to explain the interviewer variance with several sets of predictor variables (Schaeffer et al., 2010; West & Blom, 2016). There is little consistency on the significance of the predictors, which are, for example, interviewer demographics, interviewer-respondent-matching by demographics, interviewers' attitudes and personality, voice characteristics, and interviewers' experience across studies (Schaeffer et al., 2010; West & Blom, 2016). In the following section particularities of the analysis of interviewer effects among telephone surveys are presented. This section is followed by the summary of previous research on interviewer effects on nonresponse that is structured by the following characteristics: demographic characteristics, attitudes, expectations, and personality, and work experience.

### 7.2.1 Interviewer effects in centralized telephone interviews

The challenges for the analysis of interviewer effects differ between interviews that are conducted face-to-face by an interviewer in the respondents' home and interviews conducted in centralized telephone studios. The interviewer variance tends to be larger in face-to-face surveys than in telephone surveys (Groves & Magilavy, 1986; Groves, 1989). Factors that are assumed to affect the interviewer-respondent interaction differ between the two modes: the respondent knows less about the interviewer in the phone mode. Generally, the sex and the voice of the interviewer are the two characteristics that are transferred via phone (Schaeffer et al., 2010). In contrast, the physical appearance of the interviewer is more relevant in the face-to-face mode (Meier Jæger, 2016).

In addition to the visibility of interviewer characteristics, the methodological challenges for the analysis differ as well between the modes. A prerequisite for the isolation of interviewer effects is the random assignment of interviewers to sampled units (interpenetrated design) (Schnell & Kreuter, 2005). In personal interviews, one interviewer is often assigned to one area. Thus the disentanglement of interviewer and area effects (Campanelli & O'Muircheartaigh, 1999; Schnell & Kreuter, 2005) needs to be assured afterward in the analysis (e.g., Pickery, Loosveldt, & Carton, 2001). For telephone interviews that are conducted by a centralized studio, a quasi-interpenetrated design is assured (West & Blom, 2016). Usually, cases are assigned to the interviewers that are available at random and not exclusively to one interviewer prior to the survey interview. Several interviewers work on the same case until finally a contact or interview is obtained. Thus, the outcome of a contact is dependent on earlier contacts which were not made by the same interviewer (Lipps, 2008). In consequence, the calculation of interviewer-specific success indicators can be problematic. To solve this CATI-specific problem several ways of calculating success indicators were proposed for example by Lipps (2008) and West & Groves (2013).

## 7.2.2 Predictors of interviewer variance

### 7.2.2.1 *Demographic characteristics*

In many studies, sex and age of the interviewer are the only interviewers' attributes that were available from the agencies administrative records. Regarding telephone surveys, sex is the only observable attribute for the respondents, besides the interviewer's voice (Groves, O'Hare, Gould-Smith, Benki, & Maher, 2008). The question how respondents perceive the interviewer's age on the phone is an open one (Schaeffer et al., 2010).

The overall evidence suggests that there is only a weak direct connection between sociodemographic characteristics of the interviewers and cooperation (Schaeffer et al., 2010; West & Blom, 2016). Groves et al. (2008) found a small positive effect for female interviewers in personal interviews, whereas, Campanelli & O'Muircheartaigh (1999) reported higher nonresponse rates for male interviewers. For telephone surveys, Lipps (2010a) reported a positive effect for older male interviewer for both, male and female respondents. This effect remained significant after controlling for experience. Hansen (2007) also found male interviewers being more successful than female interviewers in telephone surveys. In contradiction to these findings, telephone interviewers with voices rated less masculine obtained higher response rates (Groves et al., 2008). Several voice characteristics and characteristics of the speech which are related to successful interviewers were identified (Broome, 2015; Groves et al., 2008; Oksenberg & Cannell, 2001). For instance, interviewers who spoke louder, with a higher perceived pitch and intonation (Oksenberg & Cannell, 2001), as well as female interviewers who spoke with a medium pace and a medium pitch (Steinkopf, Bauer, & Best, 2010) were more successful.

Some studies found a positive effect for age (Blom et al., 2011; Schaeffer et al., 2010), which might be confounded with experience. West & Olson (2010) found in their analysis of CATI interviewer data, different interviewers recruiting different respondents (West & Olson, 2010).

The framework of survey participation by Groves & Couper (1998) that also represents the heuristic framework of this dissertation relies on the interaction between respondent and interviewer. The combination of interviewer and respondent characteristics, as a proxy of "liking", might be more promising in explaining interviewer variance. Groves et al. (1992) assume that respondents tend to have a more positive attitude towards interviewers that are similar to them and share common attitudes. For personal interviews, Durrant et al. (2010) report a small and marginal interaction effect of sex and education on cooperation, while Lipps (2010b) does not find an interaction between interviewers' and respondents' sex in a telephone survey on cooperation.

In sum, the sociodemographic measures only explained a small share of interviewer variance (Schaeffer et al., 2010; West & Blom, 2016). That is why more recent research additionally included indicators of interviewers' attitudes towards surveys and their job (Blom & Korbmacher, 2013; de Leeuw, 2009), as well as measures of personality (Durrant et al., 2010).

### 7.2.2.2  *Attitudes, expectations, and personality*

Several studies of personal interviews showed that interviewers with positive expectations and those who were confident of doing a good job tended to reach higher response rates (de Leeuw, 1999; Groves & Couper, 1998; Singer, Frankel, & Glassman, 1983). Furthermore, de Leeuw (1999) showed that interviewers who valued professional behavior were also more successful. In line with this West & Blom (2016) found interviewers who attached importance to a professional image had higher contact rates. Interviewers who were more confident in being successful were actually more successful (Blom et al., 2011; Durrant et al., 2010). Link (2006) showed that among telephone interviewers those who were confident in combination with a realistic view on their abilities to convince respondents performed best.

In more recent studies, differences in interviewers' personalities were included as explaining factors. Overall, the results were inconclusive. Jäckle, Lynn, Sinibaldi, & Tipping (2013) showed an association of two out of five of the big-five personality traits with interviewer variance: interviewers that rated themselves as being extroverted had higher cooperation rates, whereas, interviewers who were more open obtained lower cooperation rates. After controlling for experience and additional attitudes, the effects diminished. Two studies found that agreeableness was associated with lower cooperation rates (Jäckle et al., 2013; Snijkers et al., 1999). Vassallo, Durrant, & Smith (2015) stressed that the value of interviewers personalities as explaining factor is inconclusive. They conclude that even if an association between interviewers' personality and respondents' behavior could be found, it would be hard to make use of this knowledge to improve interviewer training. A factor that is easier to assess is the work experience of the interviewer.

### 7.2.2.3  *Work experience*

Several studies showed a positive effect of interviewers' work experience on respondents' cooperation (Groves & Couper, 1998; Hansen, 2007; Hox & de Leeuw, 2002; Jäckle et al., 2013; Lipps & Pollien, 2011; Pickery & Loosveldt, 2002), while others did not find any effect or even a negative effect for very experienced interviewers (Durrant et al., 2010). In their research synthesis, West & Blom (2016) emphasize the multitude of measures used. Nevertheless, they showed that the majority of studies found a positive correlation between experience and cooperation. A general distinction can be made between general work experience as an interviewer, often measured in years, and survey-specific work experience, often measured in the number of interviews made before. Additionally, some studies took into account not only the number of previous interviews, but also previous success in the task (e.g., obtaining consent) (Sakshaug et al., 2012; Sala, Burton, & Knies, 2012). The survey-specific experience was shown to be more important in explaining unit nonresponse than general job experience (Lipps & Pollien, 2011). Other studies found no effect for the general measure years of work experience, but a negative effect of survey-specific experience and a positive effect of previous success in obtaining consent (Sala et al., 2012), and, for example, Sakshaug et al. (2012) found a marginally statistical positive effect of prior success. For the Swiss Household Panel that is delivered

via phone, Lipps (2009) showed that the negative prior experience (i.e. refusal) negatively affected the subsequent performance, while a positive prior experience had no effect. In sum, it remains unclear, whether the positive correlation of general work experience is a selection effect in the sense that less successful interviewers quit the job after a while or a direct effect of experience per se (Blom et al., 2011; Davis, Couper, Janz, Caldwell, & Resnicow, 2010). Although the findings may be confounded with interviewers' self-selection to remain as interviewers, one inference that can be drawn is that experienced interviewers' success derives from their "larger number of combinations of behaviors proven to be effective for one or more types of householders" (Groves et al., 1992: 478–479). Expert interviewers rely on two techniques that are assumed to be the key for successful interviewers: tailoring and maintaining interaction (Groves & McGonagle, 2001).

The summary of research on interviewer effects shows that the majority of most recent research is restricted to interviews that were conducted face-to-face. Less is known about interviewer effects and the determinants of interviewer variance in telephone surveys. In line with this reasoning Biemer (2010b, p. 844) stated *"Although numerous studies of nonresponse bias have been reported, relatively less is known about other sources of nonsampling error. For example, interviewer variance is rarely estimated in centralized telephone surveys, even though the cost of doing so routinely is relatively small."* The scope of the analysis of this chapter goes beyond reporting the interviewer variance. I try to explain parts of the interviewer variance by respondent and interviewer characteristics, their interaction, as well as characteristics of the survey. Previous research showed interviewer variances on various tasks. However, to the best of my knowledge, no study has hitherto investigated the effect of interviewers as mediators for survey features effects. In the present chapter, I additionally test, whether the effect of the experimentally varied features – the amount of promised incentive and the length of the recruitment interview – on the recruitment success differs by interviewers. And if so, by what extent the differences can be explained by means of interviewer experience.

## 7.3 Research questions

The basic aim of this chapter's analysis is to quantify interviewer variance on the propensity of being recruited, which is formulated in the first research question.

Research question 7.1:       Does the propensity of being recruited differ across interviewers?

The multilevel model is outlined in Figure 7.1 below. The data are structured in three levels: Respondents are nested in interviewers, and interviewers are nested in two fieldwork agencies. The predictor variables that are assumed to have an effect on the recruitment probability are shown for each level. In addition, the assumed cross-level interactions are included. The following research questions are structured accordingly.

Under the assumption that interviewers' variance is found in the propensity of recruitment the factors that explain the variance will be further investigated. Accordingly, Research question 7.2 to Research question 7.5 are formulated.

First, it will be investigated whether parts of the interviewer variance can be explained by the fact, that different interviewers are interviewing different respondents. Due to data limitations, the only respondent information that is available for all respondents is the sex.

Research question 7.2:    To what extent can the interviewer variance in the recruitment propensity be explained by characteristics of the respondents?

The second set of predictor variables is on the interviewer level. It will be investigated whether the interviewer demographics sex and education, as well as the work experience of the interviewers, can explain parts of the interviewer variance. Additionally, the interviewers own hypothetical willingness to participate in an online panel as a proxy of commitment is included.

Research question 7.3:    To what extent can the interviewer variance in the recruitment propensity be explained by characteristics of the interviewers?

On the basis of the above-summarized literature, the focus will be on interviewer expectations and interviewer experience. I assume interviewers that rate their own hypothetical willingness to participate in an online panel as high, to be more successful compared to interviewers with a lower own hypothetical willingness. In addition, I assume experienced interviewers to be more successful compared to interviewers that are less experienced.

Finally, the interaction between interviewer and respondent characteristics will be investigated. The characteristics at hand are the respondent's and interviewer's sex. Following the idea of "liking" (Groves et al., 1992) the same sex combination should result in higher response rates compared to the opposite sex combination.

Research question 7.4:    To what extent can the interviewer variance in the recruitment propensity be explained by a match of the sex of interviewer and respondent?

The third level is the level of the fieldwork agency. In the analysis of the previous chapters, some indications of differences in the performance of the two fieldwork agencies that conducted the study were found. For instance, the initial response rates for the telephone recruitment interview differed largely between the two agencies. The two agencies differed with respect to their major field of projects: The first one had a strong focus on academic projects within the social sciences, and the second one mainly focused on market research projects. In the current chapter, the role of the interviewer staff will be investigated.

Research question 7.5:    Does the performance of the interviewers of the two fieldwork agencies differ and can potential differences be explained by characteristics of the interviewers?

In the last part of the analysis, the interplay of interviewers and the effect of experimentally varied features of the survey will be investigated. The features are the amount of promised incentive and interview length. Additionally, the possibility of recruiting respondents without a prior interview, a measure that was given to the interviewers in order to give them some kind of flexibility, will be included in the analysis as well. Research question 7.6 is formulated accordingly.

Research question 7.6:    Does the propensity of being recruited vary across experimental groups, depending on interviewers?

Regarding the interviewer effects on recruitment for the experimentally varied survey features, I expect an interaction with the interviewers' experience. With respect to the experimentally varied incentive amount, I assume the incentive group differences on recruitment to decrease for experienced interviewers compared to less experienced interviewers. Accordingly, I assume the differences in recruitment between short and regular interviews to decrease with increasing experience of the interviewers. In contrast, I assume the differences in recruitment between respondents who were interviewed and those who were not interviewed will be larger for experienced compared to less experienced interviewers. This assumption is derived from the finding of Groves et al. (1992) that experienced interviewers are assumed to be able to react flexibly to the respondent and tailor their behavior to the needs of the situation (Groves & McGonagle, 2001).

*Figure 7.1: Multilevel model on recruitment*
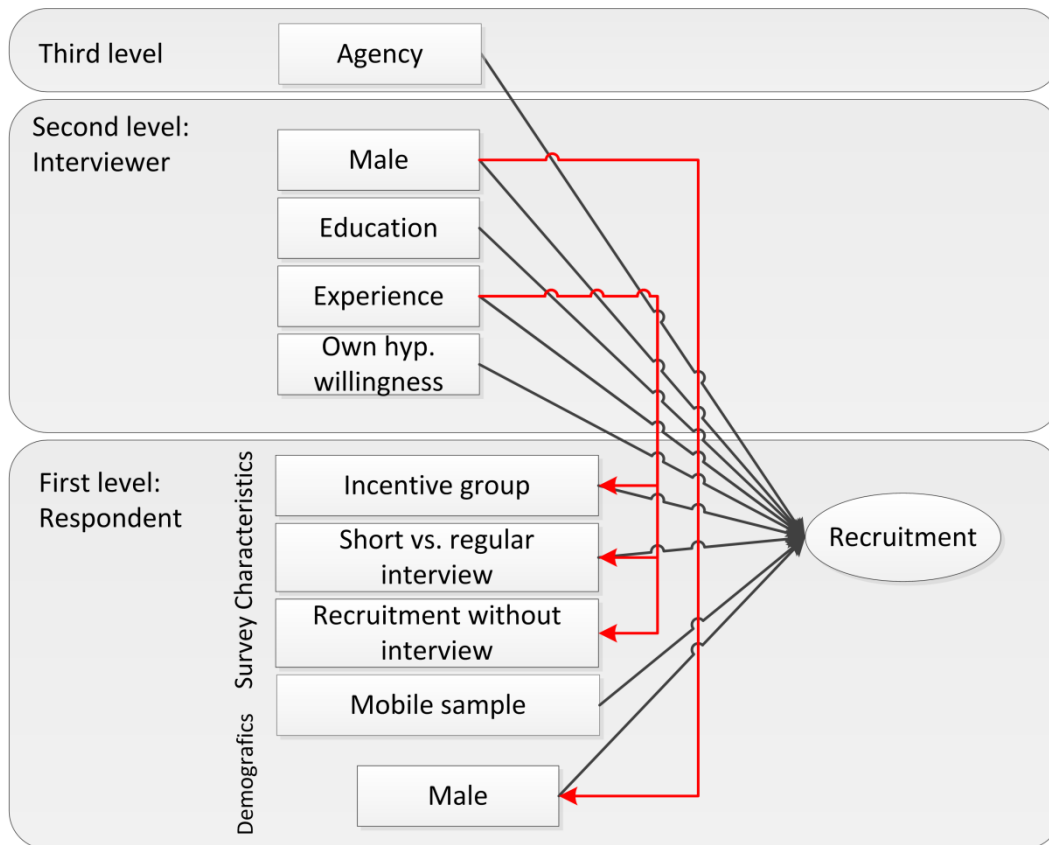


## 7.4 Method

### 7.4.1 Field Work

The interviews were conducted by two data collection agencies in centralized CATI studios. The first one was a fieldwork agency, which was specialized in high-quality studies in the field of social sciences and academia. The second one was a fieldwork agency that operated mainly in the field of market research studies. As the landscape of the fielding agencies in Germany is divided into agencies that are based in academia and agencies that mainly focus on market research, the two agencies can be seen as prototypes of typical agencies in Germany. There was no overlap of the interviewer staff of the two agencies. In both agencies, all interviewers received a survey-specific training (Stiegler & Biedinger, 2016) by the agency supervisor that lasted about two hours. In both cases, the interviewers were paid per hour and did not receive any bonus for finished interviews or recruited respondents.

### 7.4.2 Data

Overall 88 interviewers from two agencies worked for the project and completed 3422 interviews (academic, social research agency: 1487; commercial market research agency: 1935). Three interviewers did not fill out the interviewers' questionnaire. For this reason, respondents that were interviewed by the respective interviewer had to be deleted from the subsequent analysis. In the

analysis, data from 3320 interviews that were clustered within 85 interviewers will be used. Overall, 22 interviewers worked for the social research agency, and 63 interviewers worked for the market research agency.

The number of interviews per interviewer varied widely, from 2 to 176. The median interviewer of the academic agency conducted 52.5 interviews, the median interviewer of the market research agency conducted 20 interviews.

The interviewers were asked to fill out a short questionnaire that encompassed five questions before they started their work on the project. They received the paper questionnaire from their supervisors.

Table 7.1 gives a brief overview of the characteristics of the interviewers in the two agencies. It shows that the majority of the interviewers were female in both agencies. The interviewers' age shows a wide range between 19 years and 74 years of age. The mean age of the interviewers of the academic agency was almost 43 years which was more than ten years higher than the mean age of the interviewers of the market research agency (30 years). Overall more than half of the interviewers were highly educated, which means they had at least an A-level degree (German Abitur). However, there were substantive differences concerning the educational attainment between the two agencies. The majority of the interviewers of the academic agency (77%) were highly educated, whereas about half of the interviewers of the market research agency were highly educated.

The earlier work experience represents a further attribute that differs between the two agencies. On average the interviewers of the academic agency have had about two years more work experience as telephone interviewer (~5 years) compared to their colleagues of the market research agency (~3 years). Both interviewer pools included very experienced interviewers with more than ten years of specific work experience, as well as interviewers that were new on the job. However, among the interviewer pool of the market research agency about one-third had less than one year of relevant work experience, whereas only one interviewer of the academic agency was new to the job.

Finally, the interviewers were asked to rate their own hypothetical willingness of giving consent to online panel participation as a proxy of their commitment. The scale goes from 1 "very unlikely" to 10 "very likely." Those interviewers who said they would not even participate in the phone interview were coded as zero. Overall, they showed a high commitment. The interviewers of the academic agency showed a higher commitment (average 8.64) compared to their colleagues (average 7.13), with a much lower range (4 to 11 compared to 0 to 11). In sum, the small interviewer pool of the academic agency, which was specialized in social sciences research, was much more homogeneous related to the attributes that were collected in the short interviewer survey, compared to the interviewer pool of the market research agency.

*Table 7.1: Interviewer characteristics*

|  | Academic social research agency (n=22) | Commercial market research agency (n=63) |
|---|---|---|
| % female | 68.18% | 63.49% |
| Age in years | $\overline{x}$ 42.63 | $\overline{x}$ 30.32 |
|  | range: 25 - 61 | range: 19 - 74 |
| Education |  |  |
| *Low education* | 4.55% | 30.16% |
| *Middle education* | 18.18% | 9.52% |
| *High education* | 77.27% | 50.79% |
| *Still student* | - | 6.35% |
| *Other* | - | 3.17% |
| Experience in years [0-12] | $\overline{x}$ 4.95 | $\overline{x}$ 2.99 |
|  | range: 0 - 12 | range: 0 - 11 |
| Own hypothetical willingness to participate [0-11] | $\overline{x}$ 8.64 | $\overline{x}$ 7.13 |
|  | range: 4 - 11 | range: 0 - 11 |
|  | 27% in category 10 or 11 | 17% in category 10 or 11 |
|  |  | 13% in the lowest categories under 4 |

In the subsequent analyses, two additional measures of interviewer experience – study-specific experience and previous success – that were predictive in earlier studies (Sakshaug et al., 2012; Sala et al., 2012) will be included. Both are derived from the survey data: The study-specific experience is operationalized as the number of interviews one interviewer conducted prior to the interview of interest. The previous success within the study is operationalized as the number of previous successfully recruited respondents. The above-mentioned problem of calculation of success indicators for telephone interviewers because more than one interviewer is working on one case (Lipps, 2008) does not apply in this case because the focus is on recruitment for subsequent surveys at the end of a successfully conducted interview by one interviewer.

