The institutionalized population and social surveys

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Sozialwissenschaften der Universität Mannheim

vorgelegt von
Jan-Lucas Schanze

Universität Mannheim
Fakultät für Sozialwissenschaften

Dekan der Fakultät für Sozialwissenschaften: Prof. Dr. Michael Diehl

Erstbetreuer:	Prof. Dr. Christof Wolf
Zweitbetreuerin:	Prof. Annelies Blom, Ph.D.

Gutachter:	Prof. Dr. Christof Wolf (Universität Mannheim)
	Prof. Annelies Blom, Ph.D. (Universität Mannheim)
	Prof. Dr. Marek Fuchs (Technische Universität Darmstadt)

Datum der Disputation: 16. August 2021

Acknowledgements

This doctoral thesis is the sum of five years of productive work. It also reflects five years of ties and working alliances with various people supporting me and my research. I cannot overestimate the positive impact those people have had on me and on this doctoral thesis. It is my utmost pleasure to acknowledge their contribution here with deep gratitude.

First of all, I would like to express my gratitude to Christof Wolf and Annelies Blom, my PhD supervisors nearly from the beginning on. When looking at factors contributing to a successful collaboration of doctoral researchers and their PhD supervisors as reported by the Leibniz PhD Network (Beadle et al. 2020), Christof has done a great job: Our monthly meetings increased my motivation to constantly work on the tasks ahead. You reliably provided me with the right balance of substantive and methodological hints, as well as smaller linguistic corrections of written text. I enjoyed our conversations and always left meetings with you freshly encouraged. Annelies had a similarly positive impact on my PhD. Your advice comforted me after I had to swallow my first rejection by a journal. You have regularly provided me with valuable feedback, no matter whether it was about developing a research plan or improving my presentation skills. This is the best opportunity to say thank you, Annelies and Christof. I would very much appreciate to continue our scientific collaboration in the future.

I am very lucky to be surrounded by knowledgeable, cordial, and cooperative colleagues at GESIS, in the European Social Survey, and earlier in the SERISS project. I would like to express my gratitude to my close colleagues, Achim Koch and Roberto Briceno-Rosas, who both read and discussed my papers and ideas for further research various times. Stefan Zins was an excellent partner for working on a paper and extending my knowledge about sampling and simulations. Ineke Stoop always showed her interest in the topic of this doctoral thesis and shared her feedback and advice on one chapter. She is the very first person I have in mind when emphasizing the positive impact networks and a few key persons have had on my progress so far. Annette Scherpenzeel, Michael Bergmann, and Sarah Butt allowed me to grow while I worked independently on my tasks in the SERISS project. Annette and Michael also supported me with my analyses of SHARE data. Any errors left in this doctoral thesis are mine and should not have any effect on the reputation of those esteemed persons.

As almost every employer, GESIS is not entirely perfect. Nonetheless, I always perceived it as a great place to work and learn. It is a great place to connect with people within and beyond the institute and work on the challenges ahead, whether they consist in short-term or long-term problems of surveys or in finding a strategy in a global pandemic. A good number of colleagues have easily become dear friends in the last five years, through repeated afterwork drinks, weddings, rooftop parties, or hikes in the Palatinate. GESIS also allowed me to serve the Leibniz PhD Network as co-spokesperson of Katarzyna Stoltmann. Kasia, our Steering Group, and the Leibniz PhD Network taught me much about networks, voluntary work, science, and the perks and manifold challenges of pursuing a PhD in different disciplines in Germany. It was my pleasure to work

on two rounds of the Leibniz PhD Survey in interdisciplinary teams of volunteers. I very much hope that our team improved the working conditions of fellow doctoral researchers in some ways, and I wish the Leibniz PhD Network and its partner networks all the best while following up on the tasks ahead for a better scientific career development in Germany.
It is simple but true, my genuine and long-lasting happiness has its roots in my private sphere. I would like to thank my parents, Sabine and Ludwig. You always stimulated my curiosity, encouraged me to pursue my path, and constantly supported me in many ways. I say thank you, Claudia and Peter. The term parents is more appropriate for you than the term "in-laws". I say thank you to my terrific friends, who I met in Cologne, Mannheim, Stuttgart, Berlin, and elsewhere. You all have constantly empowered me with your confidence in my strengths. You all got me easily distracted and focused on other important and beautiful aspects of life. Having said this, I wholeheartedly thank the partner of my life, Charlotte. We got to know and love each other while still being in school, helped each other to prosper and develop as human beings, and always nurtured our long-lasting bond with joy, energy, and perseverance. I am so much looking forward to opening the next chapter with you.

References

Beadle, Brian et al. (2020). *Being a Doctoral Researcher in the Leibniz Association: 2019 Leibniz PhD Network Survey Report*. Tech. rep. Leibniz PhD Network.
URL: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-69403-1

Contents

List of Figures

List of Tables

Chapter 1

The institutionalized population and social surveys

1.1 Introduction

Once I started to work on the topic of this doctoral thesis in 2016, I was primed to make the observation of how often residents of prisons, refugee accommodations, retirement homes and other institutions appear in literature, films, fine arts, and the media. Harry Potter lived in a boarding school for seven years, while Charles Dicken's Oliver Twist spent a miserable time living in an almshouse. Max Liebermann painted men in a retirement home and women in a Dutch orphanage.[1] Prisoners are often protagonists, for instance, in the Netflix production "Orange is the new black", or in one of the most successful films in film history, "The Shawshank Redemption", which is based on a novella by Stephen King. German and international media report about individual cases of refugees, prisoners, or residents of retirement homes, or treat related subjects on a higher level when publishing on international humanitarian crises, the judiciary system, or demographic aging.[2]

This list is far from being exhaustive, but it already shows the great artistic interest in people in institutions. In contrast to literature or films, survey researchers working on social surveys rather neglected this part of the population in the last decades, often excluding them from their surveys. Three empirical observations can be made: First, the institutionalized population is very heterogeneous, including diverse groups such as pupils and students, refugees, prisoners, or retirees. Second, most of us "learn" more about the institutionalized population from films, literature and the media than from survey research. Those sources shape our perspective on this population as

[1] "Altmännerhaus in Amsterdam" (1880) and "Freizeit im Amsterdamer Waisenhaus" (1882).

[2] Recent examples from the New York Times Topic on "Refugees and Displaced People" (available online) showing various articles on the refugee camp Moria in Greece, the German weekly journal ZEIT (No. 38, 2020, "Zwölf Leben" (Twelve lives)) portrayed inhabitants living in a retirement home in Hamburg, an editorial by the Guardian on the 24th of May 2020 described implications of the Covid-19 lockdown on prisoners, while additional articles are available online in a Guardian collection on prisons and probation.

citizens and survey researchers. In line with Goffman's (1957) rather gloomy description of institutions, which might definitely be justified for the 1950s, the artistic depiction of institutions and their residents potentially leads to prejudices, such as highly dangerous prisoners who cannot be interviewed.[3] In addition to methodological and financial considerations described below, those prejudices could prevent social surveys to start covering this population. Third, the high visibility of institutionalized residents in art and media does not match with their small demographic share in countries worldwide (cf. Chapter 2). Social surveys are less interested in exceptional cases and more dedicated to show the bigger picture of the entire society or large subgroups in the population (cf. Stoop 2014). However, their nature as outliers in many ways might also imply that survey programs are missing a small, but disparate part of the population, whose inclusion could even alter results derived from those surveys.

Respondents differ in the level of challenge they impose on survey programs and their interviewers. Tourangeau and colleagues (2014) published a pivotal edited volume on *hard-to-survey* populations, putting forward a definition of this label, which can be broken down into *hard-to-sample*, *hard-to-identify*, *hard-to-reach*, *hard-to-persuade*, and *hard-to-interview* (Tourangeau 2014). The edited volume provides multiple examples of groups that can be labeled as hard-to-survey. Among others, displaced persons (Brodie et al. 2014; Pennell et al. 2014), people with intellectual disabilities (Malam et al. 2014), economically deprived groups (Nicaise and Schockaert 2014) or the homeless population (Glasser et al. 2014) are all hard-to-survey for different reasons (cf. Bacher et al. 2019; Willis et al. 2014). Except for the homeless population, those groups can live in private households, but they also have a high likelihood of living in institutions.

The hard-to-survey label applies for various reasons to the institutionalized population. According to a majority among 40 survey researchers with experiences in surveying institutionalized residents, they are hard-to-reach due to gatekeepers operating the institutions, and hard-to-interview because of their special living circumstances, potential language barriers, or mental and physical health limitations (Schanze and Levinson 2019, 65, cf. Feskens 2009). Additionally, roughly four out of ten survey researchers think they are hard-to-sample due to their very small relative size or the missing sampling frame, and hard-to persuade to take part in the survey interview (Schanze and Levinson 2019). Still, even in the light of those difficulties a significant number of survey programs have covered institutionalized residents in recent decades (cf. Section 1.3).

This doctoral thesis is not intended to win survey researchers round to let them redefine their target population in order to include institutionalized residents. The thesis can hopefully be used to justify the inclusion *or* exclusion of institutionalized residents, not by taking the easy way out or by simply following traditional practice, but by looking at the facts and developing evidence-based expectations. Decisions about the institutionalized population should be made by considering the target population and the topics the survey is dealing with. Which research questions would data users like to answer? Would it potentially make a difference to include institutionalized residents in those analyses? Those questions deal with the *necessity* to cover institutions. Even though it is

[3] In fact, a survey carried out end of the 1990s in the UK identified 15 out of 3563 prisoners (0.4%) as too dangerous or too disturbed for a survey interview (Singleton et al. 1998).

not the core topic of this doctoral thesis, I also urge survey researchers to think about the methodological *feasibility* of an inclusion, in the light of cost-benefit assessments and given budgets.
This doctoral thesis developed out of my participation in the SERISS project, which was financed by the EU Horizon 2020 framework program for research and innovation.[4] This collaborative research project conducted by several large cross-national survey programs, namely the ESS, EVS, GGP, and SHARE, aimed to tackle various fundamental questions social surveys are confronted with. Among other topics, the SERISS project gathered experiences by survey programs covering institutionalized residents, listed country-specific sampling frames, and developed recommendations on how to optimize the coverage of institutionalized respondents. Although the project itself did not have any resources to collect data in institutions, two deliverables and an inventory of survey programs covering and excluding the institutionalized population are the results of this task in the SERISS project (Schanze 2017; Schanze and Levinson 2019). Work done in this project also fed into the research for this doctoral thesis.
The doctoral thesis covers the relation of the institutionalized population and social surveys. Social surveys are defined as quantitative data collections based on large samples of respondents. In terms of content, social surveys cover a broad range topics of societal, economic, or political relevance, sometimes dedicating a part of their questionnaires to delve into more specific topics without loosing its broad perspective. In the present chapter, I advance a definition of institutions and the institutionalized population relying on previous work done in the SERISS project (cf. Schanze and Levinson 2019). Following the definition, I summarize how various survey programs deal with the population living in institutions. Section 1.4 presents the Total Survey Error framework (TSE) as a concept to describe and estimate errors of observation and non-observation in surveys. Elements of the TSE covered in this doctoral thesis are introduced while relating them to survey research in institutions. The TSE is very well suited to guide survey researchers while thinking about trade-offs of a potential exclusion or inclusion of the population living in institutions, considering costs and benefits of adaptations in survey methodology. Following the setting of the scene, the four main chapters forming this doctoral thesis are summarized, namely their research questions, their research methodology, and a summary of main findings. The remaining chapter is dedicated to point out opportunities for future research, namely regarding the necessity and feasibility to cover institutions and potential adaptations of survey instruments. The chapter concludes with a discussion of the relation of social surveys and the institutionalized populations.

1.2 Defining the institutionalized population

Belonging to the institutionalized population is determined by the housing situation, which is why institutionalized residents are usually not hard-to-identify (Schanze and Levinson 2019). Institutionalized people have their usual place of residence within institutions and spend most of their

[4] Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS), more information available at https://seriss.eu

time at those addresses.[5] The OECD glossary of statistical terms advances a list of institutions, containing educational institutions, health care institutions, institutions for retired and elderly persons, military institutions, religious institutions, and other institutions (OECD 2006). The latter includes prisons, refugee accommodations, or worker's dormitories.

Figure 1.1: Defining institutions and the institutionalized population (Schanze and Levinson 2019, 10)

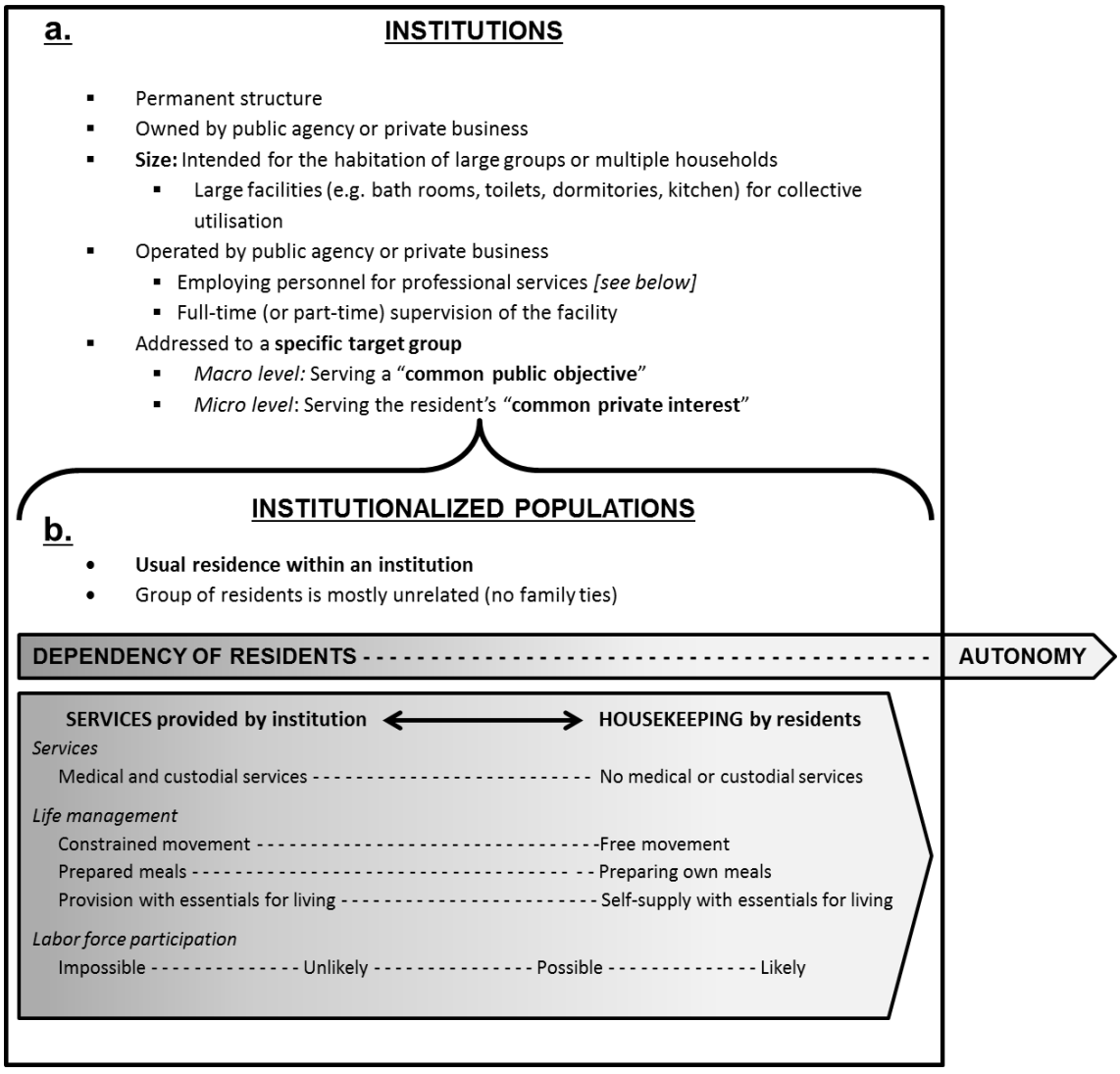


Figure 1.1 was developed by Schanze and Levinson (2019, 10) to define institutions and the population living in those institutions. Institutions are permanent places *"intended for habitation by large groups of individuals or several households"* (United Nations 2008, 196), which is why they are equipped with *"common facilities, such as cooking and toilet installations, baths, lounge*