How and whether these differences affect interviewer performance will be the object of further examinations in this chapter.

## 7.5 Results

The subsequent analyses section is organized into two parts. First, a descriptive analysis of the interviewer performance is presented, and, second, logistic multilevel models on recruitment that include both interviewer and respondent-level information is presented. The multivariate analyses first include all interviewers of both agencies in order to identify possible interviewer effects, the determinants of the interviewer variance, and the effect of the agency (Research question 7.1 to Research question 7.5). In order to answer Research question 7.6, whether the propensity of being recruited for respondents of the different experimental groups differs across interviewers, the subset of interviewers and respondents of the market research agency is part of the analysis.

### 7.5.1 Interviewer performance

Overall, 54.48% of the respondents who finished the recruitment interview and were asked the panel question were willing to provide their email address to participate in further online surveys. The recruitment rates for the two fieldwork agencies differ significantly (academic agency: 57.20%; market research agency: 52.43%). However, as shown in Chapter 6 above, the high incentive groups are associated with higher recruitment rates. The high incentive conditions were exclusively used by the social research agency. The difference between the agencies diminishes when the interviews of the high incentive groups are excluded.

Next, the performance of the individual interviewers is investigated, separately for the two telephone agencies. The performance rate is calculated for each individual interviewer. Performance is measured in terms of the recruitment success rates. The recruitment success rate is calculated as the ratio of recruited respondents in comparison to conducted interviews. Figure 7.2 shows the distribution of the recruitment success rate for both fieldwork agencies. The figure shows greater variation for the market research agency. Some interviewers performed far above average, while others performed far under average. The interviewer of the social research agency, who marks the 50%-percentile, has a recruitment rate of almost 0.61. The interviewer of the same quantile for the market research agency has a recruitment rate of 0.46. The most successful interviewer of the academic agency has a recruitment rate of 0.73, the most successful interviewer of the market research agency has a recruitment rate of 1. However, this recruitment rate is reached by an interviewer who conducted only two interviews. The same applies for the least successful interviewer of the market research agency. The association between the number of conducted interviews and performance is investigated in the next step. In conclusion, the interviewers of the academic agency are more homogeneous in their performance.

*Figure 7.2: Performance indicator recruitment success rate by agency*
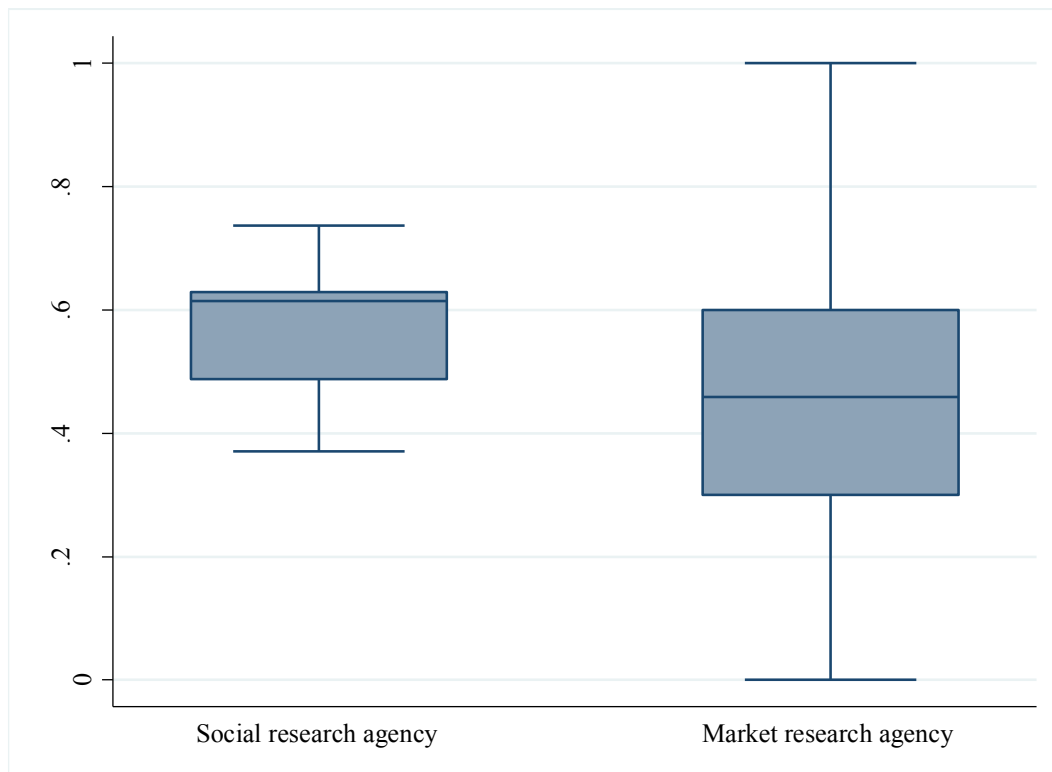


Figure 7.3 illustrates the relation between the performance indicator recruitment success rate and the number of realized interviews separated by fieldwork agency. In both cases, interviewers with the highest number of interviews range in the middle of the distribution of the recruitment success indicator. A potential explanation for this fact could be that highly productive interviewers were assigned to more difficult cases. There is no correlation between the number of interviews and recruitment success for the social research agency. For the market research agency, the figure shows, that interviewers who performed below average conducted fewer interviews. However, among the interviewers who performed better, interviewers with many interviews as well as interviewers with few interviews can be found.

*Figure 7.3: Number of interviews and performance indicator recruitment success rate by agency*



Graphs by Fieldwork agency
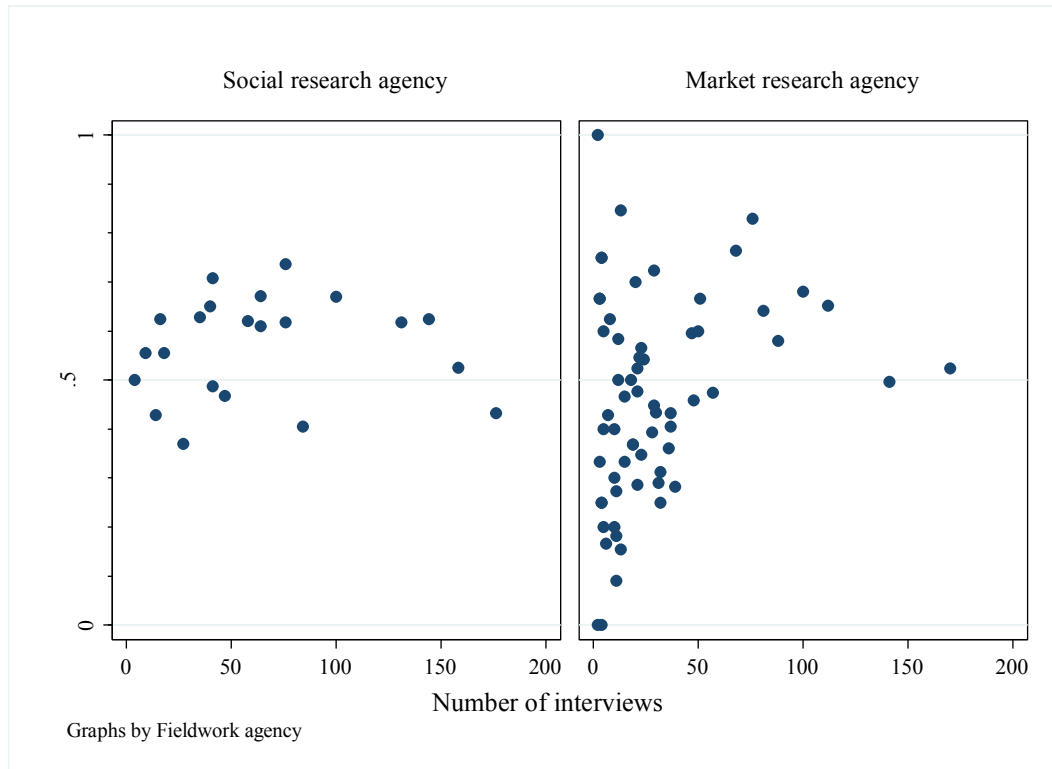
## 7.5.2 Multilevel analyses

In order to take into account the dichotomous nature of the dependent variable recruitment, as well as the hierarchical structure of the data, where respondents are nested in interviewers, which in turn are nested into fieldwork agencies, a multilevel logistic regression model is applied for the multivariate analyses. The above-described differences in the interviewer composition and performance factors, as well as the results from the previous chapters, indicate that differences between the two fieldwork agencies have to be taken into account. Due to the limited number of two fieldwork agencies, the fieldwork agency cannot be included as a third level into the model, but it is included as a dummy variable on the interviewer level. The two-level model is shown in Figure 7.4.

*Figure 7.4: Analysis model*
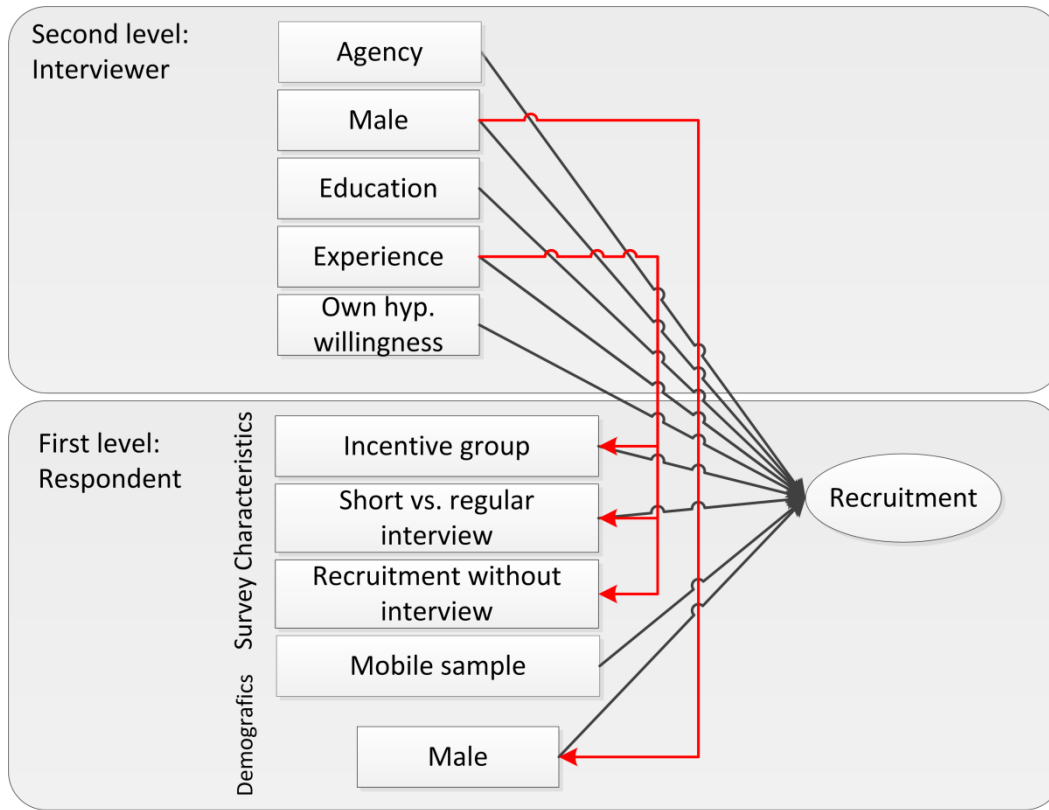


### 7.5.2.1    *Interviewer variance and its determinants*

***Null model***

To answer Research question 7.1, whether the recruitment propensity differs across interviewers, first of all, a null model without any predictor variables is estimated to quantify the amount of variance that is located on the interviewer level. The estimation procedure of all of the following multi-level models is based on a maximum likelihood estimation using adaptive quadrature with seven integration points. In order to get deeper insights into the structure of the data, three null models are estimated, one for all interviewers of both agencies and one for each of the two agencies[8]. The null model for being recruited can be formalized as follows:

$$(1) \qquad \ln\left(\frac{p}{1-p}\right)_{ij} = \gamma_{00} + u_{0j}$$

$$\text{with } u_{0j} \sim N(0, \sigma_{u_0}^2)$$

The intercept $\gamma_{00}$ is shared by all interviewers while the random effect $u_{0j}$ is specific to interviewer j. Table 7.2 shows the null models of the two-level logistic regression model on recruitment.

In the pooled data about eight percent of the total variance of being recruited is located on the interviewer level (Intraclass correlation (ICC = 0.274/0. 274+ $\pi^2$/3 = 0.077). In a study of centralized telephone interviews with a comparable setting, Vehre (2011) found a slightly higher interviewer effect on recruitment of 13% (ICC = 0.133 = 0.502681/(0.502681+ $\pi^2$/3).

---

[8] For the academic agency, respondents of the two incentive groups that included the bonus were excluded from the analysis to obtain comparability with the market research agency.

The comparison of the variance components of the data separated for the two agencies illustrates large differences. While for the social research agency only about three percent of the total variance is located on the interviewer level (ICC = $0.094/0.094 + \pi^2/3 = 0.028$), for the commercial market research agency about ten percent of the total variance is located on the interviewer level (ICC = $0.360/0.360 + \pi^2/3 = 0.1$). In conclusion, while for the market research agency there are significant and relevant differences between interviewers in the propensity of recruiting respondents, for the social research agency interviewer effects on the outcome variable recruitment are smaller, but also significant. For all three models, the likelihood-ratio test of the multilevel logistic model against a simple logistic model shows that a multilevel model is statistically different from the ordinary model that does not account for the clustered data structure.

*Table 7.2: Two-level logistic regression on recruitment, null models for each agency (odds ratios)*

| | M0 - overall | M0 – social research agency | M0 – market research agency |
|---|---|---|---|
| Intercept | 0.981 | 1.150 | 0.905 |
| $\sigma^2$ respondents | $\pi^2/3$ | $\pi^2/3$ | $\pi^2/3$ |
| $\sigma^2$ interviewers | 0.274 | 0.094 | 0.360 |
| $\rho$ interviewer | 0.077 | 0.028 | 0.099 |
| AIC | 3551.908 | 1927.228 | 2546.514 |
| Likelihood ratio test ($\chi^2$bar(1)) | 76.24*** | 6.03** | 75.88*** |
| Number of respondents | 2619 | 727 | 1892 |
| Number of interviewers | 82 | 19 | 63 |

*Note.* $+ \; p < 0.10$, $* \; p < 0.05$, $** \; p < 0.01$, $*** \; p < 0.000$

Figure 7.5 illustrates the interviewer effect ($\hat{u}_{0j}$ = estimated residuals) in terms of predicted random intercepts in rank order together with 95% confidence intervals for the null models for both agencies (Rabe-Hesketh & Skrondal, 2012).

*Figure 7.5: Caterpillar plot of residuals for both agencies*



The variance is greater for the interviewer staff of the market research agency. For seven interviewers of the market research agency, the 95% confidence interval is above the horizontal line at zero, indicating that the recruitment success for those interviewers is significantly above average. Only two interviewers are significantly below average.

The more homogenous, highly specialized interviewer staff of the social research agency is overall more successful in its task. The necessity to explain interviewer effects by context characteristics of the higher level for the social research agency is limited, as the variance between the interviewers is smaller and none of the interviewers' recruitment success is either significantly above average, nor significantly below average. Therefore, for Research question 7.1, whether interviewers differ in their ability for recruiting respondents, follows that they differ and that the amount of interviewer variance differs between the two fieldwork agencies.

### *Including predictor variables*

The predictor variables of the two levels are included in a step-wise approach following the procedure that is proposed by Hox (2010, p. 11). First, respondent-level variables are integrated. Those include the type of sample (mobile vs. landline) and respondents' sex. The sex is the only demographic information that is available for all respondents including the group of respondents that were recruited without an interview. The recruitment without interview strategy was one measure for the interviewers in order to react flexibly to respondents that were not willing to participate in the telephone recruitment interview. The details are described in Chapter 3.

Second, interviewer-level variables are added to the model. Those include the fieldwork agency, and the sex, education, and age of the interviewer as demographic information. Furthermore, the

interviewers' own hypothetical willingness to participate in an online panel as a proxy for their survey attitude is included, and the general work experience as an interviewer in years, as well as the survey-specific experience in terms of interviews that were conducted before the interview. Third, I test for random slopes, and, finally, include the cross-level interactions. Table 7.3 gives an overview of the results of the step-wise extended models.

Table 7.3: Two-level logistic multilevel regression, odds ratios, variance components, and model fit statistics for the dependent variable recruitment

| | M0 | M1 | M2 | M3 | M4a | M4b | M5 | M6 |
|---|---|---|---|---|---|---|---|---|
| Type of sample: Mobile | | 1.296** | 1.311** | 1.240* | 1.243* | 1.244* | 1.243* | 1.245* |
| Male | | 1.249** | 1.256** | 1.259** | 1.264** | 1.245* | 1.087 | 1.084 |
| Fieldwork agency (Ref.: social research agency) | | | 0.651* | 0.781 | 3.057 | 2.720 | 3.099 | 3.007 |
| Interviewer: male | | | 1.310+ | 1.254 | 0.768 | 0.732 | 0.645 | 0.662 |
| Interviewer: high education | | | 1.051 | 1.086 | 1.166 | 1.119 | 1.161 | |
| Interviewer: Age | | | 0.992 | 0.992 | 0.994 | 0.995 | 0.994 | |
| Interviewer: hypothetical participation | | | 0.960 | 0.957 | 1.124 | 1.108 | 1.125 | 1.123 |
| Number of prev. interviews conducted by interviewer | | | | 1.004** | 1.004** | 1.004** | 1.004** | 1.004** |
| Interviewer: hypothetical participation*agency | | | | | 0.834+ | 0.846+ | 0.834+ | 0.842+ |
| Male*agency | | | | | 1.998* | 1.929* | 1.983* | 1.965* |
| Interviewer: male*respondent: male | | | | | | | 1.430* | 1.436* |
| Constant | 0.981 | 0.789* | 1.703 | 1.334 | 0.334 | 0.394 | 0.355 | 0.313 |
| $\sigma^2$ respondents (rescaled) | 3.290 | 3.262 | 3.262 | 3.233 | 3.233 | 3.233 | 3.215 | 3.215 |
| $\sigma^2$ interviewer | 0.274 | 0.278 | 0.250 | 0.234 | 0.192 | 0.134 | 0.190 | 0.194 |
| $\sigma^2$ RS male | | | | | | 0.026 | | |
| $\rho$ interviewer | 0.077 | 0.078 | 0.071 | 0.067 | 0.055 | 0.039 | 0.055 | 0.056 |
| ICC interviewer | 7.688 | 7.781 | 7.064 | 6.653 | 5.523 | 3.909 | 5.460 | 5.567 |
| AIC | 3551.908 | 3538.720 | 3539.314 | 3533.412 | 3527.980 | 3530.269 | 3525.224 | 3523.235 |
| McKelvy & Zavoina's $R^2$ | 0.000 | 0.008 | 0.023 | 0.026 | 0.039 | 0.035 | 0.041 | 0.039 |
| Chi² LR-test against previous model (degrees of freedom) | - | 17.19** (2) | 9.41+(5) | 7.90(1)* | 9.43*(2) | 1.71 (2) | 4.76*(1) (M4a-M5) | 2.01(2) |
| Number of Respondents | 2619 | 2619 | 2619 | 2619 | 2619 | 2619 | 2619 | 2619 |
| Number of Interviewers | 82 | 82 | 82 | 82 | 82 | 82 | 82 | 82 |

*Note.* $^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

The variance for the respondent level is rescaled following the method of McKelvey and Zavoina (1975) which is described for multilevel models by Hox (2010). Technically the Stata user-written program (ado) meresc written by Dirk Enzmann and Ulrich Kohler is used. It is available under https://ideas.repec.org/c/boc/bocode/s457400.html.