[5] A few private households are also located within institutions, usually to host staff employed by the operators of institutions (cf. Bajekal et al. 2006).

rooms or dormitories, [...] shared by the occupants" (ibid.). In contrast to private households, institutions are always owned and operated by a public agency or private company, which employs staff for the daily business and full-time supervision of the institution (cf. Schanze and Levinson 2019 and Figure 1.1). The operators and their staff take over responsibilities on behalf of their residents (Goffman 1957). By doing so, operators of institutions fulfill collective demands by society or individual interests, such as providing education, medical services and care, or national security (Schanze 2017; Schanze and Levinson 2019).

Residents having their usual place of residence in institutions are described in the lower half of Figure 1.1. The place of residence is the housing "*where a person normally spends the daily period of rest, regardless of temporary absences for purposes of recreation, holidays, visits to friends and relatives, business, medical treatment or religious pilgrimage*" (European Union 2008, 15). This census definition is amended in Europe and elsewhere with a temporal qualification, stating that people should have spent most of the recent year or intend to stay most of the coming year at this respective place (ibid., United Nations 2008). Surveys can adapt the reference period according to their needs. In any case, the temporal qualification is a highly relevant question for the institutionalized population (cf. Beukenhorst et al. 2011). Short-term institutionalized residents, such as inpatients in hospitals, refugees living in refugee accommodations, or prisoners in pretrial imprisonment, might be defined as living in private households instead, being only temporarily not available at their usual place of residence (Schanze 2017).

Private households usually host single persons or groups of persons, who might be related or not, might pool their incomes or not, but definitely keep their households autonomously, providing themselves with food and all essentials for living (Eurostat 2011). This is not the case for the institutionalized population. The group of residents living within an institution is not related and does not have economic ties with one another. Institutionalized residents rely on the services offered by the operators because they are not able to maintain their household independently. Schanze and Levinson (2019) described the degree of dependency on the services as a continuum that differs across institutions and within institutions (cf. Figure 1.1). Residents living in student dorms face very little restrictions of their movement, while the opposite is the case for prisoners. In a retirement home, inhabitants are less relying on the medical services compared to residents living in a nursing home. Many institutionalized residents are not able or required to work, for instance, due to the status of their application for asylum, their health condition, or their advanced age, whereas some types of institutions are defined by the participation of their participants in the labor force, e.g., in worker's dormitories, or military barracks (cf. U.S. Census Bureau 2012). Difficulties in the definition of institutions emerge with new hybrid types of housing arrangements. Assisted living residences for the elderly population are one example of blurred lines between institutions and private households (Lewinter 2004; UNECE 2020). The continuum shown in Figure 1.1 can help to disentangle those hybrid types of housing by determining to what extent residents rely on services provided by operators.

This definition has several implications for social surveys. First, the institutionalized population is not a stable group in a sociological sense. Every member of the population living in private

households can move to an institution for various reasons in the course of a life, potentially leaving the institution after a while and continue living in a private household.[6] Second, moving to an institution implies a strong (self-)selection mechanism. Most of the following chapters emphasize this self-selection mechanism because it creates the peril of coverage bias for social surveys. The strong heterogeneity of the various types of institutions implies very different reasons for moving to an institution. Section 3.3 summarizes the respective reasons for retirement and nursing homes. Third, the dependency of residents on services as described in Figure 1.1 shape their ability to participate in surveys. Certainly, institutionalized residents lose their autonomy to a greater or lesser extent when moving to an institution, which is why social surveys need to get in touch with operators of institutions or legal custodians before contacting their respondents. The degree to which institutionalized residents need to rely on the services can be "*relevant to figure out the extent of adaptations they* [survey researchers] *need to make in the course of fieldwork*" and earlier in the design phase (Schanze and Levinson 2019, 11).

1.3 The status quo of survey research in institutions

The label of institutionalized residents as hard-to-survey and potentially rising costs associated with their inclusion are two reasons why many large cross-national and national survey programs exclude institutionalized residents from their definition of the target population (cf. Schanze and Levinson 2019). To name only a few examples, health surveys like the *Canadian Community Health Survey* (CCHS), the *Health Survey for England* (HSE), or the *Swiss Health Survey* (SGB) all exclude institutionalized residents, just as large cross-national survey programs like the *European Social Survey* (ESS), the *European Value Study* (EVS), the *European Quality of Life Surveys* (EQLS), or the *Programme for International Assessment of Adult Competencies* (PIAAC) in all countries except in the USA.

Reasons for not covering the institutionalized population are diverse. In 2002, the *Health Survey for England* (HSE) fielded a pilot survey trying to cover institutionalized residents. In the following waves, the HSE again covered only those residents living in private households. A researcher working on the project named (1) higher costs of interviews in institutions due to proxy interviews, (2) a missing high quality and up to date sampling frame of institutions with auxiliary information on the size of the institutions and (3) a lack of interest by policy makers as reasons for not continuing the effort (Schanze and Levinson 2019). For the *British Crime Survey* (BCS), Pickering and colleagues (2008) concluded that an adequate coverage of institutions would not be feasible without increasing the budget substantively.

On the other hand, many survey programs already covered institutions or started to cover them recently. Among others, the SERISS Survey Inventory lists a few cross-national survey programs which cover the institutionalized population partially or in some countries, like the *European Health Interview Survey* (EHIS), the *Labour Force Survey* (LFS), or the *Survey of Health, Ageing*

[6] See Chapter 2 for an analysis of average duration in institutions.

and Retirement in Europe (SHARE). National or regional surveys like the German *NRW80+* survey, the Swiss *Vivre/Leben/Vivere* (VLV), the French *Enquêtes Capacités, Aides et Ressources des seniors* (CARE), *Onderzoek naar ouderen in instellingen* (OII) in the Netherlands, or the *English Longitudinal Study of Ageing* (ELSA) all cover residents living in retirement or nursing homes. In a collaborative project with the IAB and the Federal Office for Migration and Refugees, the German *Socio-Economic Panel* (SOEP) covers refugees living in refugee accommodations (cf. Brücker et al. 2017), just like another German panel survey dealing with the labor market and social security (PASS). In the UK, the *Survey of New Refugees* (SNR) conducts survey interviews with immigrants shortly after their arrival in the country (Cebulla et al. 2010), while the REHEAL survey focused on refugees arriving in Greece after 2014 (Stathopoulou and Eikemo 2019). The *Adult Psychiatric Morbidity Survey* (APMS) and the *Millenium Cohort Study* (MCS) in the UK, or the *National Survey of Youth in Custody* (NSYC) in the USA are examples for survey programs covering prisoners. Also in the USA, the PIAAC National Supplement assessed the skills of a sample of more than 1,300 respondents in prisons (Rampey et al. 2016). Those 16 surveys and more than 100 further survey programs listed in the SERISS Survey Inventory collect data in institutions (Schanze and Levinson 2019). Thus, the claim of a general ignorance of institutionalized residents by the world of survey researchers would definitely be misleading.

In survey programs that include institutionalized people, the inclusion of institutions alongside private households is most common (cf. Schanze and Levinson 2019). A smaller group of survey programs focuses entirely on institutions without interviewing any residents living in private households. Since the focus of this doctoral thesis is on social surveys, the joint inclusion of private households and institutions is at the heart of all chapters. However, evidence from surveys only covering institutions can be relevant for social surveys when considering the feasibility and specific methodological challenges.