M0: null model without predictor variables; M1: additional respondent level variables kind of sample and sex

M2: additional interviewer level variables: fieldwork agency, demographic information of the interviewer (sex, education, age), their hypothetical participation

M3: additional interviewer level variables: the number of previous interviews of this study

M4a: additional interaction variables with fieldwork agency (male interviewer and the hypothetical participation probability of the interviewer); M4b: random slope for the variable sex; M5: cross-level interaction of interviewers' sex and respondents' sex; M6: parsimonious final model

The null model that includes cases from both agencies is shown in the first column. To answer Research question 7.2, whether different interviewers recruit different respondents, the respondent-level variables are included in a stepwise approach. First, the respondent-level variables – type of sample (mobile vs. landline) and the sex of the respondents – are included (M1). Comparing regression coefficients and variance components across logistic regression models is problematic. The lowest level variance is fixed to $\pi^2/3$ and thus, the scale of the outcome variable changes with the inclusion of additional predictor variables. In order to compare the step-wise extended logistic regression models, the variance for the respondent level needs to be rescaled. In the analysis of this chapter, I follow the approach of McKelvey and Zavoina (1975) which is described for multilevel models by Hox (2010).

Being male and being interviewed on the mobile phone both increase the probability of being recruited. The inclusion of the respondent level variables does not decrease the variance component on the interviewer level. Therefore, for Research question 7.2, whether different interviewers recruit different respondents, follows that this is not the case for the characteristics included in the analyses. Nevertheless, the LR-test of model 1 against the null model indicates a significant improvement by the inclusion of respondent level variables.

In order to answer Research question 7.3, whether the interviewer variance can be explained by characteristics of the interviewers, Model 2 and Model 3 show the step-wise inclusion of interviewer-level variables: fieldwork agency, demographic information of the interviewer (sex, education, age), as well as their own hypothetical willingness of giving consent to online panel participation and the number of previous interviews of this study. The coefficient for the fieldwork agency is significant for Model 2, indicating a lower recruitment probability for the market research agency. Additionally, the sex of the interviewer is significant on the 10%-level in Model 2. However, with the inclusion of the survey-specific experience, the significant effect of the agency and sex of the interviewer diminishes. With increasing number of previously conducted interviews, the recruitment probability increases significantly[9]. The variance component of the interviewers is reduced compared to the null model by about 14% (1-(0.234/0.274)). In Model 4a two interaction variables of being a male interviewer and the own participation probability of the interviewer with the fieldwork agency are additionally included. For interviewers of the market research agency with a more positive attitude towards the panel participation decreases the recruitment probability on the 10%-significance level. The odds of being recruited are almost doubled for male interviewers of the market research agency. The interviewer variance component is decreased by the inclusion of the interaction variables. Finally, in Model 4b a random slope component for the respondents' sex is included. That means the difference

---

[9] The effect of general work experience was also tested, as well as the study-specific previous recruitment success. The general work experience was not significant. The study-specific previous recruitment success is highly correlated (r = 0.9477) with previous number of interviews and thus cannot be included together with this information. As the previous number of interviews is the more general information it was decided to include this measure of study-specific experience.

in the recruitment probability for man and women varies across interviewers. With the inclusion of the random slope for respondents' sex the variance component for the interviewer level is reduced by about one-third compared to the model without random slope. However, the LR-test does not indicate a significant improvement of the model by the inclusion of the random slope, indicating that the inclusion of a random slope for respondents' sex does not increase the model fit. Hence, in the subsequent models, a random slope is not included. To answer Research question 7.4, in the next step, a cross-level interaction of interviewers' sex and respondents' sex is included in Model 5. The interaction term is significant and indicates that the odds of being recruited are higher for male respondents that are interviewed by a male interviewer. The interviewer variance is not reduced by the inclusion of the interaction term. In a final step, the set of included variables is reduced in order to have a parsimonious model. The age and education of the interviewer did not add to the model and are excluded from the final model that can be formalized as follows, with i=respondents and j=interviewers:

$$(2) \quad \ln\left(\frac{p}{1-p}\right)_{ij} = \gamma_{00} + \gamma_1 mobilesample_{ij} + \gamma_2 male_{ij} + \gamma_3 agency_j + \gamma_4 int\_male_j + $$

$$\gamma_5 int\_participation_j + \gamma_6 number\_interviews_j + \gamma_7 int\_participation_j * $$

$$agency_j + \gamma_8 int\_male_j * agency_j + \gamma_9 male_{ij} * int\_male_j + u_{0j}$$

$$\text{with } u_{0j} \sim N(0, \sigma_{u_0}^2)$$

In order to facilitate interpretation of selected research questions, the predicted probabilities are computed based on the final model. First, in Figure 7.6, the predicted probability is shown dependent on the number of interviews that an interviewer already finished within the project. The graph clearly shows a monotonous positive effect for each additional interview up to about 60 interviews. After the turning point, the probability of being recruited decreases with additional interviews. For the interpretation of the decrease, several considerations have to be taken into account. First of all, the number of interviewers that conducted a high number of interviews is decreasing. There are fewer than 10 interviewers that conducted more than 80 interviews. Hence, the prediction of the recruitment probability is based on less than 10 interviewers for the higher numbers of previous interviews. Furthermore, several studies (e.g., Durrant et al., 2010) showed that the interviewers with the highest workload were not the most successful. One possible explanation is the fact that the interviewers with the highest workload are often also working on the more reluctant respondents.

*Figure 7.6: Predicted probability of being recruited by number of previous interviews*
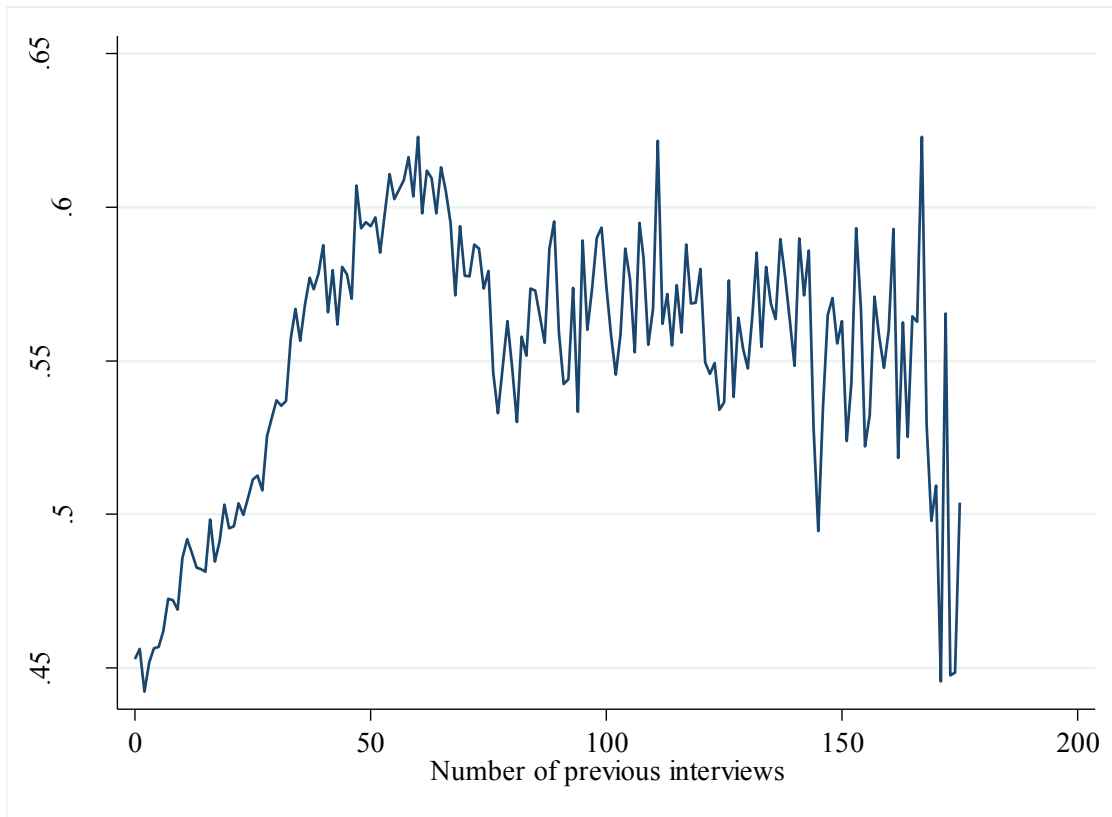


Next, the recruitment probabilities are calculated depending on the interviewer-respondent sex combination for both agencies. Table 7.4 shows for male respondents of both agencies an about 12%-point higher probability of being recruited compared to female respondents if they were interviewed by a male interviewer. For female interviewers, the probability of recruitment is the same for male and female respondents. The structure of the effect of the interviewer-respondents sex combination is different for the two agencies. While for the academic agency the combination of a male interviewer and a female respondent result in the lowest recruitment probability (44.36%), all other combinations resulting in a higher probability recruitment of over 50%; for the market research agency the combination of a male interviewer and a male respondent results in the highest recruitment probability (61.25%), all other combinations resulting in a lower probability of recruitment of about 50%. To sum it up, the match of the sex of interviewer and respondent results in contradicting patterns for the two field work agencies and does not allow for generalization.

*Table 7.4: Predicted probability of being recruited by sex of interviewer and respondent based on the final model*

|  | Social research agency | | Market research agency | |
| --- | --- | --- | --- | --- |
|  | Respondent: male | Respondent: female | Respondent: male | Respondent: female |
| Interviewer: male | 0.5604 | 0.4436 | 0.6125 | 0.4995 |
|  | [0.5454, 0.5754] | [0.4276, 0.4597] | [0.6024, 0.6227] | [0.4889, 0. 5100] |
| Interviewer: female | 0.5577 | 0.5342 | 0.5114 | .4923 |
|  | [0.5460, 0.5694] | [0.5239, 0.5446] | [0.5017, 0.5211] | [0.4817, 0.5029] |

### 7.5.2.2 *Analysis on interviewer effects in experiments*

In order to answer Research question 7.6, whether different interviewers cope differently with the varying features of the survey (promised incentive amount, the length of recruitment interview, and recruitment without interview), two-level logistic regression models that allow for random slopes for the survey features will be conducted. In a stepwise approach, random slopes are tested for the three features incentives, type of interview, and recruitment without interview separately. If the random slope model is adequate further context variables on the interviewer as well as the respondent level will be included (as illustrated in Figure 7.1). For the second part of the multivariate analysis, only interviewers and respondents of the commercial market research agency are included as the information about the type of interview and recruitment without interviewer is not available for the social resarch agency. In sum, 1892 respondents nested in 63 interviewers are part of the subsequent analyses.

**Incentives**

The amount of the incentive that was promised during the telephone recruitment interview for participation in subsequent online surveys is the first factor of the survey design. The respondents were randomly assigned to one of three incentive groups: no incentive control group, 2 Euros promised incentive for participation in each subsequent online survey, and 5 Euros promised incentive for survey participation. The detailed analysis of the incentive experiment can be found in Chapter 6.

First, a random intercept model that includes the amount of promised incentive as predictor variables is estimated:

$$(3) \qquad \ln \left( \frac{p}{1-p} \right)_{ij} = \gamma_{00} + \gamma_1 two_{ij} + \gamma_2 five_{ij} + u_{0j}$$

$$\text{with } u_{0j} \sim N(0, \sigma_{u_0}^2)$$

The model shows a minor change in the estimate of the between-interviewer variance compared to the null model (see Table 7.5), suggesting that the distribution of incentives is similar across interviewers. Nevertheless, we fit the random slope coefficients for the incentive groups to allow their effect to vary across interviewers:

$$(4) \qquad \ln \left( \frac{p}{1-p} \right)_{ij} = \gamma_{00} + \gamma_1 two_{ij} + \gamma_2 five_{ij} + u_{1j} two + u_{2j} five + u_{0j}$$

$$\text{with } u_{0j} \sim N(0, \sigma_{u_0}^2), u_{1j} \sim N(0, \sigma_{u_1}^2) \text{ and } u_{2j} \sim N(0, \sigma_{u_2}^2)$$

The likelihood ratio test shows that the random slopes model does not fit the data better compared to the random intercept model. No indication can be found that interviewers differ in how they use the different amount of incentives ($\chi^2(5) = 6.94$; $Pr > \chi^2 = 0.2254$) for recruiting respondents.

**Type of interview**

Next, a dummy for the kind of recruitment interview is included in the null model. The length of the recruitment interview was randomly assigned to the numbers before the call. 90% of the respondents were assigned to the regular interview of ten minutes duration, and 10% of the respondents were assigned to the short interview group (for details see Chapter 4). Following the procedure of the incentive amount, first, the random intercept model including the type of interview is estimated:

(5) $\quad \ln\left(\frac{p}{1-p}\right)_{ij} = \gamma_{00} + \gamma_1 short\ interview_{ij} + u_{0j}$

$\quad\quad$ with $u_{0j} \sim N(0, \sigma_{u_0}^2)$

It shows that for respondents of the short interview group the probability of being recruited increases by about nine percentage points (prob(recruit)$_{regular\ interview}$= 0.5350 vs. prob(recruit)$_{short\ interview}$= 0.4425). However, the inclusion of the interview type does not change the variance component compared to the null model.

In the second step, the random slope coefficient for the type of interview to allow their effect to vary across interviewers is estimated:

(6) $\quad \ln\left(\frac{p}{1-p}\right)_{ij} = \gamma_{00} + \gamma_1 short\ interview_{ij} + u_{1j} short\ interview + u_{0j}$

$\quad\quad$ with $u_{0j} \sim N(0, \sigma_{u_0}^2)$ and $u_{1j} \sim N(0, \sigma_{u_1}^2)$

The likelihood ratio test indicates no improvement by allowing for random slopes for the type of interview ($\chi^2(2) = 0.09$; $Pr > \chi^2 = 0.9537$). On the basis of the analysis, there is no indication that interviewers can be distinguished by how they recruit respondents in the different interview types. To sum it up, the test with two experimental varied features of the survey reveal no differences between the interviewers how they cope with the varying experimental groups.

*Table 7.5: Two-level logistic multilevel regression, odds ratios, variance components, and model fit statistics for the dependent variable recruitment*

| | Incentives | Incentives random slope | Short interview | Short interview random slope |
|---|---|---|---|---|
| Incentive | | | | |
| *2 Euro* | 1.174 | 1.072 | | |
| *5 Euro* | 1.217+ | 1.210 | | |
| Type of interview: Short | | | 0.702* | 0.688* |
| Constant | 0.803+ | 0.826 | 0.944 | 0.609** |
| Variance respondents (rescaled) | 3.283 | 3.283 | 3.281 | 3.281 |
| Variance interviewer | 0.357 | 0.475 | 0.358 | 0.349 |
| Variance RS five | | 0.269 | | |
| Variance RS two | | 0.287 | | |
| Variance RS short interview | | | | 0.004 |
| Rho interviewer | 0.098 | 0.126 | 0.098 | 0.096 |
| ICC interviewer | 9.782 | 12.622 | 9.803 | 9.586 |
| AIC | 2547.330 | 2550.393 | 2542.889 | 2546.794 |
| McKelvy & Zavoina's R² | 0.002 | 0.002 | 0.003 | 0.003 |
| Chi² LR-test against previous model (degrees of freedom) | – | 6.94(5) | – | 0.09(2) |
| Number of Respondents | 1892 | 1892 | 1892 | 1892 |
| Number of Interviewers | 63 | 63 | 63 | 63 |

*Note.* The variance for the respondent level is rescaled following the method of McKelvey and Zavoina (1975) which is described for multilevel models by Hox (2010); $^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

## Recruitment without interview

Next, the effect of the non-experimental feature of recruitment without interview is examined. Asking the recruitment question without prior interview represented a measure for the interviewers to maximize the recruitment success. Interviewers were allowed to recruit respondents without interview if predefined conditions were met. Those were: the target person declined to participate in telephone interviews in general or had no time at all, at least three appointments failed before, or an appointment within the field period was not possible. The option of recruitment without prior interview was the element of the interview where interviewers were given some flexibility. For this reason, stronger interviewer effects are assumed in comparison to the experimentally varied features. Overall, 18.92% (358) of the recruitment requests of the market research agency were conducted without an interview. Thirty of those recruitment requests were conducted during a short interview, 328 during a regular interview.

First, the random effects model which includes recruitment without interview as an explanatory variable is examined:

(7) $$\ln\left(\frac{p}{1-p}\right)_{ij} = \gamma_{00} + \gamma_1 recruitment\ wo\ interview_{ij} + u_{0j}$$

with $u_{0j} \sim N(0, \sigma_{u_0}^2)$

The model shows a significant positive effect on recruitment for respondents that were asked the recruitment question without an interview. Respondents who were asked the recruitment question without prior interview have almost the double probability of being recruited compared to those who were asked for further participation after having passed the interview (prob(recruit)$_{\text{regular recruitment}}$= 0.4541 vs. prob(recruit)$_{\text{without interview}}$= 0.8243). On the basis of the result, it cannot be concluded that the recruitment without prior interview is superior compared to the regular recruitment with an interview in general. The groups were not randomly assigned, and the small share of respondents that were recruited without interview represents a special group of potential nonrespondents, where interviewers had the freedom to decide to use the measure of recruitment without an interview.

Controlling for recruitment without interview decreases the variance component for interviewers by about one third (1-(0.243/0.360) = 0.325) compared to the null model. The variance reduction is a first indication that the effect of recruitment without interview on recruitment is not similar across interviewers.

In a second step, the random slope for recruitment without interview is included into the model:

$$(8) \qquad \ln\left(\frac{p}{1-p}\right)_{ij} =$$

$$\gamma_{00} + \gamma_1 recruitment\ wo\ interview_{ij} + u_{1j} recruitment\ wo\ interview + u_{0j}$$

$$\text{with } u_{0j} \sim N(0, \sigma_{u_0}^2) \text{ and } u_{1j} \sim N(0, \sigma_{u_1}^2)$$

The likelihood ratio test shows that the random slope model leads to an improvement compared to the random intercept model ($\chi^2(2)$ = 34.32; Pr > $\chi^2$ = 0.0000). This result indicates that interviewers are using the instrument in different manners.

In the next step, additional context variables on the respondent level are included: the set of survey features (incentives and type of interview) and, in a second step, the available respondent characteristic sex. In the third model, the interviewer characteristics are included. And finally, cross-level interactions are included.

The inclusion of the additional survey features does not change the variance component on interviewer level compared to the random slopes model that only includes the dummy variable for recruitment without an interview. This result is not surprising because the predictor variables are located on the lower level. Nevertheless, the likelihood-ratio test shows that including the additional variables improve the model fit. By the inclusion of the interviewer information sex and potential own participation, the interviewer variance is reduced by almost 30% (1-(0.147/0.211) = 0.3033). As in the pooled model for both agencies the general work experience of the interviewers does not significantly add to the explanation and is not included in the shown models. The survey-specific work experience cannot be included in the model along with the information about recruitment without an interview because it is highly correlated by design. Two of the prerequisites of the use of the measure recruitment without interview were the failure of at least three appointments before, and the

impossibility realizing the interview within the field period. Hence, the mean number of interviews that were conducted before was 30.95 (SD = 37.79) for regular recruitment interviews and 43.22 (SD = 32.05) for calls that ended in a recruitment without interview[10]. Finally, the cross-level interaction of interviewers' sex and respondents' sex, that was included in the final model of the pooled analysis above, is included. The positive effect of the combination of a male interviewer and a male respondent is significant and positive as well. The interviewer variance is not reduced additionally by the inclusion of the cross-level interaction. Finally, it is tested whether the random intercept model fits the data as well as the model with a random slope for the recruitment without a prior interview. This is not the case (LR chi²(2) 36.65, prob > chi² 0.000). In consequence, we are not able to explain the interviewer differences in the use of the option recruitment without an interview with the variables that are at hand.

*Table 7.6: Two-level logistic multilevel regression, odds ratios, variance components, and model fit statistics for the dependent variable recruitment and factor recruitment without interview*

| | M1 | M2 (random slope) | M3 (random slope) | M3 (random slope) | M4 (random slope) |
|---|---|---|---|---|---|
| Recruitment without interview | 5.572** | 4.050** | 4.046** | 3.886** | 3.872** |
| Incentives | | | | | |
| 2 Euro | | | 1.204 | 1.200 | 1.212+ |
| 5 Euro | | | 1.233+ | 1.235+ | 1.234+ |
| Type of interview: short | | | 1.244* | 1.261* | 1.255* |
| Mobile | | | 0.747* | 0.746* | 0.752+ |
| Male | | | 1.205+ | 1.218* | 1.035 |
| Interviewer: male | | | | 1.432* | 1.178 |
| Interviewer: hypothetical participation | | | | 0.939* | 0.940* |
| Interviewer: male*respondent: male | | | | | 1.468* |
| Constant | 0.735** | 0.760** | 0.572** | 0.777 | 0.843 |
| $\sigma^2$ respondents (rescaled) | 2.929 | 2.929 | 2.912 | 2.901 | 2.901 |
| $\sigma^2$ interviewer | 0.243 | 0.225 | 0.211 | 0.147 | 0.147 |
| $\sigma^2$ RS recruitment without interview | | 2.724 | 2.737 | 2.878 | 2.871 |
| $\rho$ interviewer | 0.069 | 0.064 | 0.060 | 0.043 | 0.043 |
| ICC interviewer | 6.880 | 6.405 | 6.040 | 4.284 | 4.278 |
| AIC | 2398.909 | 2368.592 | 2362.426 | 2355.977 | 2354.082 |
| McKelvy & Zavoina's R² | 0.112 | 0.043 | 0.049 | 0.057 | 0.058 |
| Chi² LR-test against previous model (degrees of freedom) | | 34.32(2)*** | 16.17(5)** | 10.45(2)** | 3.89(1)* |
| Number of Respondents | 1892 | 1892 | 1892 | 1892 | 1892 |
| Number of Interviewers | 63 | 63 | 63 | 63 | 63 |

*Note.* The variance for the respondent level is rescaled following the method of McKelvey and Zavoina (1975) which is described for multilevel models by Hox (2010); + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.000$

---

[10] The median for regular interviews was 18, the median for recruitment without interview was 38.

## 7.6 Summary and discussion

The aim of this chapter was to quantify the interviewer variance on the propensity of being recruited (Research question 7.1). It was aimed to explain parts of the interviewer variance by means of respondent information (Research question 7.2), interviewer characteristics (Research question 7.3), by the combination of interviewer and respondents characteristics (Research question 7.4), and the type of fieldwork agency the interviewers worked for (Research question 7.5). The innovative part of the analysis was to investigate whether interviewers differ in their ability to use the experimentally varied features of the survey (incentive amount, the length of recruitment interview) and the way they used the non-experimental feature recruitment without interview (Research question 7.6).

I showed that interviewers differed considerably in their ability to recruit respondents (Research question 7.1) and the magnitude of the effect differs between the two agencies. While for the academic fieldwork agency, only about three percent of the total variance is located on the interviewer level, for the market research agency, about ten percent of the total variance is located on the interviewer level. This is also reflected by the composition of the interviewer staff. The interviewers of the academic, social research agency were more homogeneous, highly educated, very experienced and highly specialized in the field of social sciences research. In contrast, the staff of the market research agency was more heterogeneous, less experienced and with a focus on market research projects. These findings may have implications for the interviewer recruitment of agencies, as well as for interviewer training. These aspects are discussed below.

In order to explain parts of the interviewer variance on the recruitment propensity (Research question 7.2 to Research question 7.4), the pooled data including interviewers and respondents from both field work agencies were analyzed. The inclusion of respondent level variables respondent sex (Research question 7.2) and type of sample did not explain any of the residual interviewer variance.

On the interviewer level (Research question 7.3), the focus was on work experience. In contradiction to the assumption, the general work experience of the interviewers does not significantly contribute to the explanation of the recruitment. In contrast, the survey-specific experience significantly adds to the explanation of recruitment. With each additional interview, the recruitment probability increases up to the threshold of about 60 interviews. The finding shows, that for the specific task of gaining consent at the end of a successfully conducted interview the interviewers learn while conducting the recruitment. One possible reason for the u-shaped relation may be that interviewers that performed well during the field phase were assigned to more complicated cases at the end.

Regarding Research question 7.4, whether the combination of interviewer and respondents characteristics contributes to the explanation of recruitment, the results are not homogeneous for the two fieldwork agencies. The structure of the effect of the interviewer-respondents sex combination differs between the agencies. While for the social research agency, the combination of a male

interviewer and a female respondent results in the lowest recruitment probability (44.36%), all other combinations result in a higher recruitment probability of above 50%; for the market research agency the combination of a male interviewer and a male respondent results in the highest recruitment probability (61.25%), all other combinations result in a lower probability of recruitment of about 50%. The contradicting findings for the matching of respondents' and interviewers' sex reflect the findings of earlier studies (Durrant et al., 2010; Lipps, 2010b) that are also equivocal. Matching based on limited demographic characteristics as an easy heuristic for liking is not sufficient. On the basis of the results, fieldwork agencies should rather enhance their recruitment and training strategies for telephone interviewers, instead of trying to enhance the success by applying matching strategies.

In Research question 7.5, I asked, whether the performance of the interviewers of the two fieldwork agencies differ. The analyses showed that the interviewer variance on recruitment is larger for the market research agency compared to the social research agency. However, none of the included characteristics on interviewer level was able to explain the differences between the two agencies. Crucial factors that could not be included on a third level because of the limited number of agencies are the strategy of interviewer selection and interviewer training. A thorough recruitment of the interviewer staff (Alcser, Clemens, Holland, Guyer, & Hu, 2016) and sound basic training is the key to minimizing the interviewer error component (West & Elliot, 2014). The differences in the recruitment performance are reinforced by the differences in the initial telephone survey response rate (AAPOR RR 3) that was 26.14% for the social research agency and only 14.08% for the market research agency.

In the second part of the analysis, only interviewers and respondents of the market research agency were taken into account due to data restrictions. In Research question 7.6, I asked whether interviewers cope differently with varying features of the survey. I showed that the effect of the experimentally varied factors "amount of promised incentive" and the "type of recruitment interview" is uniform across interviewers. The effect of the experimental conditions on recruitment does not depend on the experience of the interviewers as it was assumed above.

However, for the possibility of recruiting respondents without a prior interview, I showed that interviewers used the measure in different ways. The possibility of skipping the recruitment interview under predefined circumstances was a measure that gave interviewers some freedom to decide. The interviewers of the market research agency made use of it in different ways. In contradiction to the formulated assumption, the general job experience does not contribute to the explanation of this effect. The survey-specific experience could not be included due to collinearity. The innovative analysis shows in sum, that in cases where interviewers are free to decide on the use of a measure interviewer differences occur. In order to assure the quality, even in non-standardized parts of the telephone survey, a consequent supervision is recommended. On the other hand, it also shows that the effect of experimentally varied survey features is uniform across interviewers. A result that is highly desirable from a data quality perspective.

On the basis of the analysis of interviewer effects on the recruitment propensity in a telephone recruitment interview, I derive several recommendations for agencies that are conducting telephone interviews. First of all, invest in interviewer recruitment and training strategies. Besides a thorough basic training, survey-specific training is of particular importance. For tasks that go beyond the everyday task of interviewing and gaining cooperation, such as the panel recruitment, the integration of training interviews under most realistic circumstances in the study-specific interviewer training is recommended. Researchers that are looking for a high-quality fieldwork agency are recommended to rely on a provider with experience in the field of the respective study. Ideally, the agency is able to provide information about their training and recruitment concept of the interviewers.

The analyses and results of this chapter are limited in a number of ways. First of all, the respondent information is limited to the sex of respondents due to the experimental design on the interview length. Hence, the analyses of the respondent-interviewer matching had to be limited to the sex, a factor that might not be the best proxy for "liking." Furthermore, the data do not reveal any information on the interaction and strategies that interviewers applied during the process of the recruitment. Hence, a tape recording of the interaction of the recruitment process would be of great value. It can be assumed that interviewers use the information they gain during the prior interview to tailor their request.

The second part of the the analysis suffered from the exclusion of the interviewer staff of the social research agency due to data limitations. The findings on interviewer success depending on experimental variations would profit from the inclusion of interviewers that are specialized in the field of social research and highly skilled. Even though the two field work agencies can be interpreted as prototypical examples of the German survey environment, a generalization about what is desirable regarding the interviewer staff is not possible.

To systematically evaluate the effect of interviewer recruitment, the basic training, as well as supervising strategies on the interviewer variance, a greater number of agencies that ideally conduct the same project have to be studied in a next step. As a future avenue of research, I see a study that includes several agencies in order to empirically derive factors that are the key for interviewer success. Those factors include recruitment strategies, general interviewer training strategies, study-specific training strategies, as well as supervising strategies. On the basis of such a study, practical recommendations for general interviewer training could be derived, and existing guidelines for general interviewer training that are based on best-practices (e.g. Alcser et al., 2016; Wengrzik, Silber, Bosnjak, Zabal, & Martin, 2017) could be supplemented.

# 8. Summary, discussion, and future research

Online surveys have become the predominant survey mode in commercial survey research (Callegaro et al., 2014). Thanks to the many advantages of online surveys, scientific communities are increasingly interested in using them as well. To meet the requirements of social sciences research, online surveys need to be based on a probability sample of the population of interest. A probability sample is a prerequisite for drawing conclusions that are generalizable to the reference population. However, for online surveys, there is neither a sampling frame available (e.g., email lists of all country residents), nor the possibility of generating email addresses that would cover the whole population equivalent to RDD sampling procedures for telephone surveys. Probability-based online panel infrastructures are one possibility to overcome the problems of sampling and make use of the advantages of online surveys (Blom, Bosnjak, et al., 2016; Bosnjak et al., 2016; Hays et al., 2015). They represent a pool of potential respondents that are recruited in an offline mode, for instance, mail, telephone, personal (Couper, 2000), and then are invited, in a second step, to the regular online survey waves. Due to the multi-step approach of recruitment, probability-based online panels are particularly prone to nonresponse (Bartsch, 2011; Bosnjak, 2002; Lensvelt-Mulders, Lugtig, & Hox, 2008). The resulting cumulative response rates across stages are, thus, low compared to more traditional modes of data collection (Callegaro & DiSogra, 2008; Tourangeau, Conrad, & Couper, 2013, p. 41).

There are several panels which share the trait of multistep recruitment process (Blom, Bosnjak, et al., 2016; Bosnjak et al., 2016; Hays et al., 2015). Besides the common base of a probability sample, the recruitment protocols are as numerous as the panels are. Some design decisions, such as the sampling procedure, depend on country-specific peculiarities; other design decisions are based on best-practice knowledge from survey research. However, only a minority of decisions is tested experimentally (Bosnjak, Das, & Lynn, 2016).

In cases where experimental evidence exists, the main focus of experiments on panel recruitment has so far been on the recruitment mode (e.g., Rao et al., 2010; Scherpenzeel & Toepoel, 2012), the amount and timing of incentive at the different nonresponse stages (e.g., Blom et al., 2015; Scherpenzeel & Toepoel, 2012), and on the logistics around the recruitment interview, for instance, the use and timing of a reminder letter (e.g., Martinsson & Riedel, 2015; Rao et al., 2010). To date, there is a lack of experimental evidence on the design of the recruitment interview.

With this dissertation, I aimed to fill the gap and contribute experimental evidence about design decisions of a recruitment interview for a probability-based online panel. The central question of this dissertation was: How can we optimize the telephone recruitment process of a probability-based online panel in Germany? Based on the framework of survey participation (Groves & Couper, 1998), the four studies of this dissertation focus on several aspects of the recruitment process that researchers can decide upon and have control about. In three survey experiments that I conducted as the basis for this dissertation, I examine the effect of varying survey features on the success of the recruitment process.

The experimental factors are: the length of the recruitment interview, the inclusion of a sensitive question, and incentives. In addition, I incorporate the role of the interviewers in the panel recruitment process. The common data source of all studies of this dissertation is the data from the recruitment of a probability-based online panel: the GESIS Online Panel Pilot.

In the subsequent section, I give an overview of the dissertation and summarize the main findings of each of the four studies of the dissertation. In the last section, I briefly discuss limitations of the dissertation and future research.

## 8.1 Summary of the results

Chapter 1, provides an overview on the motivation of this dissertation. In Chapter 2, I describe the framework of survey participation as the conceptual framework of this dissertation. This framework captures the multitude of factors that are assumed to influence the decision to participate in a panel. Additionally, I provide a review of the pertinent literature, identify research gaps, and formulate the research questions. In Chapter 3, I present the central data source of this dissertation: GESIS Online Panel Pilot. In addition, I define the central dependent variables that are used throughout the analyses of the experiments and briefly describe the experiments. In Chapters 4, 5, and 6, I present the design, analysis, and results of the survey experiments of the three factors: length of the recruitment interview, the inclusion of a sensitive question, and incentives. The analyses of the experiments follow the logic of the multi-step recruitment process. Three research questions guided the analyses of the survey experiments:

Research question 1: Does the experimental variation of the three selected survey factors−length of the recruitment interview, inclusion of a sensitive question, and amount of promised incentive−affect the recruitment probability?

Research question 2: Does the experimental variation of the amount of promised incentive affect the probability for online participation in a probability-based online panel?

Research question 3: Does an experimental variation of the selected survey design features−length of the recruitment interview, inclusion of a sensitive question, and amount of promised incentive−introduce sample composition bias at the stage of recruitment and the stage of online participation?

In Chapter 4, the effect of the first factor of survey design, the length of the recruitment interview, on the quality indicators was examined. Questionnaire length is one factor that is assumed to contribute to the respondent burden. The study tested two versions of the telephone recruitment interview: one with the duration of 3 minutes vs. 10 minutes duration. The tested versions of the recruitment interview were both short compared to other surveys. Thus two competing hypotheses were formulated based on the factor−burden and trust−that influences the decision to participate in the panel. The shorter interview is assumed to be superior compared to the longer one if the decision of the respondent is based on the perceived burden of the interview. Given the short interview time of both versions, it

could, in contrast, also be assumed that the short interview time is not sufficient for the interviewer to establish trust in the survey organization. The analysis revealed that the shorter interview does not increase the recruitment probability compared to the longer version. At the stage of the recruitment, the longer version is more successful in terms of recruitment with an eight-percentage-point increase in the recruitment rate. However, the difference between the two groups does not stay significant when controlling for differences in sample composition at the stage of the interview. There is some indication that at the stage of the decision to participate in the telephone interview, different respondents are attracted by the differing length of announced interviews. The respondents of the short interview group had a significantly lower education level compared to the respondents from the longer interview group. However, the sample composition of the respondents that were recruited into the panel and that of the resulting online panel was not affected by the experimental treatment. In order to improve sample composition regarding education, a targeted strategy could be applied. However, prior knowledge about the level of educational is needed to target the interview length. In sum, based on the analysis, I would recommend choosing the longer version of the interview, as researchers may benefit from the additional information that is collected about the respondent.

In Chapter 5, the effect of the second factor of survey design, the inclusion of a sensitive question, on the recruitment probability and the sample composition was investigated. During the process of designing a recruitment interview, researchers are often concerned about including questions that are perceived as being sensitive by the respondents. The split-half experiment tested the effect of including a question about the household income versus not including it in the recruitment interview. The analysis revealed that respondents of the experimental group did not have a lower recruitment probability compared to the respondents in the control group. However, respondents that refused to provide the income information had a recruitment probability that was almost halved compared to the respondents that provided the income information. Income item nonresponse is known as being an indicator for reluctant respondents. It could be that reluctant respondents have a lower recruitment probability because they are asked about the income, or they refuse to provide the income information and the panel consent because they are reluctant respondents. For the recruitment process, the direction of causality was not evident. The analysis revealed that the second interpretation is more likely. This knowledge can be used in future multi-step recruitment procedures for applying targeted recruitment strategies to reluctant respondents. As possible measures to increase the recruitment probability of respondents with privacy concerns, interviewers can stress confidentiality issues and the trustworthiness of the institution.

In Chapter 6, the effect of the third factor of survey design, the promised incentive, on the recruitment probability, the online participation probability, and the sample composition was analyzed. Incentives represent a response-enhancing measure that is also used at the various stages of the multi-step recruitment process. This study investigated the potential of promised incentives for online participation to affect nonresponse at several stages of the multi-step recruitment process.

First of all, the study revealed that in the context of a multi-step recruitment, promising an incentive increases the probability of providing consent to become a panel member. This finding is promising for panel recruitment strategies that are based on an RDD-sample without addresses available for sending a prepaid incentive. The analyses show that at the stage of recruitment, the incentive amount needs to be at least 5 Euros promised for each finished online survey to increase the recruitment probability. At the stage of panel participation where the incentive is directly connected with the action of the respondent (i.e., survey participation), increasing the incentive increases the probability of online participation. In addition, the control group provides a good baseline estimation of the recruitment success that can be expected without an incentive. The cumulative response rate that summarizes the recruitment and panel participation stages results in 21%. The innovative part of the study was to test the effect of promising a bonus for loyal respondents that participate in all online surveys of the project. It was shown that adding a bonus for loyal respondents had a double positive effect: first, it increased the participation probability in the online surveys, and second, it increased the proportion of loyal respondents that participated in all eight surveys of the study. Particularly in longitudinal studies, one of the aims is to collect measurements for all respondents at every occasion (Lynn, 2009). As a consequence, promising a bonus for participation could be a flexible instrument for a provider of probability-based access panels to assure data with fewer missing waves and high quality.

The comparison of the sample composition across the experimental groups did not reveal systematic differences at both stages under consideration. In contradiction to the assumption, I did not find a higher share of respondents that are usually underrepresented in surveys of the general population among the higher incentive groups. For the given circumstances the promised incentive of 5 euros with an additional bonus was shown to be most effective.

In contrast to the three previous chapters that focused on factors of the survey design, Chapter 7 focused on the interviewer as an additional factor that influences the recruitment success. While the design factors are specific to the recruitment process, the analysis of the interviewer success is of broader relevance. The chapter was divided into two separate analysis parts. The aim of the first part was to quantify and explain interviewer variance on the propensity of recruiting respondents. The second part focused on the research question of whether interviewers differed in their ability to use the experimentally varied features of the survey (incentive amount, the length of the recruitment interview).

The first part of the analysis included interviewers from two fieldwork agencies that conducted the recruitment interviews. The two agencies represented prototypes of agencies of the German survey landscape. The first one was an academic agency with a strong focus on social research and a homogenous and highly-qualified interviewer staff. The second one was a market research agency with less experienced and less qualified interviewers that were more heterogeneous in their characteristics. The study revealed major differences in the interviewer variance on panel recruitment

between the two fieldwork agencies. The variance in the ability to recruit respondents was smaller for the interviewers of the social research agency compared to the interviewers of the market research agency. In the analysis, none of the available characteristics on interviewer level were able to explain the differences between the two agencies. In contradiction to the formulated assumptions, general work experience of the interviewers does not explain the differences in their recruitment abilities. In contrast, the survey-specific experience in terms of number of interviews conducted significantly adds to the explanation of recruitment propensity. Once a certain number of interviews has been reached, the recruitment success decreased. Thus, the interviewers learned to optimize their recruitment approach in the course of the project. One possible explanation for the diminishing additional gain for interviewers that conducted a large number of interviews may be that interviewers that performed well during the initial field phase were assigned to more complicated cases at the end.

The effect of the interviewer-respondents sex combination on recruitment differed between the two fieldwork agencies. Hence, interviewer-respondent matching based on one single characteristic is not a promising strategy. Possible explanations for the overall difference between the agencies can be found in the different quality and composition of the interviewer staff of the two agencies.

In an ideal study, characteristics of the agency would be included into the multilevel model as a third level. To systematically evaluate the effect of factors, such as the interviewer recruitment strategies, the basic training, as well as supervising strategies on the interviewer variance, a greater number of agencies that ideally would be conducting the same project need to be studied in a next step.