Depending on the definition of the target population, the survey design might be subjected to methodological changes. If a social survey decides to cover institutions alongside private households, it could either abstain from excluding institutions without making any methodological adaptations, or it could implement modifications to improve coverage and interviews of hard-to-survey populations. The latter option could be further expanded to allow separate analyses for the group of institutionalized residents. Given the small relative size of the institutionalized population and the challenges associated with its inclusion, this option would lead to stronger methodological adaptations, like using oversampling of institutions or specific demographic groups with a high share of institutionalized people (Schanze 2017).

Cross-national surveys in institutions are rare (Schanze and Levinson 2019). The definition of the target population and decisions about the eligibility of institutionalized residents are especially challenging for survey programs if they need to take into account national particularities while ensuring cross-national comparability. The LFS and EHIS define private households as their target population and leave it to the participating member states to extend their definition to institutions (Schanze 2017). This is why some of the countries include institutionalized residents - sometimes with proxy responses in the LFS - while others limit their definition to private households (ibid.).

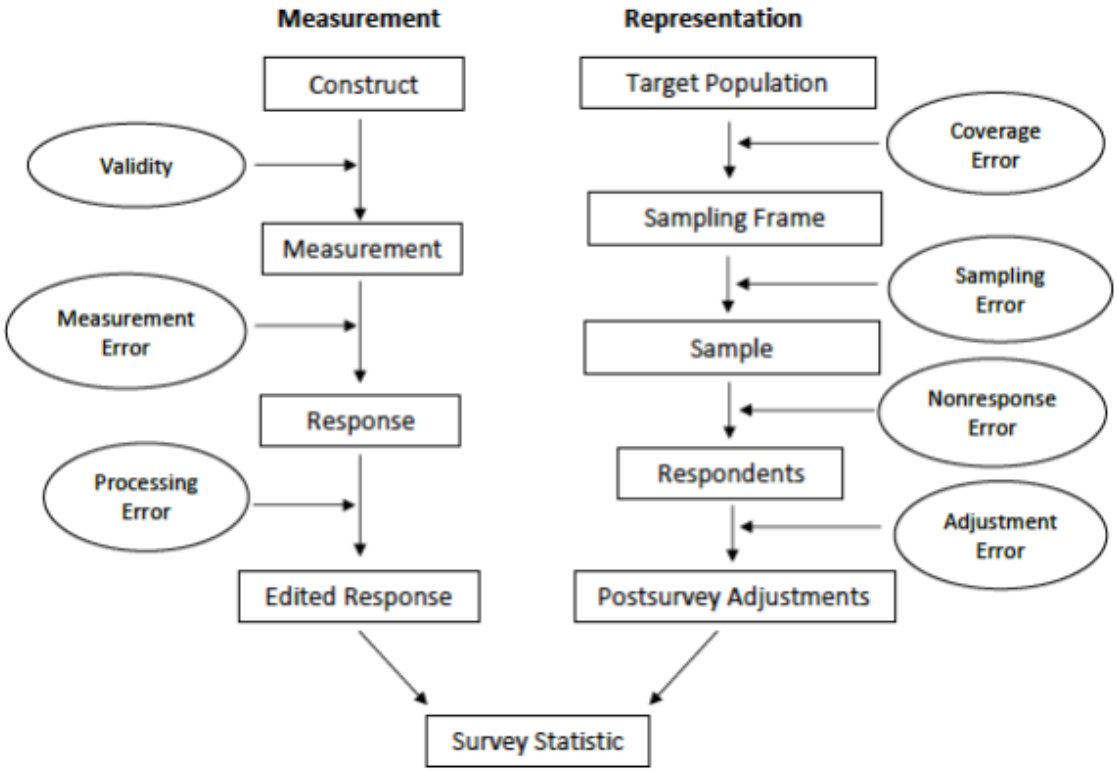
The EHIS implemented a task force to assess the methodological implications of including institutionalized residents and how it is conducted in single countries (Beukenhorst et al. 2011). In contrast, SHARE advances central guidelines on how to define the target population. This survey program excludes residents who are incarcerated or hospitalized, but wants to include respondents who live in institutions for the elderly (Bergmann et al. 2019). Interviewers are provided with a brief definition to identify nursing homes in the fieldwork (cf. SHARE ERIC 2020).
Obviously the intention to cover institutionalized residents in a survey program does not imply a successful realization of the plan. Multiple potential sources of error exist in the survey cycle. The following section introduces the Total Survey Error framework and explains which errors survey programs are dealing with when excluding institutionalized people *and* when aiming to include them.

1.4 Total Survey Error framework and the institutionalized population

This doctoral thesis approaches various aspects of survey quality related to a specific subgroup of the population. To systematize and assess survey quality conceptually and practically, survey researchers rely on the Total Survey Error framework (Groves et al. 2009; Groves and Lyberg 2010). This framework has multiple roots and knows various configurations (Groves and Lyberg 2010; Lyberg 2012). Figure 1.2 shows a common version of the Total Survey Error framework (TSE), which consists of two branches (Groves et al. 2009, 48). The right branch deals with *errors of non-observation* stemming from insufficient inclusion of the inferential population in a survey sample due to coverage error, errors in the sampling design, or nonresponse error. The left branch deals with *errors of observation* in the measurement of concepts and topics, while developing survey items and administering them to respondents. Errors accumulate over the successive steps, from the construct that should be measured with a survey item to the final edited data, as well as from the target population which should be covered to the final net sample of respondents (Biemer 2010; Groves et al. 2009).
When considering to cover institutionalized residents, two aspects need to be evaluated: First, the *necessity* of including institutionalized residents due to potential bias for specific topics. Second, if the inclusion is considered as necessary, questions on the *feasibility* of doing so with appropriate survey methods and a given budget will follow. The latter step needs to take into account the hard-to-survey nature in the case of institutionalized residents. Questions about the necessity can be located at the representation branch of the TSE, while the topic of feasibility includes elements of both branches.
In all methodological considerations, survey costs always play a crucial role in decision making. Survey researchers have limited budgets and resources and need to evaluate any methodological decision with a cost-benefit perspective. In other words, designing a social survey is about "*optimizing surveys by minimizing the accumulated size of all error sources, given budgetary con-*

straints" (Lyberg 2012, 109). After considering necessity, feasibility, and the costs of including institutionalized people, it might make sense to limit the definition of the target population to the population living in private households and invest more resources in other measures to improve survey quality, such as interviewer training, incentives, or better data processing.

Figure 1.2: The Total Survey Error framework (cf. Groves et al. 2009, 48)



Groves and Lyberg (2010) name the absence of key quality characteristics as developed by a number of international organizations and National Statistical Offices as a weakness of the TSE framework (cf. Biemer 2010; Lyberg 2012). The framework focuses on indicators of accuracy and neglects other quality indicators, such as relevance, timeliness, or accessibility, which are all essential measures for the user-friendliness of a survey program (Biemer 2010). As Feskens (2009) points out, the exclusion of institutionalized residents could be classified as a violation of the *completeness criterion*, which is a subcategory of relevance. It is up for discussion in this doctoral thesis and beyond, whether this violation of completeness is severe or negligible. Moreover, methodological accuracy is *the* key element of survey quality and deserves the utmost attention (Groves and Lyberg 2010; Lyberg 2012). The topic of completeness as another quality indicator is still discussed in the present doctoral thesis whenever dealing with the necessity of an inclusion and the peril of bias.

The following sections introduce survey errors of high relevance for this doctoral thesis. Chapter 2

uses administrative benchmark data to guide the definition of the target population. Chapters 3 and 5 both deal with undercoverage and noncoverage, which can be caused by the insufficient definition of the target population, sampling error, or nonresponse error. Section 1.4.1 contains relevant information for those chapters. Chapter 4 deals with the measurement branch, focusing on respondents' behavior and data quality (cf. Section 1.4.2). The overall focus lies on face-to-face surveys, since this mode administered by interviewers is still the prevailing mode in social surveys and in survey programs including the institutionalized population (cf. Schanze and Levinson 2019). All survey data analyzed in the various chapters of this doctoral thesis were collected in a face-to-face made. Even though Chapter 5 contrasts probability-based samples with non-probability quota samples, the following section primarily covers probability-based sampling, which is still the default method in most survey programs, also in those survey programs covering institutionalized people (cf. Schanze and Levinson 2019).