The second part of the analysis focused on the interviewers of the market research agency and their ability to conduct the experiments in a standardized way. The analysis revealed that the effect of experimentally varied survey features on recruitment is uniform across interviewers. This is a result that is highly desirable from a data quality perspective in standardized interviews. In contrast, for the possibility of recruiting respondents without a previous recruitment interview, I showed that interviewers used the measure in different ways. The possibility of skipping the recruitment interview under predefined circumstances was a measure that was not varied experimentally. Interviewers were given some freedom to decide. The different interviewers made use of it in different ways. In contradiction to the formulated assumption, the general job experience does also not contribute to the explanation of this effect. To assure the quality, even in non-standardized parts of the telephone survey, a consequent supervision is recommended. On the basis of the results, I recommend a thorough assessment of a fieldwork agency before selection. Ideally, an agency would disclose its interviewer training strategies, field-specific experience as well as the detailed information about the qualifications of its interviewer staff.

## 8.2 Limitations and suggestions for future research

This dissertation is a work that has practical relevance for survey research and aims to contribute to evidence-based knowledge on the recruitment process of probability-based online panels. The four studies of this dissertation have some limitations that will be discussed in this section along with suggestions for future research that arise from this work and go beyond its scope.

The pilot nature of the GESIS Online Panel Pilot imposed some restrictions on the design of the survey experiments. The aim of the project was not only to test factors that affected recruitment but also to reveal insights in maintaining a scientific probability-based online panel. As a consequence, treatment conditions that were assumed to result in very low recruitment rates were not implemented in order to achieve a sufficient number of respondents in the active panel.

The main limitation of the overall experimental design is the fact that the tested factors are not fully crossed, which results in fractional factorial design (Shadish, Cook, & Campbell, 2002). A full factorial design that would have allowed for all possible combinations would have required 2x2x3x2x2=48 (questionnaire length x sensitive question x incentive amount x bonus condition x fieldwork agency) conditions, which would have required too many cases for a pilot study. Thus the three experimental factors were tested separately, and wherever possible the fieldwork agency was additionally controlled for in the models.

The main drawback of the experimental design on the length of recruitment interview that was presented in Chapter 4, is the focus on two versions of a comparably short interview. An ideal study would investigate the effect of more than two versions of the recruitment interview. It would be worth investigating the effectiveness of calling respondents and immediately asking them to consent to panel participation. This strategy was offered as an exit option for reluctant respondents under predefined circumstances. The recruitment rate for respondents that were recruited without prior interview was high compared to respondents that were asked at the end of the interview. As this exit option was not varied experimentally, the effectiveness of this measure on general recruitment success cannot be evaluated. Concurrently, the effect of a recruitment interview that was longer than ten minutes would also be worth investigating. Implementing an experiment that includes, for instance, four versions of the recruitment interview lasting from immediate recruitment without interview to 20 minutes of interview time, would more appropriately allow the investigation of the relation between questionnaire length and panel consent. In combination with measures of respondent trust in the survey organization and the perceived survey burden, it would provide the opportunity to reveal insights into the competing mechanisms trust vs. burden that affect the respondents' decisions.

The treatment of the experiment on the sensitive question in Chapter 5 was designed as one question about the household income with bracketed answer options. It needs further investigation whether the manipulation was sufficiently powerful (Mutz, 2011). The bracketed answer option is known to reduce income item-nonresponse (Juster & Smith, 1997). Asking about the income by means of an open-ended question or in a more extensive manner was not tested. The experiment was a first step in

investigating the interrelation between topic-specific item nonresponse and next wave nonresponse. The emergence of probability-based access panels with high-frequency data collections provides the opportunity for subsequent research. The connection between item and unit nonresponse in high-frequency panels needs to be tested. In this context, the effect of additional types of sensitive questions, such as for instance, questions about racial prejudice or deviant behavior, on next wave nonresponse could be tested. Furthermore, not only sensitivity is an issue in the context of high-frequency data collection but also the effect of repeated topics. In access panels where external researchers submit their questionnaire modules, the panel provider has limited control over the topics. In consequence, several consecutive waves can include questions that are perceived as being similar and redundant by the respondents. This challenge, that is specific to access panels, opens the opportunity for a new thread of research that investigates the interrelation of topic-specific nonresponse and unit nonresponse in the subsequent panel waves.

Chapter 6 provided insights into possible strategies of incentive schemes for promised incentives. Due to the pilot nature of the study, the incentive experiment was implemented in a stepwise approach. This fact results in a design without fully crossed factors "incentive amount" and "additional bonus." The bonus condition was implemented only along with the higher incentive groups. In consequence, there is no evidence on the effect of the bonus in combination with the lower incentive groups. The use of a bonus is an aspect that is worth following up. Probability-based online access panels include cross-sectional as well as longitudinal studies. For longitudinal studies minimizing unit nonresponse is a major concern. Promising respondents a bonus for participation in all waves that include a specific longitudinal study is a strategy that is worth to be tested.

Chapter 7 that investigated interviewer effects on recruitment suffered from limited information about respondents, as well as information about interviewers that could be used as explaining variables. Furthermore, the data do not reveal any information on the interaction and strategies that interviewers applied during the process of the recruitment. Hence, a tape recording of the interaction of the recruitment process would be of great value. It can be assumed that interviewers use the information they gain during the prior interview to tailor their panel participation request.

Furthermore, the second part of the analysis suffered from the exclusion of the interviewer staff of the social research agency due to data limitations. The findings that interviewers do not cope differently with experimental variations would profit from the inclusion of interviewers that are specialized in the field of social research and are highly skilled. Even though the two field work agencies can be interpreted as prototypical examples of the German survey environment, a generalization about what is desirable regarding the interviewer staff is not possible. To systematically evaluate the effect of interviewer recruitment, the basic training, and supervising strategies on the interviewer variance, a greater number of agencies, that ideally conduct the same project, would have to be studied in a next step. As a future avenue of research, I envision a study that includes several agencies in order to empirically derive factors that are key for interviewer success. Those factors include recruitment

strategies, general interviewer training strategies, study-specific strategies, as well as supervising strategies. Based on such a study, practical recommendations for general interviewer training could be derived, and existing guidelines for general interviewer training that are based on best-practices could be supplemented (e.g. Alcser et al., 2016; Wengrzik et al., 2017).

In sum, the dissertation provides empirical evidence on the effect of selected factors of the survey design of the telephone recruitment interview for a probability-based online panel on the success of the panel. There are many other aspects influencing the success of recruitment that were not considered in this dissertation, for instance, the overall topic and content of the recruitment interview. Based on general recommendations from survey research, researchers choose questions of a broad appeal that are assumed to be interesting for the majority of respondents. The aim is to design the interview in such a manner that the majority of respondents have a good survey experience and provide respondents with a preview of what they can expect in the panel (see for example Bosnjak et al., 2017). To date, there is a lack of experimental evidence specific to panel recruitment interview that could guide the choice of topics and questions for a recruitment interview.

In the last section, I focus on research on probability-based online panels beyond the scope of the present work. The topic of the dissertation was motivated by the fact that the majority of the probability-based access panels recruit respondents by means of a separate recruitment interview. A second recruitment strategy is to use an established survey as a vehicle for asking respondents for panel consent. The American Gallup Panel and American Life Panel are the two examples that follow this strategy. In Europe, there are three panels that recently tested the possibility of recruiting respondents without a separate recruitment interview: the CROss-National Online Survey Panel (CRONOS), the NatCen panel (NatCen, 2017), and the GESIS Panel. The CRONOS panel used the European Social Survey (ESS) round 8 as a vehicle for recruiting respondents at the end of the interview in selected countries (UK, Estonia, Slovenia) (Villar & Sommer, 2017). Respondents without internet access are provided with access and a device. In the course of the pilot project, panel maintenance and quality enhancement strategies were tested experimentally. The project aims to publish an integrated data set that combines the interview data from the ESS and the online surveys.

The NatCen panel was recruited during the British Social Attitudes survey, and the GESIS Panel conducted a pilot study together with the German General Social Survey 2016 (ALLBUS) for its refreshment sample (Schaurer, Weyandt, Bauer, Blohm, & Weiß, 2017).

This strategy provides several opportunities for both sides, the panel provider, as well as the established surveys, which serve as a vehicle for panel recruitment. The panel provider has the opportunity for recruiting respondents without the need of a separate recruitment interview. As the personal interview represents a large share of the costs, this strategy is cost-effective. The established survey has the opportunity to enrich the data set of the face-to-face survey. Researchers have the opportunity to collect additional information, test for stability and change of the measures over time, and can implement experiments more easily. However, the strategy raises questions on data quality

that need to be evaluated. What are the recruitment rates after a long personal interview that was not optimized for recruitment compared to the stand-alone recruitment interview? Is there additional selectivity introduced by the long face-to-face interview and its topics? What is the attrition behavior of respondents that were recruited after an established survey compared to respondents that were recruited by means of a separate recruitment interview? All these research questions need to be answered in order to evaluate the success of the new approaches of online panel recruitment. The experimental research provided in this dissertation can be useful for designing experimental studies aimed to gain further insights into these newly developed recruitment strategies.

# Appendices

## Appendix Chapter 3

### CATI recruitment interview GOPP

**Introduction landline**

Guten Tag/ mein Name ist XXX von der wissenschaftlichen Forschungseinrichtung GESIS in Mannheim. Wir führen zurzeit eine Umfrage zum Leben und Internetnutzung in Deutschland durch und Ihr Haushalt wurde dafür zufällig ausgewählt.

Für diese wissenschaftliche Befragung, die maximal 10 (maximal 3) Minuten dauert, müssen wir diejenige Person in Ihrem Haushalt auswählen, die mindestens 18 Jahre alt ist und zuletzt Geburtstag hatte. Sind Sie das oder ist es eine andere Person?

**Introduction cell phone**

Guten Tag/ Hallo mein Name ist XXX von der wissenschaftlichen Forschungseinrichtung GESIS in Mannheim. Wir führen zurzeit eine wissenschaftliche Umfrage zum Leben und Internetnutzung in Deutschland durch und Sie wurden dafür zufällig ausgewählt. Es dauert maximal zehn (3) Minuten . Darf ich fragen, wo ich Sie gerade erreiche?

- 1 zuhause
- 2 Auf der Arbeit
- 3 Im Auto
- 4 In einem öffentlichen Verkehrsmittel
- 5 In einem Geschäft, beim Einkaufen
- 6 im Ausland
- 7 anderer Ort, sonstiges
- 8 Telefon weggedrückt, legt sofort auf

Wichtig ist, dass Sie mindestens 18 Jahre alt sind. Trifft das auf Sie zu?

- 1 ja, mindestens 18 Jahre alt
- 2 nein, unter 18 Jahre
-

**Frage 1 (regular and short interview)**

Ich lese Ihnen nun eine Aussage vor. Bitte sagen Sie mir, ob Sie der Aussage voll und ganz zustimmen, eher zustimmen, eher nicht zustimmen oder gar nicht zustimmen.

Alles in allem kann ich in einem Land wie Deutschland gut leben. [Stimmen Sie der Aussage voll und ganz zu, eher zu, eher nicht zu oder gar nicht zu?]

- 1 stimme voll und ganz zu
- 2 stimme eher zu
- 3 stimme eher nicht zu
- 4 stimme gar nicht zu

- 5 Teils-teils [nicht vorlesen ]

**Frage 2 (regular and short interview)**

In den letzten Jahren hat das Internet im Alltag immer mehr an Bedeutung gewonnen. Wie ist das bei Ihnen, nutzen Sie privat das Internet?

- 1 ja
- 2 nein → exit

**Frage 3**
Können Sie mir in etwa sagen, seit welchem Jahr Sie das Internet nutzen?

Jahreszahl vierstellig gültige Werte: 1980-2011]

**Frage 4**
Wie häufig nutzen Sie privat das Internet?
Ist das…
- 1 Täglich
- 2 Mehrmals die Woche
- 3 Mindestens einmal die Woche
- 4 Mindestens einmal im Monat
- 5 Seltener als einmal im Monat

**Frage 5**
Haben Sie innerhalb der letzten zwölf Monate an einer Umfrage im Internet teilgenommen?
[Int.: Denken Sie dabei an alles was Sie als Internetumfrage ansehen würden. z.B. Kundenbefragungen oder Websitebefragungen oder Befragungen zu denen Sie per Email, Weblink oder Popup eingeladen wurden.]

- 1 ja, an einer Umfrage teilgenommen
- 2 ja, an mehreren Umfragen teilgenommen
- 3 nein

**Frage 6**
Kommen wir nun zu ein paar Fragen zum Leben in Deutschland.
Manche Leute sagen, dass man den meisten Menschen vertrauen kann. Andere meinen, dass man nicht vorsichtig genug sein kann im Umgang mit anderen Menschen. Was ist Ihre Meinung dazu? Denken Sie,
- 1 Man kann den meisten Menschen vertrauen
- 2 Man kann nicht vorsichtig genug sein
- 3 Oder kommt es darauf an

**Frage 7**
Denken Sie jetzt einmal an Ihre persönliche Situation: wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?
Bitte antworten Sie auf einer Skala von 0 bis 10. Wobei 0 bedeutet „ ganz und gar unzufrieden" und 10 „ganz und gar zufrieden".

Wert 0-10

**Frage 8**
Wenn Sie in Ihrer Wohngegend draußen alleine sind, wie sicher fühlen Sie sich?
Fühlen Sie sich…
- 1 Sehr sicher
- 2 Ziemlich sicher
- 3 Ziemlich unsicher
- 4 Oder sehr unsicher
- 5 Teils-teils [nicht vorlesen]

**Frage 9**
Welche der folgenden Kategorien beschreibt am besten, wo Sie wohnen?
Wohnen Sie in…
[Int.: Bei Unsicherheit bzgl. der Zuordnung:

* Großstadt ab 100.000 Einwohner
* Rand- oder Vorort einer Großstadt: eingemeindetes Dorf, Vorort etc. einer Stadt ab 100.000 Einwohnern
* Mittel- oder Kleinstadt: jede Stadt bis 100.000 Einwohner]

- 1 Einer Großstadt
- 2 Einem Rand- oder Vorort einer Großstadt
- 3 Einer Mittel- oder Kleinstadt
- 4 Oder in einem Dorf

**Frage 10**
Was meinen Sie, wird die heutige junge Generation, im Vergleich zu ihren Eltern, später einmal einen höheren, einen niedrigeren oder denselben Lebensstandard haben?

- 1 Eher höheren Lebensstandard
- 2 Eher niedrigeren Lebensstandard
- 3 Denselben Lebensstandard

**Frage 11**
Ich nenne Ihnen nun verschiedene Lebensbereiche. Bitte sagen Sie mir jeweils, ob Sie erwarten, dass in den nächsten 5 Jahren die Situation in Deutschland in dem jeweiligen Bereich eher besser oder eher schlechter wird oder ob Sie keine Veränderung erwarten?
Wie ist das mit dem **Arbeitsmarkt**? Wird die Situation in den nächsten fünf Jahren eher besser, eher schlechter oder erwarten Sie keine Veränderung?

- 1 Eher besser
- 2 Eher schlechter
- 3 Keine Veränderung

**Frage 12**
Wie ist das mit den Lebensbedingungen für Familien und Kinder? Wird die Situation in den nächsten fünf Jahren eher besser, eher schlechter oder erwarten Sie keine Veränderung?
- 1 Eher besser
- 2 Eher schlechter
- 3 Keine Veränderung

**Frage 13**
Und im Bereich Bildung?
[Wird die Situation in den nächsten fünf Jahren eher besser, eher schlechter oder erwarten Sie keine Veränderung?]

- 1 Eher besser
- 2 Eher schlechter
- 3 Keine Veränderung

**Frage 14**

Und im Bereich des Gesundheitssystems?

[Wird die Situation in den nächsten fünf Jahren eher besser, eher schlechter oder erwarten Sie keine Veränderung?]

- - 1 Eher besser
- - 2 Eher schlechter
- - 3 Keine Veränderung

**Frage 15**

Ich nenne Ihnen nun einige Freizeitaktivitäten. Welche der folgenden Aktivitäten üben Sie in Ihrer freien Zeit aus. Sagen Sie mir bitte zu jeder Tätigkeit, wie oft Sie das machen: jede Woche, jeden Monat, seltener oder nie?

Am Abend ausgehen bspw. ins Kino, ins Theater, zu Konzerten oder Essen gehen

[Int.: alle weiteren Abendveranstaltungen fallen auch unter diese Kategorie bspw. mit Freunden etwas trinken gehen, Diskobesuche, Sportveranstaltungen, Vorträge etc.]

- - 1 Jede Woche
- - 2 Jeden Monat
- - 3 Seltener
- - 4 nie

**Frage 16**

Und wie oft treiben Sie aktiv Sport?

- - 1 Jede Woche,
- - 2 Jeden Monat,
- - 3 Seltener,
- - 4 oder nie?

**Frage 17**

Wie oft helfen Sie mit, wenn bei Freunden, Verwandten oder Nachbarn etwas zu tun ist. Machen Sie das …

- - 1 Jede Woche,
- - 2 Jeden Monat,
- - 3 Seltener,
- - 4 oder nie?

If landline → Frage 18 if cell phone Frage 24

**Frage 18**

Nun zu einem ganz anderen Thema. Ich habe noch ein paar Fragen zu Ihrem Telefonanschluss

Ich habe Sie unter der Nummer 0621-xxxx angerufen.

Hätte ich Ihren Haushalt zusätzlich noch über andere private Festnetznummern erreichen können?

Denken Sie bitte auch an Rufnummern, die Sie niemandem mitteilen. Und an Rufnummern, die Sie von Providern erhalten haben für Telefonate über das Internet (VoIP).

[Int.: Homezoneanschlüsse zählen hier nicht]

*[Int. Wozu wird das benötigt? Wir wählen zufällige Nummern an. Deshalb benötigen unsere Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*

- - 1 ja
- - 2 nein

**Frage 19**

Wie viele Festnetznummern haben Sie insgesamt?

Zählen Sie bitte diese Nummer auch dazu.

[Int.: Homezoneanschlüsse zählen hier nicht]

*[Int. Wozu wird das benötigt? Wir wählen zufällige Nummern an. Deshalb benötigen unsere*
*Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*

**Frage 20**

Besitzen Sie ein Handy, das Sie privat nutzen?

[Int.: Geschäftshandys, die teilweise auch privat genutzt werden zählen auch. Reine Diensthandys
nicht.]

*[Int. Wozu wird das benötigt? Wir wählen zufällige Nummern an. Deshalb benötigen unsere*
*Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*

- 1 ja
- 2 nein

**Frage 21**

Sind Sie über mehr als eine Handynummer erreichbar?

- 1 ja
- 2 nein

*[Int.: Ungenutzte und inaktive Sim-Karten zählen Sie bitte nicht dazu.]*

*[Int. Wozu wird das benötigt? Wir wählen zufällige Nummern an. Deshalb benötigen unsere*
*Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*

**Frage 22**

Wie viele Handynummern haben Sie insgesamt?

*[Int.: Ungenutzte und inaktive Sim-Karten zählen Sie bitte nicht dazu.]*

*[Int. Wozu wird das benötigt? Wir wählen zufällige Nummern an. Deshalb benötigen unsere*
*Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*

**Frage 23**

Hat Ihr Handy eine zusätzliche Festnetznummer?

[Int.: Wir meinen die Zusatzoption wie Sie unter den Bezeichnungen o2-Homezone, Vodafone
Zuhause und T-Mobile-at-home angeboten wird]

- 1 ja
- 2 nein

**Frage 24**

Nun zu einem ganz anderen Thema. Ich habe noch ein paar Fragen zu Ihrem Telefon.

Sind Sie über mehr als eine private Handynummer erreichbar?

*[Int.: Ungenutzte und inaktive Sim-Karten zählen Sie bitte nicht dazu.]*

*[Int. Wozu wird das benötigt? Wir wählen zufällige Nummern an. Deshalb benötigen unsere*
*Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*

- 1 ja
- 2 nein

**Frage 25**

Wie viele Handynummern haben Sie insgesamt?