1.4.1 Representation and errors of non-observation

The term "non-observation" implies that certain people are missing in a survey sample because they were lost at some point as a consequence of coverage error, sampling error, or nonresponse error. When determining the extent of the resulting bias, the size of the group of missing people and their statistical distinctiveness are the two relevant factors (Groves et al. 2009; Stoop 2014). How large is a group that belongs to the population of interest but is missing in the final net sample? Does this group differ in the variables measured in the survey when comparing it to those people who took part as respondents? This subsection discusses various reasons that could lead to incomplete net samples, especially with respect to the institutionalized population. A *single* estimate of bias for a net sample is not sufficient; any distortion caused by the definition of the target population, insufficient coverage, or nonresponse varies across survey estimates depending on their topic (cf. Groves and Peytcheva 2008).
The following paragraphs outline some of the considerations survey researchers have to make in the process of defining their target population. This is the step where errors due to the deliberate exclusion of institutionalized people are considered, whereas the subsequent parts on sampling and recruitment of respondents rather deal with errors when institutionalized people *should* be included according to the definition of the target population but are missing for other reasons.

Defining the target population

The definition of the target population specifies which finite population is eligible to partake in the survey in a given period of time (Groves et al. 2009). It should be mainly determined by questions of necessity. As mentioned in chapters 2 and 3, social survey programs usually aim to allow their users to draw conclusions for the general population living in a specific region or country. Any group in the general population that reaches a significant relative size and is expected to differ on specific topics included in the survey should be defined as eligible to take part in the survey.

Chapter 2 uses census data for calculating the relative size of institutionalized residents in different age and gender groups. It also delves into the statistical distinctiveness of institutionalized people, as does Chapter 3. It is a key assumption of the present doctoral thesis that the (self-)selection of moving to an institution increases the statistical distinctiveness in various ways and thereby also increases the peril of bias if institutionalized residents are not included in the definition of the target population.

Sometimes the definition is already driven by the feasibility of conducting the survey, for instance, resulting in the exclusion of geographically remote parts of a country in face-to-face surveys (Smith 2014). The same rationale can play a crucial role when considering the inclusion of institutionalized people or other hard-to-survey groups. For the sake of lower costs and higher response rates, the easiest method would be to exclude any hard-to-survey respondents in this step of the survey cycle (Stoop 2014). Under a more rigorous methodological approach, only questions about the necessity would determine whether people living in institutions are included or excluded from the definition of the target population. If it is deemed necessary, all the following steps of the TSE only deal with questions of feasibility.

Sampling design and sampling frame

Once the target population has been determined, thoughts about the sampling design and potential sampling frames ensue. Kalton (2009, 128) developed a few guiding questions for sampling rare populations (cf. Schanze 2017): The type of sampling design, the availability and quality of sampling frames, the size of the population of interest, and their regional distribution all have an influence on the quality of the gross sample and potential noncoverage or undercoverage.

The type of sampling frame, namely whether it provides addresses, dwellings, or names of individuals, makes a difference for the selection procedure and the bias in the resulting gross and net sample (Kölln et al. 2018). As Schanze (2017) points out, an address-based sampling frame or a random route procedure could result in lower selection probabilities of institutionalized residents, because a large number of residents is clustered in single addresses (cf. Kalton 2009). Moreover, address-based sampling frames come with a greater responsibility for interviewers during the selection procedures, which can also lead to sampling error induced by interviewers (Eckman and Koch 2019; West and Blom 2017).

If a sampling frame is readily available, survey programs should always assess whether it matches their definition of the target population to be in better control of errors caused by undercoverage or overcoverage (cf. Lohr 2010). This check also applies to addresses of institutions, irrespective of whether they should be included in the survey or not.[7] Various methods can be used to identify those addresses in a sample, for instance, by matching addresses to a database of institutions (Gaertner et al. 2019). If a survey aims to include institutions, the sampling frame should not suffer from noncoverage or undercoverage. Sometimes institutions are fully or partially missing in sampling frames, for instance in Bulgaria, Estonia, Finland, Italy, Poland, or Switzerland (Bergmann

[7] If a sampling frame contains addresses of institutions and the survey aims to exclude them, interviewers need to be informed about characteristics of ineligible units (cf. Section 1.2) to discard them during fieldwork.

et al. 2019, 10, De Luca et al. 2015, Schanze and Levinson 2019, 33). However, the main purpose of any population register does not consist in serving as a sampling frame for social surveys (cf. Poulain and Herm 2013). This is why information on their coverage might be incomplete or misleading (cf. De Luca et al. 2015).

As another source of undercoverage, institutionalized residents might still be registered at their last address and cannot be easily found when conducting the contact attempt (cf. Schanze and Levinson 2019). The issue of outdated addresses is especially pronounced in the highly mobile group of refugees. A recent survey on refugees in Germany could not locate 30% of their sampled cases, even though the survey aimed to reduce the time between sampling and fieldwork by drawing the sample in six tranches (Kühne et al. 2019). To mitigate the negative impact of undercoverage, oversampling might be considered to increase the number of hard-to-survey respondents in the gross sample. Kalton (2009) presents related methods such as screening samples of other survey programs, using disproportionate stratification or multiple sampling frames to reduce undercoverage of rare populations.

Unit nonresponse

As a next source of error leading to incomplete survey samples, selected persons in the gross sample might drop out before the interview can take place. Various reasons can explain unit nonresponse: Some of the selected persons cannot be reached during fieldwork, another group refuses to respond, while a third group is not capable of responding due to language barriers, health, or cognitive impairment (cf. Beullens et al. 2018; De Leeuw et al. 2018; Lohr 2010). If residents living in institutions are harder to contact and harder to interview, a lower share of this group might end up in the final net sample due to non-contact and limited capability.

Before considering nonresponse by individuals, operators of institutions add another layer of complexity for the institutionalized population (cf. Groom et al. 2009). A refusal at the level of institutions can occur if the survey program needs to rely on the operators for getting access to the institution, sometimes for drawing a sample of respondents, e.g., if it uses a sampling frame of addresses. In those cases, institutions need to cooperate and some of them are not willing or able to do so (Schanze and Levinson 2019). A refusal of institutions makes institutionalized respondents hard-to-contact. To give a few empirical examples, 98% of the U.S. prisons sampled for PIAAC cooperated with the survey program (Rampey et al. 2016), while 30.7% out of 200 retirement and nursing homes did not take part in a Dutch pilot study (Feskens 2009). Most probably the latter number will be much lower if a survey program already knows the respondents after drawing them from a sampling frame of individuals (cf. Wagner et al. 2018). Once the survey program has access to the institution, nonresponse due to non-contact will be less problematic, because the institutionalized people are easy to locate and gatekeepers can even help to schedule an appointment (Kammerer et al. 2019).[8]

Why do individuals refuse to participate in a survey interview? In Round 9 of the ESS, many

[8] As an exception, more mobile institutionalized group, such as students living in dorms, might still be difficult to locate (Groom et al. 2009).

refusing people justified their refusal with a lack of interest, bad timing of the invitation, or the fact that they would never participate in surveys.[9] A lack of interest and a general refusal to survey invitations might also occur in institutions. However, due to special living circumstances, a lack of time might be less relevant in institutions compared to private households. Probably a request to participate in a survey attracts greater interest in institutions, as it could be perceived as a welcome distraction. In the expert survey carried out for the SERISS project, a majority of survey researchers confirmed that institutionalized people are not hard-to-persuade to take part (Schanze and Levinson 2019, 65). Empirical evidence from a range of surveys in different types of institutions seems to confirm this position (cf. Brücker et al. 2017; Feskens 2009; Kühne et al. 2019; Rampey et al. 2016; Wagner et al. 2018).