*[Int.: Ungenutzte und inaktive Sim-Karten zählen Sie bitte nicht dazu.]*

*[Int. Wozu wird das benötigt?  Wir wählen zufällige Nummern an. Deshalb benötigen unsere*
*Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*

**Frage 26**

*Hat Ihr Handy eine zusätzliche Festnetznummer?*
*[Int.: Wir meinen die Zusatzoption wie Sie unter den Bezeichnungen o2-Homezone, Vodafone Zuhause und T-Mobile-at-home angeboten wird]*
- *1 ja*
- *2 nein*

**Frage 27**

*Haben Sie zusätzlich zu Ihrem Handy noch einen Festnetzanschluss zu Hause?*
*[Int. Wozu wird das benötigt? Wir wählen zufällige Nummern an. Deshalb benötigen unsere Statistiker die Angabe unter wie vielen Nummern man Sie noch hätte erreichen können.]*
- *1 ja*
- *2 nein*

**Frage 28**

*Ist Ihr Haushalt über mehr als eine private Festnetznummer erreichbar?*
*Denken Sie bitte auch an Rufnummern, die Sie niemandem mitteilen. Und an Rufnummern, die Sie von Providern erhalten haben für Telefonate über das Internet(VoIP).*
- *1 ja*
- *2 nein*

**Frage 29**

*Unter wie vielen privaten Festnetznummern ist Ihr Haushalt insgesamt erreichbar?*

**Frage 30**

Zum Schluss habe ich kurz ein paar wenige Fragen für unsere Statistik.

Geschlecht der/ des Befragten *[Int.: Bitte eintragen]*
- *1 männlich*
- *2 weiblich*

**Frage 30 (regular and short interview)**

Können Sie mir sagen, in welchem Jahr Sie geboren sind?

**Frage 31**

Welchen Familienstand haben Sie?
- 1 verheiratet und leben mit Ihrem/Ihrer Ehepartner/in zusammen
- 2 in eingetragener gleichgeschlechtlicher Lebenspartnerschaft zusammenlebend
- 3 verheiratet und leben von Ihrem/Ihrer Ehepartner/-in getrennt
- 4 ledig
- 5 geschieden
- 6 verwitwet
- 7 in einer eingetragenen Lebenspartnerschaft und leben getrennt
- 8 in einer eingetragenen Lebenspartnerschaft, die aufgehoben wurde
- 9 in einer eingetragenen Lebenspartnerschaft, bei der der/die Partner/-in  verstorben ist

**Frage 32**

Haben Sie einen festen Lebenspartner/ eine feste Lebenspartnerin?
[Int.: Unter festem Lebenspartner wird auch der Partner verstanden, mit dem man nicht zusammen wohnt!]
- 1 ja
- 2 nein

**Frage 33**
Führen Sie mit ihrem Partner/ ihrer Partnerin einen gemeinsamen Haushalt?
-     1 ja
-     2 nein

**Frage 34 (regular and short interview)**
Welchen höchsten allgemein bildenden Schulabschluss haben Sie?
[Int.: nur bei Bedarf vorlesen:] Sind Sie…
-     1 Schüler/in und besuchen eine allgemein bildende Vollzeitschule
-     2 von der Schule abgegangen ohne Hauptschulabschluss (Volksschulabschluss)?
-     3 Haben Sie einen Hauptschulabschluss (Volksschulabschluss)?
-     4 einen Realschulabschluss (Mittlere Reife)?
-     5 einen Abschluss der Polytechnischen Oberschule der DDR, 8. oder 9. Klasse
-     6 einen Abschluss der Polytechnischen Oberschule der DDR, 10. Klasse?
-     7 eine Fachhochschulreife, den Abschluss einer Fachoberschule?
-     8 eine Allgemeine oder fachgebundene Hochschulreife/ Abitur (Gymnasium bzw. EOS, auch EOS mit Lehre)?
-     9 das Abitur über den zweiten Bildungsweg nachgeholt
-     10 einen anderen Schulabschluss? und zwar…

**Frage 35 (regular and short interview)**
Sind Sie zurzeit erwerbstätig?
Sind Sie…
[Int.: Auf Nachfrage: Unter Erwerbstätigkeit wird jede bezahlte bzw. mit einem Einkommen verbundene Tätigkeit verstanden, egal welchen zeitlichen Umfang sie hat.]

-     1 Vollzeit-erwerbstätig
-     2 Teilzeit-erwerbstätig
-     3 in Altersteilzeit (unabhängig davon, ob in der Arbeits- oder Freistellungsphase befindlich)
-     4 geringfügig erwerbstätig, 400-Euro-Job, Minijob
-     5 „Ein-Euro-Jobber" (bei Bezug von Arbeitslosengeld II)
-     6 gelegentlich oder unregelmäßig beschäftigt
-     7 in einer beruflichen Ausbildung/Lehre
-     8 in Umschulung
-     9 im Wehrdienst/Zivildienst
-     10 im Freiwilligen Sozialen Jahr/Freiwilligen ökologischen Jahr, Bundesfreiwilligendienst
-     11 in Mutterschafts-, Erziehungsurlaub, Elternzeit oder sonstiger Beurlaubung (Altersteilzeit unter Code 3 angeben)
-     12 nicht erwerbstätig (einschließlich: Schüler(n)/-innen oder Studierende, die nicht gegen Geld arbeiten, Arbeitslosen, Vorruheständler(n)/-innen, Rentner(n)/-innen ohne Nebenverdienst)

**Frage 36 (regular and short interview)**
Sind Sie zurzeit…

-     1 Schüler(in) an einer allgemeinbildenden Schule
-     2 Student(in)
-     3 Rentner(in)/ Pensionär(in), im Vorruhestand
-     4 arbeitslos
-     5 dauerhaft erwerbsunfähig
-     6 Hausfrau/ Hausmann
-     7 Sonstiges, und zwar:

**Frage 37**

Können Sie mir sagen, wie viele Personen ständig in Ihrem Haushalt leben, Sie selbst mit eingeschlossen? Denken Sie dabei bitte auch an alle im Haushalt lebenden Kinder.
[Int.: ein Haushalt ist eine wirtschaftliche Einheit]

**Frage 38**

Wie viele Personen in Ihrem Haushalt, sie selbst eingeschlossen, sind 18 Jahre oder älter?

**Frage 39 (Experimental split)**

Wie hoch ist das monatliche Nettoeinkommen Ihres Haushalts insgesamt? Ich lese Ihnen jetzt 4 Einkommensgruppen vor. Zu welcher Gruppe zählen Sie?
Gemeint ist die Summe, die nach Abzug der Steuern und Sozialversicherungsbeiträge übrigbleibt.
[Int.: Bei Selbständigen nach dem durchschnittlichen monatlichen Netto-Einkommen, abzüglich der Betriebsausgaben fragen]

- 1 Bis einschließlich 1.500 Euro
- 2 Zwischen 1.500 und 2.500 Euro
- 3 Zwischen 2.500 und 4.000 Euro
- 4 4.000 Euro oder mehr

Bei fehlender Antwort:
- Verweigerung
- Weiß nicht

**Frage 40**

Haben Sie die deutsche Staatsangehörigkeit?
- 1 ja
- 2 nein

**Frage 41**

In welchem Land sind Sie geboren?
- Liste

**Frage 42**

In welchem Land ist Ihre Mutter geboren?
- Liste

**Frage 43**

In welchem Land ist Ihr Vater geboren?
- Liste

**Frage 44 (recruitment, identical for regular and short interview)**

Um weiterhin zu erfahren, was die Menschen über das Leben in Deutschland denken, wollen wir Fragen zu wichtigen Themen auch im Internet stellen.
Und wir möchten Sie einladen an diesen kurzen Internet-Befragungen teilzunehmen.
Diese dauern wenige Minuten (höchstens 10) und Sie können den Fragebogen ausfüllen, wann es Ihnen möglich ist.
Die Teilnahme ist natürlich auch wieder freiwillig.
Wenn Sie Interesse haben und uns unterstützen möchten, dann würde ich Ihre E-Mail-Adresse aufnehmen.

[Int: weitere Infos
- 1x im Monat würden wir Ihnen dann eine Befragungseinladung zuschicken.
- verschiedene Themen: Beruf, Familie, Freizeit, Medien, Politik
- Als Dankeschön erhalten Sie für jede Teilnahme #incentive# Euro.

- Das Geld erhalten Sie dann in Form von Amazon-Gutscheinen. Wir können Ihnen auch, wenn Sie keinen Gutschein möchten, das Geld auf Ihr Konto überweisen. Oder Sie können das Geld auch spenden [Spendenoption, nur bei Bedarf erwähnen.]
  Aber das können Sie noch später entscheiden.

Alle Informationen finden Sie natürlich auch auf unserer Homepage www.deutschlandstudie.de.

- 1 teilnahmebereit, Emailadresse
- 2 Befragte zögert (Homepage)
- 3 hat keine Email-Adresse
- 4 Kein Interesse, will nicht teilnehmen

**Frage 45**
Können Sie mir bitte Ihre E-Mail-Adresse angeben, unter der wir Sie am besten erreichen können?
[INT.: Falls mehrere Email-Adressen vorhanden sind, dann die Adresse notieren, unter der die Befragungsperson am besten erreichbar ist.]

Liste häufigster Provider. Präfix wird eingegeben.
- gmx.de
- gmx.net
- web.de
- t-online.de
- hotmail.de
- hotmail.com
- aol.de
- aol.com
- 1und1.de
- yahoo.de
- yahoo.com
- googlemail.com
- andere

**Frage 46**
Bitte die E-Mail-Adresse nochmals ganz genau vorlesen. Auch auf „.“ „_“, bzw. andere Sonderzeichen achten. Bitte ggf. korrigieren.
Benutzen Sie diese Adresse alleine oder zusammen mit weiteren Personen?
- 1 alleinige E-Mailadresse
- 2 geteilte E-Mailadresse

**Frage 47**
Wenn Sie gerne namentlich angeschrieben werden möchten, können Sie mir gerne Ihren Namen sagen.
[Int.: keine Pflicht, kann Vorname, Nachname oder beides sein]

**Frage 48**
Wie oft rufen Sie diese E-Mailadresse durchschnittlich ab?
Ist das…
- 1 fast ständig
- 2 Täglich
- 3 Mehrmals die Woche
- 4 Mindestens einmal pro Woche
- 5 mehrmals pro Monat
- 6 einmal pro Monat
- 7 Seltener als einmal im Monat

**Frage 49**

Vielen Dank für Ihre Teilnahmebereitschaft. Wir senden Ihnen jetzt gleich eine Willkommens-E-Mail zu, mit dem Betreff „Deutschlandstudie von GESIS".
In einigen Tagen – voraussichtlich am Mittwoch - erhalten Sie dann die Einladung für unsere erste Internetbefragung.

**Frage 50**

Die Themen für die diese Befragungen stehen schon fest.
Ich gebe Ihnen fünf Themen zu Auswahl. Für welches dieser Themen interessieren Sie sich am meisten?

-      Freizeit
-      Beruf und Familie
-      Politik
-      Medien
-      Gesundheit

**Frage 51**

Nochmals vielen Herzlichen Dank für Ihre Teilnahme an unserer Befragung. Einen schönen Tag noch. Auf Wiederhören.

**Frage 52**

Und wie einfach oder schwierig war es, die Person von der Teilnahme am telefonischen Interview zu überzeugen?

-      1 Sehr einfach
-      2 Eher einfach
-      3 Eher schwierig
-      4 Sehr schwierig

**Frage 53**

Wie gut spricht der/ die Befragte Deutsch?
-      1 sehr gut
-      2 gut
-      3 mittelmäßig
-      4 schlecht
-      5 sehr schlecht

**Screenshots from the CATI recruitment part of the interview**

```
Um weiterhin zu erfahren, was die Menschen über das Leben in Deutschland
denken, wollen wir Fragen zu wichtigen Themen auch im Internet stellen
Und wir möchten Sie dazu einladen, an diesen kurzen Internet-Befragungen
teilzunehmen. Diese dauern wenige Minuten (höchstens 10), und Sie können
den Fragebogen ausfüllen, wann es Ihnen möglich ist.
Die Teilnahme ist natürlich auch wieder freiwillig.
Wenn Sie Interesse haben und uns unterstützen möchten, dann würde ich gerne
Ihre E-Mail-Adresse aufnehmen.




                  1 WEITER ...






                <ESC>=Zurück <CTRL+N>=Note Rec:99999 Test Last Name
```

```
weitere Infos zu den Internet-Umfragen:

- Einmal im Monat würden wir Ihnen dann eine Befragungseinladung zuschicken.
- verschiedene Themen: Beruf, Familie, Freizeit
- Als Dankeschön erhalten Sie für jede Teilnahme  0 Euro.
   [INT.: Stundenlohn keinesfalls selber ansprechen!!]
  Und wenn Sie an allen Befragungen teilnehmen, zusätzlich nochmal  0 Euro.
  Das Geld erhalten Sie dann in Form von Amazon-Gutscheinen oder
  per Bank-Überweisung. Oder Sie können das Geld auch spenden.
  Aber das können Sie noch später entscheiden.
- Wir würden Ihnen zunächst eine Willkommens-E-Mail mit dem Betreff
  "Deutschlandstudie von GESIS"
  zuschicken. Darin enthalten sind dann die Einladung zur ersten
  Internet-Befragung und Ihre persönlichen Login-Informationen für
  unsere Website.
- Alle Informationen rund um die Deutschlandstudie von GESIS finden
  Sie natürlich auch auf unserer Homepage www.deutschlandstudie.de.


             1 Möchte daran teilnehmen (Email-Adresse erfassen)
             2 ZP zögert (neuer Termin oder raus)
             3 Hat keine Email-Adresse
             4 Hat kein Interesse (HARTE ABLEHNER!)
             5 unter 18J.
    <F1> Hilfe <ESC>=Zurück <CTRL+N>=Note Rec:99999 Test Last Name
```

Können Sie mir bitte Ihre Email-Adresse angeben, unter der wir Sie
am besten erreichen können?

[INT.: Falls mehrere email-Adressen vorhanden sind, dann die Adresse
        notieren, unter der die Befragungsperson am besten erreichbar ist.]

```
                                @
                        1und1.de
                        aol.com
                        aol.de
                        arcor.de
                        gmx.de
                        gmx.net
                        googlemail.com
                        hotmail.com
                        hotmail.de
                        online.de
                        t-online.de
                        web.de
                        yahoo.com
                        yahoo.de
                        ANDERES
```

Können Sie mir bitte Ihre Email-Adresse angeben, unter der wir Sie
am besten erreichen können?

[INT.: Falls mehrere email-Adressen vorhanden sind, dann die Adresse
        notieren, unter der die Befragungsperson am besten erreichbar ist.]

INT.: Bitte die email-Adresse nochmals GANZ GENAU vorlesen.
      Auch auf ".", "_" bzw. andere Sonderzeichen achten.
      Bei geringen Abweichungen (z.B. Endung ".com" statt ".de")
      des Providers diesen OFFEN eingeben!!


      ines.schaurer@gesis.org




                    1 = Adresse korrekt
                    2 = Adresse NICHT korrekt          Previous Answer: 2

```
Vielen Dank für Ihre Teilnahmebereitschaft.
Wir senden Ihnen dann dieser Tage eine Willkommens-E-Mail zu,
mit dem Betreff "Deutschlandstudie von GESIS".

Die Themen für die nächsten Befragungen stehen schon fest.
Ich gebe Ihnen fünf Themen zu Auswahl.
Für welches dieser Themen interessieren Sie sich am meisten?



        1 Freizeit
        2 Beruf und Familie
        3 Politik
        4 Medien
        5 Gesundheit
        6 EGAL
        7 KEINE ANGABE / WEISS NICHT








        <ESC>=Zurück <CTRL+N>=Note Rec:99999 Test Last Previous Answer: 1
```

```
  Nochmals vielen Dank!




                              Beliebige Taste ...















        <ESC>=Zurück <CTRL+N>=Note Rec:99999 Test Last Name
```

# Appendix Chapter 4

*Table A 1: Bivariate relation between length of interview and online participation (three categories), based on recruited*

|  | No participation [%] | Participation in 1-7 online surveys [%] | Participation in all 8 online surveys [%] | Total |
|---|---|---|---|---|
| Regular interview (10 minutes) | 46.32 (296) | 28.48 (182) | 25.20 (161) | 100 (639) |
| Short interview (3 minutes) | 47.37 (36) | 26.32 (20) | 26.32 (20) | 100 (76) |
| Total | 46.43 (332) | 28.25 (202) | 25.31 (181) | 100 (715) |

Pearson chi² (1) =0.1625 Pr = 0.922

# Appendix Chapter 5

**Question wording of the question about net household income**

Wie hoch ist das monatliche Nettoeinkommen Ihres Haushalts insgesamt? Ich lese Ihnen jetzt 4 Einkommensgruppen vor. Zu welcher Gruppe zählen Sie?

Gemeint ist die Summe, die nach Abzug der Steuern und Sozialversicherungsbeiträge übrigbleibt.

[Int.: Bei Selbständigen nach dem durchschnittlichen monatlichen Netto-Einkommen, abzüglich der Betriebsausgaben fragen]

- 1 Bis einschließlich 1.500 Euro
- 2 Zwischen 1.500 und 2.500 Euro
- 3 Zwischen 2.500 und 4.000 Euro
- 4 4.000 Euro oder mehr

- Verweigerung
- Weiß nicht

*Table A 2:Bivariate relation between income experiment and recruitment by fieldwork agency*

|  | Social research agency | Market research agency |
|---|---|---|
| Income not asked | 58.31% (372/638) | 46.13% (322/698) |
| Income asked | 53.32% (345/647) | 47.03% (317/674) |
| Total | 55.80% (717/1285) | 46.57% (639/1372) |

*Table A 3: Control of sample composition on the telephone interview stage: logistic regression on the experimental variable (1=income asked 0=income not asked)*

|  | Odds ratio |
|---|---|
| Male | 0.948 |
| Age | 0.999 |
| Education *(Ref.: low)* |  |
|     *middle* | 0.785[*] |
|     *high* | 0.874 |
| In paid work | 0.898 |
| Positive survey attitude | 0.925 |
| Mobile | 1.088 |
| Agency | 0.959 |
| N | 2657 |

+ *p* < 0.10, * *p* < 0.05, ** *p* < 0.01

# Appendix Chapter 6

*Table A 4: Effect on incentive on differences in respondent characteristics at the recruitment interview stage (N = 2869; N =  for low household income)*

| | Mean age (t) | % male | % low education | % in paid work | % Positive survey attitude | % mobile | Low household income[1] |
|---|---|---|---|---|---|---|---|
| 0 Euros | 43.48 | 50.07 | 17.09 | 71.95 | 89.96 | 41.68 | 23.19 |
| 2 Euros | 42.77 | 50.72 | 18.80 | 73.18 | 89.75 | 39.53 | 22.94 |
| 5 Euros | 42.78 | 51.04 | 14.87 | 73.59 | 89.89 | 40.75 | 24.80 |
| 5 Euros + bonus | 43.91 | 53.22 | 14.32 | 67.44 | 86.28 | 45.37 | 19.38 |
| 10 Euros + bonus | 44.62 | 51.65 | 14.50 | 70.68 | 85.37 | 48.42 | 23.62 |
| | | | | | | | |
| Δ 0 Euros - 2 Euros | 0.71 | -0.66 | -1.71 | -1.23 | 0.21 | 2.15 | 0.25 |
| Δ 0 Euros - 5 Euros | 0.70 | -0.98 | 2.22 | -1.64 | 0.07 | 0.93 | -1.62 |
| Δ 0 Euros - 5 Euros + bonus | -0.42 | -3.15 | 2.77 | 4.51 | 3.68 | -3.69 | 3.80 |
| Δ 0 Euros - 10 Euros + bonus | -1.14 | -1.58 | 2.58 | 1.28 | 4.59* | -6.74 | -0.43 |
| Δ 2 Euros - 5 Euros | -0.01 | -0.32 | 3.93 | -0.40 | -0.14 | -1.23 | -1.87 |
| Δ 2 Euros - 5 Euros + bonus | -1.14 | -2.50 | 4.48 | 5.75 | 3.47 | -5.84 | 3.55 |
| Δ 2 Euros - 10 Euros + bonus | -1.85+ | -0.92 | 4.30 | 2.51 | 4.38* | -8.89 | -0.68 |
| Δ 5 Euros - 5 Euros + bonus | -1.13 | -2.18 | 0.55 | 6.15 | 3.61 | -4.62 | 5.42 |
| Δ 5 Euros - 10 Euros + bonus | -1.84+ | -0.60 | 0.37 | 2.91 | 4.52* | -7.67 | 1.18 |
| Δ 5 Euros + bonus - 10 Euros + bonus | -0.71 | 1.57 | -0.18 | -3.24 | 0.91 | -3.05 | -4.24 |