In some types of institutions, e.g., in institutions for the elderly, the low refusal rate comes at the expense of a higher share of nonresponse due to health-related impairment (Wagner et al. 2018). Four out of ten respondents in retirement and nursing homes could not take part in the survey interview due to their capacity (Feskens 2009). The ability of respondents is a key question in this environment, which is why survey programs in many types of institutions need to determine a procedure on how to ascertain respondents' capacity. For the elderly population, operators in institutions or relatives might tend to "overprotect" respondents by notifying they cannot participate due to incapacity (Neuert et al. 2016; Sangl et al. 2007). Testing respondents with cognitive screening questions could be a more objective way of determining their ability (Sangl et al. 2007). However, ethical questions should be carefully considered, since residents might be emotionally disturbed after failing the test. In other types of institutions the situation obviously looks different: Roughly 1% of the sample of newly arrived refugees could not be interviewed, another 5% could not participate due to language barriers (Kühne et al. 2019, cf. Rampey et al. 2016 for prisoners). The latter number depends on how many languages the questionnaire is translated into.

Two chapters included in this doctoral thesis use panel data collected by SHARE. For panel surveys aiming to interview institutionalized respondents, questions emerge on whether the panel attrition or panel mortality is higher among those respondents moving to institutions or already living in institutions. Leaving a panel survey after answering one or several waves never happens completely at random (De Leeuw et al. 2003). Institutionalization can be clearly considered as an external shock, which might decrease the likelihood of taking part in the following wave (cf. Lugtig 2014). However, clear evidence is lacking, for instance for SHARE (Bergmann et al. 2019). For an English cohort study working on elderly persons, Davies and colleagues (2014) found a clear impact of health on the likelihood of panel retention, with some indication that institutionalized residents have a higher likelihood of dropping out of the sample also for other reasons than health. Again, refugees are especially mobile shortly after arriving in their host country, which creates difficulties in keeping those respondents in a panel survey (cf. Cebulla et al. 2010). The duration of the institutionalization can be a meaningful piece of information to assess the likelihood of panel attrition (see Chapter 2).

[9] Results are based on own calculation of ESS Round 9 Contact Forms data (edition 2.0)

Item nonresponse

Item nonresponse is a final error of non-observation at the level of single survey items. The rate of missing items is analyzed as a quality indicator in Chapter 4 of this doctoral thesis. Three main types of missing values can be distinguished: Don't know responses, refusals, or no answer (Koch and Blohm 2009). Difficulties of respondents with the concept or the specific type of measurement (cf. Section 1.4.2) as well as memory problems can translate into Don't know replies.[10] In the case of a refusal the respondents usually know the reply but do not want to reveal it to the interviewer (Koch and Blohm 2009). The no answer category is caused by accidentally skipping questions or routing questions. In the first two types of item nonresponse survey estimates have a higher likelihood of not missing at random (De Leeuw et al. 2003). The listwise exclusion of those cases in any analysis can potentially lead to biased estimates. Research on item nonresponse is strongly linked to research on measurement errors. This is why further implications of item nonresponse for survey research in institutions are discussed in the section below.

1.4.2 Measurement and errors of observation

The measurement branch of the Total Survey Error framework starts with a theoretical construct that should be measured by the survey item. This step, as well as the final editing step of responses and processing error will not be considered here, as they are of low relevance for the papers included in this doctoral thesis.

Measurement error

Once a survey item is included in the questionnaire, measurement error can occur in the phase of administering the items during fieldwork. Each respondent ideally performs several cognitive steps while answering a survey question: Understanding the question and identifying which information is required to answer it, retrieving relevant information from memory, formatting it to the answer scale provided, judging whether the information is correct and finally communicating the reply (Alwin and Krosnick 1991; Schwarz 2007; Tourangeau et al. 2000). Errors in the process can be caused by the type and difficulty of the survey question, respondents' characteristics and their motivation, as well as the influence of interviewers (Alwin and Krosnick 1991; Krosnick 1991; West and Blom 2017).

It is the underlying hypothesis of Chapter 4 that institutionalized residents living in retirement and nursing homes are prone to produce more measurement error and item nonresponse mainly due to their health and cognitive performance. Chapter 4 also summarizes the rich research on how those factors can influence various cognitive steps during a survey interview. For retirement and nursing homes, previous research shows how specific features of survey questions, such as

[10] It should be noted that a Don't know reply can be a meaningful response, e.g., if respondents need to answer a knowledge question or are asked which party they will vote for (De Leeuw et al. 2003; Koch and Blohm 2009).

negative phrasing, item complexity, the reference period provided, or the personal salience of a topic all influence difficulties in the cognitive process (Fox et al. 2007; Kutschar and Weichbold 2019; Sangl et al. 2007). For other types of institutions, a differences in cultural understanding or the native language could lead to similar problems.

1.5 The four papers of this doctoral thesis

As mentioned above, the four chapters included in this paper-based doctoral thesis cover different sources of error from the TSE presented in the previous section. Dealing with errors of non-observation, chapters 2, 3, and Chapter 5 examine whether it might be necessary to extend the coverage to institutionalized people by quantifying the impact of noncoverage and undercoverage. Chapter 4 deals with errors of observation, investigating whether the label of institutionalized residents as a hard-to-interview population is justified and whether the resulting data quality differs from respondents living in private households. The four papers use different data sources and different research methods. The following two sections summarize the research questions, the methodology applied, and the results obtained.

1.5.1 Research questions and methodology applied

In the first paper, Chapter 2 in this doctoral thesis, administrative census data from 30 European countries is used to offer a comprehensive overview of the institutionalized population and investigate the possibility of coverage error in social surveys if this population is excluded. The chapter contrasts institutions with private households regarding age, gender, marital status, citizenship, level of education, economic activity, size of locality, and residential mobility. It aims to provide empirical guidance in which cases the restriction of the target population might result in coverage bias in those basic socio-demographic and socio-economic variables when making inferences about the general population.

The chapter considers the size of the institutionalized population and its statistical distinctiveness as the two elements of the function that determines coverage bias. In addition to the descriptive analyses of the distributions of the variables of interest, Duncan's Index of Dissimilarity is calculated for the various variables to grasp the peril of coverage bias within the respective age groups in European countries.

The second paper, Chapter 3, continues the analysis of coverage bias for the elderly population. This demographic group has the highest likelihood of being affected by coverage bias according to results presented in Chapter 2. In joint work with Stefan Zins, we examine the potential coverage bias caused by noncoverage and undercoverage of the population living in retirement and nursing homes. If this population is excluded from survey samples, bias in variables that help to explain the likelihood of institutionalization is a possible outcome (cf. Tables 3.2 and 3.3). This chapter aims to quantify the impact of noncoverage and undercoverage for two health-related variables.

Furthermore, it also examines the potential of correcting for the bias by weighting the net samples. As research method we use a Monte Carlo simulation based on cross-national samples of the *Survey of Health, Ageing and Retirement in Europe* (SHARE). While the share of institutionalized residents varied in Chapter 2 across countries and age groups, it is steady in Chapter 3: We identify 2,500 respondents living in retirement and nursing homes in a pooled sample of more than 100,000 respondents aged 50 years or older. Thus, they amount to a small share of 2.5% in our empirical population, which can be considered a realistic proportion when cross-checking it with census data (cf. Table 2.3). We draw 5,000 random samples from this empirical population, only varying the degree to which respondents from the group of institutionalized residents are selected. Within those samples with varying conditions of coverage, we calculate point estimates for the self-rated health of respondents and for ADL, a variable measuring how many limitations in activities of daily living (ADL) respondents face. We check whether point estimates in each random sample include the true values with their 95% confidence interval. Moreover, we weight all samples for age and gender, as well as for age, gender, and institutionalization to assess whether the precision of estimates improves compared to design-weighted samples.