*Note.* Comparison of groups with Chi²-test for the categorical variables and t-test for age (applying Rubin's rules)

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

*Table A 5:* Effect on incentive on differences in respondent characteristics at the stage of recruitment (N = 1,436; N = 662 for low household income)

| | low education | in paid work | low hh-income | male | positive survey attitude | mean age | mobile |
|---|---|---|---|---|---|---|---|
| Δ 0 Euros - 2 Euros | -2.66 | -2.72 | 4.26 | 3.45 | -0.75 | -0.26 | 2.58 |
| Δ 0 Euros - 5 Euros | 2.58 | -3.05 | 1.35 | 2.77 | 0.11 | 0.48 | 4.63 |
| Δ 0 Euros - 5 Euros + bonus | 2.63 | 2.14 | 8.52 | 2.33 | 2.25 | -2.19 | -0.95 |
| Δ 0 Euros - 10 Euros + bonus | 1.53 | -2.99 | 7.97 | 2.13 | -2.75 | -1.59 | -3.77 |
| Δ 2 Euros - 5 Euros | 5.23* | -0.34 | -2.90 | -0.68 | 0.86 | 0.74 | 2.05 |
| Δ 2 Euros - 5 Euros + bonus | 5.29+ | 4.86 | 4.27 | -1.12 | 3.00 | -1.94 | -3.53 |
| Δ 2 Euros - 10 Euros + bonus | 4.19 | -0.27 | 3.72 | -1.33 | -2.00 | -1.33 | -6.35 |
| Δ 5 Euros - 5 Euros + bonus | 0.06 | 5.20 | 7.17 | -0.44 | 2.14 | -2.67* | -5.58 |
| Δ 5 Euros - 10 Euros + bonus | -1.05 | 0.06 | 6.62 | -0.64 | -2.85 | -2.07 | -8.4+ |
| Δ 5 Euros + bonus - 10 Euros + bonus | -1.10 | -5.13 | -0.55 | -0.20 | -5.00* | 0.61 | -2.82 |

*Note.* Comparison of groups with test of proportion (z-test) for the categorical variables and t-test for age (applying Rubin's rules)

$+ p < 0.10$, $* p < 0.05$, $** p < 0.01$

*Table A 6: Logistic regression on overall participation*

|  | Odds ratio |
|---|---|
| Incentive |  |
|  *2 Euros* | 1.425** |
|  *5 Euros* | 1.617** |
|  *5 Euros + bonus* | 2.704** |
|  *10 Euros + bonus* | 3.027** |
| Agency *(Ref. internal)* | 0.763 |
| Male | 1.140 |
| Age | 0.998 |
| Education *(Ref. low)* | 1.000 |
|  *middle* | 1.210 |
|  *high* | 2.093** |
| In paid work | 0.904 |
| Positive survey attitude | 2.960** |
| Short interview | 0.798 |
| N | 2869 |

*Note.* Robust standard errors (88 clusters in interviewers)
+ p < 0.10, * p < 0.05, ** p < 0.01

## Comparison between the agencies

*Table A 7: Predicted probability of being recruited by incentive group for the social research agency (n=1300)*

|  | margins | 95 CI |
|---|---|---|
| 0 Euros | 0.4476 | [0.3835, 0.5118] |
| 2 Euros | 0.4835 | [0.4008, 0.5661] |
| 5 Euros | 0.5625 | [0.4710, 0.6540] |
| 5 Euros + bonus | 0.6019 | [0.5014, 0.7023] |
| 10 Euros + bonus | 0.6266 | [0.5575, 0.6957] |

*Table A 8: Predicted probability of being recruited by incentive group for the market research agency (n= 1569)*

|  | margins | 95 CI |
|---|---|---|
| 0 Euros | 0.4140 | [0.3526, 0.4755] |
| 2 Euros | 0.4739 | [0.4124, 0.5353] |
| 5 Euros | 0.4814 | [0.4322, 0.5306] |

*Table A 9: Predicted probability of participation online by incentive group for the social research agency (n=721)*

|  | margins | 95 CI |
|---|---|---|
| 0 Euros | 0.4681 | [0.3738, 0.5623] |
| 2 Euros | 0.6325 | [0.5259, 0.7390] |
| 5 Euros | 0.6325 | [0.5597, 0.7052] |
| 5 Euros + bonus | 0.7487 | [0.6704, 0.8271] |
| 10 Euros + bonus | 0.7626 | [0.6950, 0.8302] |

*Table A 10: Predicted probability of participation online by incentive group for the market research agency (n=715)*

|  | margins | 95 CI |
|---|---|---|
| 0 Euros | 0.5045 | [0.4327, 0.5762] |
| 2 Euros | 0.5388 | [0.4799, 0.5977] |
| 5 Euros | 0.5610 | [0.4862, 0.6357] |

*Table A 11: Effect on incentive on differences in respondent characteristics at the stage of online participation (N = 872; N = 406 for low household income)*

|  | low education | in paid work | low hh-income | male | pos. surv. attitude | mean age | mobile |
|---|---|---|---|---|---|---|---|
| Δ 0 Euros - 2 Euros | -3.21 | -4.18 | 6.02 | 3.17 | -3.24 | -0.50 | 2.87 |
| Δ 0 Euros - 5 Euros | -0.99 | -4.66 | 6.65 | 1.31 | 2.16 | 1.00 | 4.94 |
| Δ 0 Euros - 5 Euros + bonus | -2.21 | 1.71 | 8.32+ | -0.65 | 3.73 | -0.27 | -3.22 |
| Δ 0 Euros - 10 Euros + bonus | -1.69 | -4.03 | 5.06 | 1.16 | -4.45* | 0.05 | -7.66 |
| Δ 2 Euros - 5 Euros | 2.22 | -0.48 | 0.63 | -1.86 | 5.40* | 1.51 | 2.07 |
| Δ 2 Euros - 5 Euros + bonus | 1 | 5.89 | 2.3 | -3.82 | 6.97* | 0.23 | -6.08 |
| Δ 2 Euros - 10 Euros + bonus | 1.52 | 0.14 | -0.96 | -2.01 | -1.21 | 0.55+ | -10.53* |
| Δ 5 Euros - 5 Euros + bonus | -1.22 | 6.38 | 1.67 | -1.96 | 1.57 | -1.28+ | -8.15 |
| Δ 5 Euros - 10 Euros + bonus | -0.7 | 0.63 | -1.59 | -0.15 | -6.61* | -0.96 | -12.6* |
| Δ 5 Euros + bonus - 10 Euros + bonus | 0.52 | -5.75 | -3.26 | 1.81 | -8.18 | 0.32 | -4.44 |

*Note.* Comparison of groups with test of proportion (z-test) for the categorical variables and t-test for age (applying Rubin's rules)
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

# Bibliography

Alcser, K., Clemens, J., Holland, L., Guyer, H., & Hu, M. (2016). Interviewer Recruitment, Selection, and Training. In Survey Rearch Center Michigan (Ed.), *Guidelines for Best Practice in Cross-Cultural Surveys* (pp. 419–468). Retrieved from http://www.ccsg.isr.umich.edu/index.php

American Association for Public Opinion Research. (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th edition.). Lenexa, KS: AAPOR.

Apodaca, R., Lea, S., & Edwards, B. (1998). The Effect of Longitudinal Burden on Survey Participation. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 906–910. Retrieved from Proceedings of the Survey Research Methods Section, ASA (1998)

Baker, R., Blumberg, S., Brick, J. M., Couper, M. P., Courtright, M., Dennis, M., Dillman, D. A., et al. (2010). *AAPOR Report on Online Panels*. Lenexa, Kansas.

Bandilla, W., Kaczmirek, L., Blohm, M., & Neubarth, W. (2009). Coverage- und Nonresponse-Effekte bei Online-Bevölkerungsumfragen. In N. Jackob, H. Schoen, & T. Zerback (Eds.), *Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung*. VS Verlag für Sozialwissenschaften.

Bartsch, S. (2012). *...würden Sie mir dazu Ihre E-Mail-Adresse verraten?" Internetnutzung und Nonresponse beim Aufbau eines Online Access Panels*. Nomos.

Biemer, P. P. (2010a). Total Survey Error Design, Implementation, and Evaluation. *Public Opinion Quarterly*, *74*(5), 817–848.

Biemer, P. P. (2010b). Overview of Design Issues: Total Survey Error. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 27–58). Bingley: Emerald Group Publishing Limited.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. New York: John Wiley & Sons.

Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A.-S., Das, M., Douhou, S., & Krieger, U. (2016). A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe. *Social Science Computer Review*, *34*(1), 8–25. doi:10.1177/0894439315574825

Blom, A. G., de Leeuw, E. D., & Hox, J. J. (2011). Interviewer Effects on Nonresponse in the European Social Survey. *Journal of Official Statistics*, *27*(2), 359–377.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting Up an Online Panel Representative of the General Population: The German Internet Panel. *Field Methods*, *27*(4).

Blom, A. G., Herzing, J. M. E., Cornesse, C., Sakshaug, J. W., Krieger, U., & Bossert, D. (2016). Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-Based Online Panels? Evidence From the German Internet Panel. *Social Science Computer Review*. doi:10.1177/0894439316651584

Blom, A. G., & Korbmacher, J. (2013). Measuring Interviewer Characteristics Pertinent to Social Surveys: A Conceptual Framework. *Survey Methods: Insights from the Field*, *1*. Retrieved from http://surveyinsights.org/?p=817

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 651–675.

Bogen, K. (1996). The Effect of Questionnaire Length on Response Rates – A Review of the Literature.

Börsch-Supan, A., Krieger, U., & Schröder, M. (2013). *Respondent incentives, Interviewer Traning and Survey Participation*. *Share Working Paper Series* (Vol. 12). Munich.

Bosnjak, M. (2002). *(Non)Response bei Web-Befragungen*. Aachen: Shaker.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2017). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review*. doi:10.1177/0894439317697949

Bosnjak, M., Das, M., & Lynn, P. (2016). Methods for Probability-Based Online and Mixed-Mode Panels: Selected Recent Trends and Future Perspectives. *Social Science Computer Review*, *34*(1), 3–7. doi:10.1177/0894439315579246

Bosnjak, M., & Tuten, T. L. (2003). Prepaid and Promised Incentives in Web Surveys. *Social Science Computer Review*, *21*(2), 208–217.

Bradburn, N. M. (1978). Respondent Burden. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 35–40.

Branden, L., Gritz, R. M., & Pergamit, M. (1995). The Effect of Interview Length on Attrition in the National Longitudinal Survey of Youth. Washington, DC: Bureau of Labor Statistics, US Department of Labor. Retrieved from http://www.bls.gov/osmr/nls_catalog.htm

Broome, J. (2015). First Impressions of Telephone Survey Interviewers. *Journal of Official Statistics*, *31*(4), 611–625.

Callegaro, M., Baker, R., Bethlehem, J. G., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Online panel research. History, concepts, aapplication and a look at the future. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research. A Data Quality Perspective* (pp. 1–22). New York: Wiley.

Callegaro, M., & DiSogra, C. (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly*, *72*(5), 1008–1032.

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). *Web Survey Methodology*. Los Angeles: Sage.

Callegaro, M., Villar, A., Yeager, D., & Krosnick, J. A. (2014). A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research. A Data Quality Perspective* (pp. 61–85). New York: Wiley.

Campanelli, P., & O'Muircheartaigh, C. (1999). Inteviewers, interviewer continuity, and panel survey nonresponse. *Quality and Quantity*, *33*(1), 59–76.

Campanelli, P., Sturgis, P., & Moon, N. (1996). Exploring the impact of survey introduction. *American Statistical Association Proceedings: Section on survey research methods* (pp. 1054–1059). Alexandria: American Statistical Association.

Cantor, D., O'Hare, B. C., & O'Connor, K. (2008). The use of monetary incentives to reduce nonresponse in random digit dial telephone surveys. In J. M. Lepkowski, C. Tucker, J. M.

Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, et al. (Eds.), *Advances in Telephone Survey Methodology* (pp. 471–498). Hoboken: Wiley.

Chang, L., & Krosnick, J. A. (2009). National surveys via rdd telephone interviewing versus the internet. Comparing sample representativeness and response quality. *Public Opinion Quarterly*, *73*(4), 641–678.

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly*, *57*(1), 62–79. Retrieved from http://poq.oxfordjournals.org

Cialdini, R. B. (1988). *Influence: Science and Practice*. Glenview, IL: Scott, Foresman.

Collins, M., Sykes, W., Wilson, P., & Blackshaw, N. (1988). Nonresponse: The UK Experience. In R. M. Groves (Ed.), *Telephone Survey Methodology* (pp. 213–231). Wiley and Sons.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in Web- or Internet-based surveys. *Educational and Psychological Measurement*, *60*, 821–836.

Cornilleau, A., Cousteaux, A.-S., Legleye, S., & Razakamanana, N. (2015). *Bilan du pilot de l'etude longitudinal par internet pour les sciences sososial*. Paris: Sciences Po.

Couper, M. P. (1997). Survey introduction and data quality. *Public Opin Quarterly*, *61*(2), 317–338. Retrieved from http://www.jstor.org/stable/pdf/2749554.pdf

Couper, M. P. (2000). Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*, *64*(4), 464–494.

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and Nonresponse in an Internet Survey. *Social Science Research*, *36*(1), 131–148. doi:10.1016/j.ssresearch.2005.10.002

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web Surveys: Perceptions of Burden. *Social Science Computer Review*, *19*(2), 146–162.

Curtin, R., Singer, E., & Presser, S. (2007). Incentives in Random Digit Dial Telephone Surveys: A Replication and Extension. *Journal of Official Statistics*, *23*(1), 91–105.

Das, M. (2012). Innovation in Online Data Collection for Scientific Research: the Dutch MESS project. *Methodological Innovations Online*, *7*(1), 7–24.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, *25*(1), 14–26.

de Leeuw, E. D. (1999). How Do Successful and Less Successful Interviewers Differ in Tactics for Combating Survey Nonresponse? *Bulletin de Méthodologie Sociologique*, *62*(1), 29–42. doi:10.1177/075910639906200104

de Leeuw, E. D. (2009). *International Interviewer Questionnaire (IQUEST): Development and Scale Properties*. Department of Methodology and Statistics, University Utrecht.

de Leeuw, E. D., & Hox, J. J. (2004). I am Not Selling Anything: 29 Experiments in Telephone Introductions. *International Journal of Public Opinion Research*, *16*(4), 464–473. doi:10.1093/ijpor/edh040

Dennis, J. M. (2017). *Technical overview of the Amerispeak Panel*. Retrieved from http://amerispeak.norc.org

Dillman, D. A. (1978). *Mail and Telephone Surveys*. John Wiley & Sons.

Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little, R. J. A. (2002). Survey Nonresponse in Design, Data Collection, and Analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 3–26). New York: Wiley.

Dillman, D. A., Reips, U., & Matzat, U. (2010). Advice in surveying the general population over the Internet (Editorial). *International Journal of Internet Science*, *5*, 1–4.

Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly*, *57*(3), 289–304. doi:10.1086/269376

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys. The Tailored Design Method*. Hoboken, New Jersey: Wiley.

DiSogra, C., & Callegaro, M. (2009). Computing Response Rates for Probability-Based Web Panels. *Proceedings of the Section on Survey Research Methods – JSM 2009*.

DiSogra, C., & Callegaro, M. (2016). Metrics and Design Tool for Building and Evaluating Probability-Based Online Panels. *Social Science Computer Review*, *34*(1), 26–40. doi:10.1177/0894439315573925

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys. *Public Opinion Quarterly*, *74*(1), 1–36.

Eckman, S. (2016). Does the Inclusion of Non-Internet Households in a Web Panel Reduce Coverage Bias? *Social Science Computer Review*, *34*, 41–58.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *British Medical Journal*, *324*(7347), 1183.

Engel, U., Bartsch, S., Schnabel, C., & Vehre, H. (2012). *Wissenschaftliche Umfragen*. Frankfurt/ New York: Campus Verlag.

Esser, H. (1986). Über die Teilnahme bei Befragungen. *ZUMA Nachrichten*, *18*, 38–47.

Fricker, S., Creech, B., Davis, J., Gonzalez, J., Tan, L., & To, N. (2012). *Does Length Really Matter? Exploring the Effects of a Shorter Interview on Data Quality, Nonresponse, and Respondent Burden*. Paper presented at the Federal Committee on Statistical Methodology 2012 Research Conference, Washington, DC, USA. Retrieved from http://fcsm.sites.usa.gov/files/2014/05/Fricker_2012FCSM_IX-B.pdf

Gabler, S., & Häder, S. (2002). *Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland*. Münster: Waxmann Verlag.

Gabler, S., Häder, S., Lehnhoff, I., & Mardian, E. (2012). Weighting for the unequal inclusion probabilities and nonresponse in dual frame telephone surveys. In S. Häder, M. Häder, & M. Kühne (Eds.), *Telephone surveys in Europe* (pp. 147–167). Heidelberg: Springer.

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, *73*(2), 349–360. doi:10.1093/poq/nfp031

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2016). Interviewing Practices, Conversational Practices, and Rapport. *Sociological Methodology*, *46*(1), 1–38. doi:10.1177/0081175016637890

Göritz, A. S. (2004a). Recruitment for Online Access Panels. *International Journal of Market Research*, *46*(4), 411–425.

Göritz, A. S. (2004b). The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels. *International Journal of Market Research*, *46*, 327–345.

Göritz, A. S. (2006). Incentives in Web Studies: Methodological Issues and a Review. *International Journal of Internet Science*, *1*(1), 58–70.

Göritz, A. S., & Neumann, B. P. (2016). The longitudinal effects of incentives on response quantity in online panels. *Translational Issues in Psychological Science*, *2*(2), 163–173.

Goudy, W. J., & Potter, H. R. (1975). Interview Rapport: Demise of a concept. *Public Opinion Quarterly*, *39*(4), 529. doi:10.1086/268250

Goyder, J. (1987). *The Silent Minority. Nonrespondents on Sample Surveys*. Camebridge: Polity Press.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R. M. (2004). *Survey Methodology*. Hoboken, NJ: Wiley.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the Decision to participate in a Survey. *Public Opinion Quarterly*, *56*(4), 475–495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of Royal Statistical Society: Series A (Statistics in Society)*, *169*(3), 439–457.

Groves, R. M., & Lyberg, L. E. (2001). An Overview of Nonresponse Issues in Telephone Surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, & J. Waksberg (Eds.), *Telephone Survey Methodology*. New York: Wiley and Sons.

Groves, R. M., & Lyberg, L. E. (2010). Total Survey Error. Past, Present, and Future. *Public Opinion Quarterly*, *54*(5), 849–879.

Groves, R. M., & Magilavy, L. J. (1986). Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys. *Public Opinion Quarterly*, *50*(2), 2251–266.

Groves, R. M., & McGonagle, K. A. (2001). A Theory-Guided Interviewer Training Protocol Regarding Survey Participation. *Journal of Official Statistics*, *17*(2), 249–265.

Groves, R. M., O'Hare, B. C., Gould-Smith, D., Benki, J., & Maher, P. (2008). Telephone interviwers voice characteristics and the survey participation decision. In J. M. Lepkowski, C. Tucker, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 385–400). New York: Wiley.

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias. A Meta-Analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Groves, R. M., Presser, S., & Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, *68*(1), 2–31. doi:10.1093/poq/nfh002

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-Saliency Theory of Survey Participation. *Public Opinion Quarterly*, *64*, 299–308.