Chapter 4 moves away from the representation branch of the Total Survey Error framework and focuses on errors of observation. In multiple scientific studies, an increasing age and cognitive impairment are usual suspects for increasing difficulties in survey interviews and a decreasing data quality. Institutionalized residents living in retirement and nursing homes are older on average, suffer from a worse health condition and live more often with cognitive impairment. This is why survey researchers tend to label those residents as hard-to-interview and exclude them from most social surveys. The chapter sheds light on the overarching questions whether institutionalized respondents are harder to interview than comparable respondents living in private households and whether the resulting quality of survey data differs between those two groups.

Again, SHARE data is analyzed. As indicators of hard-to-interview respondents, I examine the share of proxy interviews, interviewer ratings of requests for clarification by respondents and their willingness to answer. As indicators of data quality, I analyze the extent of item nonresponse in two item batteries and acquiescence in balanced Big-Five items, and correlate hard-to-interview indicators with data quality indicators to investigate their potential detrimental effect on the data. Propensity Score Matching (PSM) is used to adjust and control for confounding sociodemographic and health-related variables explaining the likelihood of being institutionalized.

The fourth and last paper, Chapter 5, adds to the previous chapters on coverage bias by analyzing survey data by another large social survey program, the *European Social Survey* (ESS). It applies a broader perspective and asks whether elderly respondents are more adequately covered in surveys if surveys rely on probability-based sampling instead of using nonprobability quota sampling. The paper was written as a reply and extension to a paper published by Jean-Yves Dormagen and Laura Michel in French Politics (2018). In their paper titled “Do surveys accurately report voters over 80 years old?”, the two authors examine deviations of the aggregated self-reported electoral turnout in three French electoral surveys from the factual turnout in the French presidential elections in 2002, 2007 and 2012. The three election studies used a quota sampling method, which has a

higher likelihood of suffering from selection bias especially in those segments of the population that are difficult or even hard-to-survey.
In my reply, I replicate their analysis with probability-based survey data and test whether the demographic bias and the bias in turnout of the elderly population also occurs in randomly drawn samples. I analyze data of three rounds of the ESS, which were collected in 2002, 2008, and 2012 shortly after the respective presidential elections. As a second research question, the chapter aims to investigate whether the exclusion of institutionalized respondents in the ESS potentially helps to explain the stronger bias in turnout estimates for the oldest age group. For this purpose, synthetic institutionalized cases are added to the ESS samples. The number of synthetic respondents is based on the share of institutionalized residents in three different age groups as reported in the French census. Regarding the electoral turnout within this group, the simulation tests three different assumptions. Given their worse health condition, it can be assumed that this part of the population has a lower probability to vote than the elderly population still living in private households. This is why I assume a lower electoral participation of institutionalized cases by 5, 25, and 50 percentage points in the simulation.

1.5.2 Summary of results

A number of conclusions drawn from Chapter 2 might serve as a basis for future decisions about the eligibility or ineligibility of the institutionalized population. According to the 2011 census, the population living in institutions amounts to 1.3% in Europe. Considering the varying relative size of the institutionalized population in various age groups, the paper distinguishes between two groups of countries: In a first group of countries labeled as *one peak countries*, the overall share of institutionalized residents is very low in the young and middle-aged groups, whereas is constantly increases in the old age groups. In the second group of countries, the pattern looks similar, except for a second, smaller peak in the group of residents aged between 10 to 29 years (labeled as *two peaks countries*). Those two groups contain 24 out of 30 European countries. Due to the very unequal distribution of the institutionalized population across age groups, I conclude that social surveys should definitely consider extending their definition of the target population to the elderly population living in retirement and nursing homes, especially if they assess their survey topics as likely candidates for coverage bias. Survey programs operating in two peaks countries should also consider the consequences of excluding young institutionalized residents. From a methodological perspective, middle-aged institutionalized groups are not relevant quantitatively.
Regarding statistical distinctiveness, institutionalized people older than 85 years are less often married, either they have never been married or are widowed. This population lives more often in small and middle-sized towns and less often in villages. Differences regarding citizenship, the level of education, and economic activity are very small or even absent in this age group. Young institutionalized people are more often from non-EU countries, economically inactive, and live more often in large cities and metropolises than the respective age group living in private households. Generally, the heterogeneity within the institutionalized population can be emphasized.

Some patterns are common across European countries, on the other hand, a number of national particularities are also mentioned.
Chapter 2 only compares basic socio-demographic and socio-economic variables, which are probably less influenced by bias than health-related variables. Our results for two health-related variables presented in Chapter 3 show that noncoverage of institutionalized residents clearly leads to biased estimates. If those respondents are not included at all, more than 41% of 1,000 random samples miss the true share of respondents without any limitations in ADL with their confidence intervals, 11% of those samples miss the true share of respondents with a poor self-rated health. Weighting for age and gender improves those estimates, but cannot reduce the bias below an acceptable error rate lower than 5%. The precision of all estimates gets better the more institutionalized residents are covered. Thus, a higher number of random samples yield precise results under the condition of undercoverage compared to noncoverage. Further improvements of the precision can be achieved by applying post-stratification weights, either for age and gender, or for institutionalization in addition to age and gender. Looking at those results, we conclude that even a small number of institutionalized respondents missing in a net sample can lead to a measurable distortion of survey estimates. A very insufficient coverage of this hard-to-survey population would definitely be preferable compared to noncoverage.
Differences in socio-demographic, socio-economic, and health-related variables can translate into differences in attitudes or behavior. The electoral participation studied in Chapter 5 belongs to the small group of behavioral variables for which external benchmark data is readily available to study bias. In French quota surveys, Dormagen and Michel found a clear underrepresentation of the population older than 85 years in all three surveys and a strong overestimation of voting in this age group. For the French probability-based samples a less distinct bias is expected since interviewers cannot select their respondents but need to follow random selection procedures. Indeed, as Chapter 5 shows, estimates derived from ESS samples are closer to true values with respect to age distribution, level of education, and self-reported turnout. However, even in probability-based surveys the bias in demographic estimates and turnout is still stronger in the oldest age group compared to younger age groups.
The authors of the initial paper replied to this chapter and judged marginal improvements caused by probability-based sampling methods as minor (cf. Section 5.8 for my discussion of their reply). The comparably higher bias in the oldest age group in probability-based samples can be caused by a higher systematic nonresponse among the elderly population. It can also be caused by the fact that the ESS does not cover more than 13% of the French population older than 80 years living in retirement or nursing homes. Their exclusion might increase the distortion of self-reported voting behavior compared to the factual turnout of old-aged residents. According to results of the simple simulation in Chapter 5, 13% of additional institutionalized respondents older than 80 years could have reduced the turnout in 2012 from 89.3% to 86% if they voted 25 percentage points less, or even to 82.6% if they voted 50 percentage points less than their fellow citizens living in private households. Given that the factual turnout for this age group was 63% in 2012, it is evident that increasing the scope of the target population would not suffice to reduce the strong bias in survey

estimates.
In their reply to Chapter 5, Dormagen and Michel (2019) partly confirmed the assumptions made in the simulation and its results. According to their analysis of French administrative data, institutionalized residents have a much lower probability to vote, in line with the strongest scenario used in Chapter 5. However, institutionalized residents are also less often registered as voters, which is why their relative size within the electorate is arguably much smaller than assumed in my simulation. As summarized in Section 5.8, the estimated reduction of bias when including institutionalized people and people living in private households in France could be smaller than presented in Chapter 5 (Dormagen and Michel 2019).
Chapter 4 investigates the hard-to-interview label of institutionalized respondents. Indeed, a higher fraction of institutionalized respondents relies on proxy interviews in SHARE. Those respondents ask more often for clarification and are rated as less motivated by their interviewers than respondents living in private households. Those hard-to-interview indicators also have a significant impact on item nonresponse, which again is higher in institutions than in private households. Differences are also strongly influenced by confounding variables, such as age, health condition, and cognitive impairment, which is why differences between institutionalized respondents and other respondents get smaller, sometimes statistically insignificant, when controlling for confounding variables with PSM. In addition to institutionalization, health and cognitive performance appear to play important roles for the data quality irrespective of the housing situation. As another important finding, acquiescent response behavior is less frequent in survey interviews in institutions. I summarize that two out of three interviews with institutionalized respondents could be conducted without great difficulties. On the other hand, hard-to-interview respondents live in private households just as in retirement and nursing homes, even though the likelihood of difficulties in survey interviews is definitely larger in the institutional types of housing.