Hansen, K. M. (2007). The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study. *International Journal of Public Opinion Research*, *19*(1), 112–121. doi:10.1093/ijpor/edl022

Hansen, K. M., & Pedersen, R. T. (2011). Efficiency of Different Recruitment Strategies for Web Panels. *International Journal of Pubilic Opinion Research*, *24*(2), 238–249. doi:10.1093/ijpor/edr020

Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behavior Research Methods*, *47*(3), 685–690. doi:10.3758/s13428-015-0617-9

Heberlein, T. A., & Baumgartner, R. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. *American Sociological Review*, *43*(4), 447–462. Retrieved from http://www.jstor.org/stable/2094771

Heeringa, S., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. CRC Press.

Hill, D. H., & Willis, R. J. (2001). Reducing Panel Attrition: A Search for Effective Policy Instruments. *Journal of Human Resources*, *36*(3), 416–438.

Hogestol, A., & Skjervheim, O. (2013). *Norwegian citizen panel. 2013, First wave*.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent

Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly*, *67*(1), 79–125. doi:10.1086/346010

Hoogendoorn, A. W., & Daalmans, J. (2009). Nonresponse in the Recruitment of an Internet Panel Based on Probability Sampling. *Survey Research Methods*, *3*(2), 59–72.

Hox, J. J. (2010). *Multilevel Analysis*. New York and Hove: Routledge.

Hox, J. J., & de Leeuw, E. D. (2002). The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An International Comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 103–120). New York: Wiley.

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2013). The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-operation with Face-to-Face Surveys. *Survey Research Methods*, *7*(1), 1–15.

Japec, L. (2008). Interviewer Error and Interviewer Burden. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, et al. (Eds.), *Advances in telephone survey methodology* (pp. 185–211). Hoboken, NJ: Wiley-Interscience.

Jónsdóttir, G. A. (2015). Sample Design for Internet Panel Surveys. Retrieved from http://www.lidata.eu/files/naujienos/2015-04-09/internet_panel.pdf

Juster, F. T., & Smith, J. P. (1997). Improving the Quality of Economic Data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association*, *92*(440), 1268–1278. doi:10.1080/01621459.1997.10473648

Kish, L. (1962). Studies of Interviewer Variance for Attitudinal Variables. *Journal of the American Statistical Association*, *57*, 92–115.

Knoef, M., & Vos, K. de. (2008). The representativeness of the LISS panel.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys. The Effect of Mode and Question Sensitivity. *Public Opin Quarterly*, *72*(5), 847–865.

Krieger, U. (2016). Using Cash Incentives to help Recruitment in a probability based Web Panel. Presented at the 71th Annual Conference of the American Association for Public Opinion Research (AAPOR). Austin, TX, USA.

Laurie, H., & Lynn, P. (2009). The Use of Respondent Incentives on Longitudinal Surveys. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys*. West Sussex: Wiley.

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for Reducing Nonresponse in a Longitudinal Panel Survey. *Journal of Official Statistics*, *15*(2), 269–282.

Lee, S. (2006). An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey. *Social Science Computer Review*, *24*(4), 460–475.

Leenheer, J., & Scherpenzeel, A. C. (2013). Does it pay off to include non-Internet households in an Internet panel? *International Journal of Internet Science*, *8*(1), 17–28.

Lensvelt-Mulders, G. J. L. M., Lugtig, P. J., & Hox, J. J. (2008). Assembling an access panel: a study of initial nonresponse and self-selection bias. In I. Stoop & M. Wittenberg (Eds.), *Access panels and online research, panacea or pitfall? Proceedings of the DANS symposium* (pp. 79–86). Amsterdam: aksant.

Lepkowski, J. M., & Couper, M. P. (2001). Nonresponse in the Second Wave of Longitudinal Household Surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (1st ed., pp. 259–272). New York: John Wiley & Sons.

Link, M. W. (2006). Predicting the Persistence and Performance of Newly Recruited Telephone Interviewers. *Field Methods*, *18*, 305–320.

Lipps, O. (2008). A Note on Interviewer Performance Measures in Centralised CATI Surveys. *Survey Research Methods*, *2*(2), 61–73.

Lipps, O. (2009). Cooperation in Centralised CATI Household Panel Surveys - A Contact-based Multilevel Analysis to Examine Interviewer, Respondent, and Fieldwork Process Effects. *Journal of Official Statistics*, *25*(3), 323–338.

Lipps, O. (2010a). Effects of different incentives on attrition and fieldwork effort in telephone household panel surveys. *Survey Research Methods*, *4*(2), 81–90.

Lipps, O. (2010b). Interviewer-Respondent Socio-Demographic Matching and Survey Cooperation. *Survey Practice*, *3*(4). Retrieved from

http://surveypractice.org/index.php/SurveyPractice/article/view/140/html

Lipps, O., & Pollien, A. (2011). Effects of Interviewer Experience on Components of Nonresponse in the European Social Survey. *Field Methods*, *23*(2), 156–172. doi:10.1177/1525822X10387770

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd edition.). Hoboken, New Jersey: John Wiley & Sons.

Long, J. S., & Freese, J. (2014). *Regression Models for Categorial Dependent Variables Using Stata* (Third Edition.). College Station, Texas: STATA Press.

Loosveldt, G., Pickery, J., & Billiet, J. (2002). Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics*, *18*(4), 545–557.

Lynn, P. (2001). The impact of incentives on response rates to personal interview surveys: role and perceptions of interviewers. *International Journal of Public Opinion Research*, *13*(3), 326–336.

Lynn, P. (2008). The problem of nonresponse. In E. D. de Leeuw, J. Hox, & D. A. Dillmann (Eds.), *International Handbook of Survey Methodology* (pp. 35–55). Taylor & Francis Group.

Lynn, P. (2009). Methods for longitudinal surveys. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 1–19). West Sussex: Wiley.

Lynn, P. (2013). What can we do to encourage participation in web surveys? Retrieved from http://www.natcen.ac.uk/genpopweb/blog.htm#sthash.YOtePW3P.dpuf

Lynn, P. (2014). Longer Interviews May Not Affect Subsequent Survey Participation Propensity. *Public Opinion Quarterly*, *78*(2), 500–509. doi:10.1093/poq/nfu015

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schütz, A. (2007). Compensating for Low Topic Interest and Long Surveys. A Field Experiment on Nonresponse in Web Surveys. *Social Science Computer Review*, *25*(3), 372–383.

Martinsson, J., & Riedel, K. (2015). *Postal recruitment to a probability based web panel. Long term consequences for response rates, representativeness and costs*. *LORE Methodological Notes* (Vol. 1).

Mason, R., Lesser, V. M., & Traugott, M. (2002). Effect of item nonresponse on nonresponse error and inference. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 149–161). New York, NY: John Wiley & Sons.

Meier Jæger, M. (2016). Hello Beautiful? The Effect of Interviewer Physical Attractiveness on Cooperation Rates and Survey Responses. *Sociological Methods & Research*, *Published online first in October 2016*. doi:10.1177/0049124116672677

Mercer, A., Caporaso, A., Cantor, D., & Townsend, R. (2015). How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly*, *79*(1), 105–129.

Moore, D. L., & Tarnai, J. (2008). Evaluation Nonresponse Error in Mail Surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 197–211). New York: John Wiley & Sons.

Moore, J. C., Stinson, L. L., & Welniak, E. J. (1999). Income Reporting in Surveys: Cognitive Issues and Measurement Error. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and Survey Research* (pp. 155–174). New York: Wiley-Interscience.

Morton-Williams, J. (1993). *Interviewer Approaches*. Camebridge: Cambridge University Press.

Müller, B., & Castiglioni, L. (2015). Attrition im Beziehungs- und Familienpanel pairfam. In J. Schupp & C. Wolf (Eds.), *Nonresponse Bias* (pp. 383–408). Wiesbaden: Springer Fachmedien.

Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton and Oxford: Princeton University Press.

NatCen. (2017). Technical information NatCen Panel. Retrieved from http://www.natcen.ac.uk/our-expertise/methods-expertise/surveys/probability-panel/

Nock, S. L., & Guterbock, T. M. (2010). Survey experiments. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 838–864). Bingley: Emerald Group Publishing Limited.

Oksenberg, L., & Cannell, C. (2001). Effects of Interviewer Vocal Characteristics on Nonresponse. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, & J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 257–272). New York: Wiley and Sons.

Olsen, R. J. (2005). The Problem of Respondent Attrition: Survey Methodology is Key. *Monthly Labour Review*, *128*(2), 63–70.

Olson, K. (2006). Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias. *Public Opinion Quarterly*, *70*(5), 737–758.

Pew Research Center. (2017). American Trends Panel. Retrieved from http://www.pewresearch.org/methodology/u-s-survey-research/american-trends-panel/

Pforr, K., Blohm, M., Blom, A. G., Erdel, B., Felderer, B., Fräßdorf, M., Hajek, K., et al. (2015). Are incentive effects on response rates and nonresponse bias in large-scale, face-to-face surveys generalizable to Germany? Evidence from ten experiments. *Public Opinion Quarterly*, *online first*.

Pickery, J., & Loosveldt, G. (2002). A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse. *Quality and Quantity*, *36*(4), 427–437.

Pickery, J., Loosveldt, G., & Carton, A. (2001). The Effects of Interviwer and Respondent Characteristics on Response Behavior in Panel Surveys. A Multilevel Approach. *Sociological Methods & Researach*, *29*(4), 509–523.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata* (3rd ed.). STATA Press.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology*, *27*(1), 85–95.

RAND. (2017). American Life Panel. Retrieved from https://alpdata.rand.org

Rao, K., Kaminska, O., & McCutcheon, A. L. (2010). Recruting Probability Samples for a Multi-Mode Research Panel with Internet and Mail Components. *Public Opinion Quarterly*, *74*(1), 68–84.

Revilla, M., Cornilleau, A., Cousteaux, A.-S., Legleye, S., & Pedraza, P. de. (2015). What Is the Gain in a Probability-Based Online Panel of Providing Internet Access to Sampling Units Who Previously Had No Access? *Social Science Computer Review*, *34*(4), 479–496. doi:10.1177/0894439315590206

Revilla, M., & Saris, W. E. (2010). *A comparison of different surveys using different modes of data colletion: European Social Survey versus Liss Panel*. Universitat Pompeu Fabra - Research and Expertise Centre for Survey Methodology.

Roberts, C., Eva, G., Allum, N., & Lynn, P. (2010). Data quality in telephone surveys and the effect of questionnaire length: A cross-national experiment. *ISER Working Paper Series, No. 2010-36*. Retrieved from http://hdl.handle.net/10419/65960

Rookey, B. D., Hanway, S., & Dillman, D. A. (2008). Does a Probability-Based Household Panel benefit from Assignment to Postal Response as an Alternative to Internet-only? *Public Opinion Quarterly*, *72*(5), 962–984.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, New Jearsey: John Wiley & Sons.

Ryu, E., Couper, M. P., & Marans, R. W. (2006). Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonrespnse Error. *International Journal of Pubilic Opinion Research*, *18*(1), 89–106.

Sakshaug, J., Couper, M. P., Ofstedal, M. B., & Weir, D. R. (2012). Linking Survey and Administrative Records : Mechanism of Consent. *Sociological Methods & Research*, *41*(4), 535–569.

Sakshaug, J., & Kreuter, F. (2011). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. Swiss Foundation for Research in Social Sciences.

Sala, E., Burton, J., & Knies, G. (2012). Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview, and Interviewer Characteristics. *Sociological Methods & Research*, *41*(3), 414–439. doi:10.1177/0049124112457330

Saris, W. E. (1998). Ten years of interviewing without interviews: The telepanel. In M. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J.Martin, W.L.Nicholls, & J. O'Reilly (Eds.), *Computer assisted survey information collection* (pp. 409–429). New York: Wiley.

Schaeffer, N. C., & Dykema, J. (2011). Questions for Surveys: Current Trends and Future Directions. *Public Opinion Quarterly*, *75*(5), 909–961.

Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and Interviewing. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 437–470). Bingley: Emerald Group Publishing Limited.

Schafer, J. L., & Olsen, M. K. (1998). Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, *33*(4), 545–571.

Schaurer, I., & Bosnjak, M. (2016). *Does Changing Monetary Incentive Schemes in Panel Studies Affect Cooperation? A Quasi-experiment on Switching from Prepaid to Promised Incentives on Participation in a Probability-based Mixed-mode Panel*. Presented at the 71th Annual Conference of the American Association for Public Opinion Research (AAPOR). Austin, TX, USA.

Schaurer, I., Weyandt, K. W., Bauer, R., Blohm, M., & Weiß, B. (2017). *Vergleich zweier Rekrutierungsvarianten für das GESIS Panel*. Presented at the 11. Workshop der Panelsurveys in Deutschland. Nürnberg, Germany.

Scherpenzeel, A. C., Bergmann, M., Friedel, S., & Bristle, J. (2016). Can we use the relationship between income item nonresponse and panel attrition in an adaptive fieldwork design? A study in the Survey of Health, Ageing and Retirement in Europe (SHARE).

Scherpenzeel, A. C., & Bethlehem, J. G. (2010). How Representative Are Online Panels? Problems of Coverage and Selection and Possible Solutions. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies* (pp. 105–132). Oxford: Taylor& Francis Group.

Scherpenzeel, A. C., Das, M., Kapteyn, A., & Soest, A. van. (2008). *Start and Status of the LISS Panel: Design, Recruitment and first Year fo Data Collection*.

Scherpenzeel, A. C., & Toepoel, V. (2012). Recruting a Probability Sample for an online panel. *Public Opinion Quarterly*, *76*(3), 470–790.

Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen*. Opladen: Leske + Budrich.

Schnell, R., & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics*, *21*(3), 389.

Schonlau, M., Soest, A. van, & Kapteyn, A. (2007). Are "Webographic" or attitudinal questions useful for adjusting estimates from Web surveys using prpensity scoring? *Survey Research Methods*, *1*(3), 155–163.

Schräpler, J.-P. (2004). Respondent Behavior in Panel Studies - A Case Study for Income-Nonresponse by Means of the German Socio-Economic Panel (SOEP). *Sociological Methods & Research*, *33*(1), 118–156.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Boston: Houghton Mifflin Company.

Sharp, L. M., & Frankel, J. (1983). Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly*, *47*(1), 36–53. doi:10.1086/268765

Sheehan, K. B. (2001). E-mail Survey Response Rates: A Review. *Journal of Computer-Mediated Communication*, *6*(2), 0–0. doi:10.1111/j.1083-6101.2001.tb00117.x

Singer, E. (2011). Toward a Benefit-Cost Theory of Survey Participation: Evidence, Further Tests, and Implications. *Journal of Official Statistics*, *27*(2), 379–392.

Singer, E., Frankel, M. R., & Glassman, M. B. (1983). The Effect of Interviewer Characteristics and Expectations on Response. *Public Opinion Quarterly*, *47*(1), 68–83. doi:10.1086/268767

Singer, E., Hoewyk, J. van, & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, *64*, 171–188.

Singer, E., & Kulka, R. A. (2002). Paying Respondents for Survey Participation. In M. V. Ploeg, R. A. Moffitt, & C. F. Citro (Eds.), *Studies of Welfare Populations: Data Collection and Research Issues* (pp. 105–128). Washington: Committee on National Statistics Division of Behavioral and Social Sciences and Education National Research Council.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T. E., & McGonagle, K. A. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Official Statistics*, *15*(2), 217–230.

Singer, E., Von Thurn, D. R., & Miller, E. R. (1995). Confidentiality assurance and response: a quantitative review of the experimental literature. *Public Opinion Quarterly*, *59*(1), 66–77. doi:10.1086/269458

Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 112–141.

Snijkers, G., Hox, J. J., & de Leeuw, E. D. (1999). Interviewers' Tactics for Fighting Survey Nonresponse. *Journal of official statistics*, *15*(2), 185–198.

Steinkopf, L., Bauer, G., & Best, H. (2010). Nonresponse und Interviewer-Erfolg im Telefoninterview: Empirische Untersuchungen zum Einfluss stimmlicher Eigenschaften der Interviewer. *Methoden - Daten - Analysen*, *4*(1), 3–26.

Stiegler, A., & Biedinger, N. (2016). *Interviewer Skills and Training*. GESIS Survey Guidelines. Mannheim: GESIS - Leibniz Institute for the Social Sciences.

Stinchcombe, A. L., Jones, C., & Sheatsley, P. (1981). Nonresponse Bias for Attitude Questions. *The Public Opinion Quarterly*, *45*(3), 359–375.

Struminskaya, B. (2014). *Data quality in probability-based online panels*. Retrieved from https://dspace.library.uu.nl/bitstream/1874/301751/3/struminskaya.pdf

Struminskaya, B. (2016). Respondent conditioning in online panel surveys: Results from two field experiments. *Social Science Computer Review*, *34*(1), 95–115.

Struminskaya, B., Kaczmirek, L., Schaurer, I., & Bandilla, W. (2014). Assessing representativeness of a probability-based online panel in Germany. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research. A Data Quality Perspective* (pp. 61–85). New York: Wiley.

Tourangeau, R. (2004). Experimental Design Consideration for Testing and Evaluating Questionnaires. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. L. E. Martin, J. Martin, & E.

Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 209–224). Wiley-Interscience.

Tourangeau, R., Conrad, F. G., & Couper, M. P. and. (2013). *The Science of Web Surveys*. Cary, US: Oxford University Press.

Tourangeau, R., Groves, R. M., Kennedy, C., & Yan, T. (2009). The Presentation of a Web Survey, Nonresponse and Measurement Error among Members of Web Panel. *Journal of Official Statistics*, *25*(3), 299–321.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive Topics and Reluctant Respondents: Demonstrating a Link between Nonresponse Bias and Measurement Error. *Public Opinion Quarterly*, *74*(3), 413–432.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, England: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, *133*(5), 859–883.

Uhrig, S. C. N. (2008). *The nature and cause of attrition in the British Household Panel Survey*. *ISER Working Paper* (Vol. 2008–05).

Understanding America Study. (2017). Sample and Recruitment. Retrieved from https://uasdata.usc.edu/recruitmentoverview/1

Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Boca Raton: CRC Press.

van der Laan, J. (2009). *Representativity of the LISS panel The Hague/Heerlen, 2009* (Vol. 09041). Statistics Netherlands.

Vassallo, R., Durrant, G. B., & Smith, P. W. F. (2015). Interviewer effects on non-response propensity in longitudinal surveys: a multilevel modelling approach. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *178*(1), 83–99.

Vehovar, V., Lozar Manfreda, K., Zaletel, M., & Batagelj, Z. (2002). Nonresponse in Web Surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 229–242). New York: Wiley.

Vehre, H. (2011). *Sie wollen mir doch was verkaufen!". Analyse der Umfrageteilnahme in einem offline rekrutierten Access Panel*. Wiesbaden: VS Verlag der Sozialwissenschaften.

Villar, A., & Sommer, E. (2017). *Web recruitment design plans and experimental testing*. Report on work package 7 of SERISS. London: City University.

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

Watson, N., & Wooden, M. (2009). Identifying Factors Affecting Longitudinal Survey Response. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 157–179). John Wiley & Sons.

Wengrzik, J., Silber, H., Bosnjak, M., Zabal, A., & Martin, S. (2017). *A General Interviewer Training Curriculum for Computer-Assisted Personal Interviews (GIT-CAPI)*. GESIS Survey Guidelines. Mannheim: GESIS - Leibniz Institute for the Social Sciences.

West, B. T., & Blom, A. G. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *online first*, 1–37.

West, B. T., & Elliot, M. R. (2014). Frequentist and Bayesian approaches for comparing interviewer variance components in two groups of survey interviewers. *Survey Methodology*, *40*(2), 163–188.

West, B. T., & Groves, R. M. (2013). A Propensity-Adjusted Interviewer Performance Indicator. *Public Opinion Quarterly*, *77*(1), 352–374.

West, B. T., & Olson, K. (2010). How Much of Interviewer Variance is Really Nonresponse Error Variance? *Public Opinion Quarterly*, *75*(5), 1004–1026.

Yammarion, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding Mail Survey Response Behavior. *Public Opinion Quarterly*, *55*(4), 613–639. doi:10.1086/269284

Yan, T., & Curtin, R. (2010). The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective. *International Journal of Pubilic Opinion Research*, *22*(4), 535–551.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, *75*(4), 709–747.

Zabel, J. E. (1998). An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior. *Journal of Human Resources*, *33*(2), 479–506.