1.6 Further fields for reflection and research

As pointed out above, the necessity and feasibility of an inclusion of institutionalized residents are important factors guiding decisions about the definition of a target population. As described in the sections below, both factors require further reflection and scientific research when it comes to considering the inclusion of institutionalized people. While the sections on necessity and feasibility have a strong focus on applied survey research, the third section briefly advances ideas for substantive research.

1.6.1 On the necessity of an inclusion

Questions on the necessity to include institutionalized residents are not finally answered by chapters 2, 3, and 5 of this doctoral thesis. Finding answers to those questions require a case-by-case

assessment, taking the respective research topics and target population into consideration. Nevertheless, for most European countries, Chapter 2 concludes that social surveys should mostly focus on the population living in retirement or nursing homes when assessing potential bias. Chapter 3 summarizes multiple previous studies investigating institutionalized elderly persons and their transition to institutions. Any statistically significant differences between elderly residents living in institutions and private households are likely to cause coverage bias in samples restricted to private households.

Still, further analyses of bias are necessary, for different survey topics, the overall population, and for demographic subgroups. The availability of data often limits those analyses. Generally, analyzing bias, as well as response behavior, within given survey data without further external benchmark data as done in Chapters 3 and 4 comes with the limitation of potentially missing the truly critical residents and respondents in institutions *and* private households. This is why administrative data, notably census data, are a crucial resource for the assessment of coverage bias. They operate with very large sample sizes, participation is often mandatory. On the downside, those data sources only contain a few hard-facts without any behavioral or attitudinal variables. Moreover, they are collected at large time intervals, namely 10 years in Europe at the moment. Further administrative data are usually available at the national level. Other survey programs that cover institutionalized people can also provide valuable information about the institutionalized population and their empirical features to be used in analyses of bias (cf. Schanze and Levinson 2019).

Chapter 5 shows an easy way to estimate the potential coverage bias caused by the exclusion of institutionalized population. Information about the relative size of institutionalized groups is usually available, even for demographic subgroups. As a second step, researchers need to develop a set of assumptions how institutionalized residents could potentially differ with respect to any variable of interest. Information about correlations of those variables with confounding variables, such as health condition, mental performance, citizenship, marital status, or social networks might help to arrive at assumptions which are closer to reality. Artificial respondents could be added to results derived from survey responses to estimate the bias in point estimates. As Section 5.8 shows, it might be a good approach to also vary the relative size of institutionalized residents in this simple simulation, because not all of them will participate in the interview leading to a smaller relative size of institutionalized respondents in the net sample compared to their share in the overall population. As a limitation, this strategy is much more difficult to apply for any inferential statistics.

Except for the analysis in Chapter 5, all papers included in this doctoral thesis use cross-national data, even though they not delve into cross-national differences in bias, due to their scope and due to the small size of the institutionalized population in SHARE. Still, survey programs operating in multiple countries need to consider the peril of bias from a comparative perspective (cf. Smith 2014). Chapter 2 highlighted that the high degree of homogeneity regarding the institutionalized population in European countries does not imply perfect congruence across countries. A survey in private households in Hungary would miss 6% of the population aged between 15 to 29 years living in institutions, while it would only miss a little more than 1% of the respective age group in its

neighbor country Austria. The differences are even larger in the elderly population. In Belgium, France, or the Netherlands, every fourth women older than 85 years lives in a retirement or nursing home, while the respective share amounts to clearly less than 10% in Spain, Italy, or Poland. All things being equal, it is easy to imagine how conclusions about the health condition of elderly women would be more favorable in the first group of countries, simply because a critical fraction of the population with a worse health condition does not belong to the target population. Further analyses of bias in a cross-national context due to the exclusion of institutionalized residents are required.

Chapters 3 and 4 use panel data without making use of the longitudinal dimension for analyses at the level of individuals or at the macro level. At the micro level, panel attrition caused by the external shock of moving to an institution - or leaving it again - is a matter that requires further scientific attention. At the macro level, any societal or political changes in a country, or external events like the humanitarian crisis in Syria or a global pandemic can lead to rapid changes in the number of people living in institutions and their statistical distinctiveness, as well as the challenges they pose to survey programs. Unfortunately, we will have to wait at least half a decade until new cross-national data from the 2021 European census will be publicly available. In the meantime, researchers could compare the 2011 census with the less extensive 2001 census for some European countries. Tentative analyses conducted by the author of this doctoral thesis show some temporal instability regarding the relative size of institutionalized residents in some age and gender groups. Disentangling substantive changes from changes due to differences in the underlying definitions or methodology could be a task for future research.

Another important open question concerns the judgmental evaluation of the severity of bias in survey estimates. The most straightforward interpretation of bias consists in following statistical conventions: Any distortion which leads to statistically significant differences compared to a reliable benchmark value matters. In fact, those differences might still be small in absolute terms but statistically significant due to large sample sizes and narrow confidence intervals. Looking at effect sizes and putting the bias in relation help to evaluate the severity of bias. In Chapter 3, the true share of respondents without any limitations in activities of daily living is 84.3%. Excluding 2.5% of institutionalized residents causes more than 40% of the random samples to miss this true value with their confidence intervals. This share is well above the 5% error rate as suggested by statistical conventions. Samples missing the true value estimate a share of respondents without any limitations between 85.6% to 87.6%. Would a statistically significant deviation of 3.3 percentage points *at the maximum* call for an extension of coverage? As another example from Chapter 5, the European Social Survey overestimated the turnout among French voters older than 80 years by about 26 percentage points in 2012. This extent of bias definitely requires attention, even though it is only one variable in a dataset of more than 250 variables and this bias only affects a single age group and not the entire sample. In contrast to the well-controlled simulation study, multiple sources of error in addition to the exclusion of institutionalized people could cause this distortion in turnout estimates (cf. DeBell et al. 2020). In addition to judging the severity of bias, the difficulty consists in finding the most important source for bias once a distortion has been detected.

It could be an added value of this doctoral thesis if the definition of the target population will be considered as a potential source of bias.

1.6.2 On the feasibility of an inclusion

In the process of evaluating the severity of bias, the assessment of the feasibility of including missing respondents will presumably play a decisive role for many survey researchers. A small bias could lead to changes if the measures to mitigate it appear to be feasible and not too costly. It is beyond the scope of this doctoral thesis to investigate comprehensively under which circumstances it would be feasible for social surveys to sample, contact, and interview institutionalized respondents. The institutionalized population is a highly heterogeneous group with varying demands and requirements. Simply not excluding institutionalized residents without any methodological adaptations could be considered as inadequate to actually reach those residents (Gaertner et al. 2019). On the other hand, successfully covering prisoners calls for different adaptations than covering nursing homes residents or students living in dorms. For social surveys operating in Europe, a broader inclusion of the institutionalized population *with* related methodological adaptations might not make sense, especially under the constraint of limited resources. As mentioned in the previous section, more thoughts focused on how to include residents of retirement and nursing homes might be more heavily needed.

The following section highlights opportunities for future research related to the feasibility of including institutionalized people. Aspects from both branches of the TSE are covered. The section emphasizes potential spill-overs of methodological adaptations for the institutionalized population to other hard-to-survey populations living in private households.

In all processes described below, the overarching challenge for social survey programs consists in finding the right balance between the standardization of procedures while moving away from a one-size-fits all approach to more flexible and targeted survey designs. Any methodological change should be carefully implemented, ideally within an experimental design. To control for error sources, costs of methods and their impact on the data, responsive designs can be a valuable method (Groves and Heeringa 2006; Tourangeau et al. 2017): While a random group of the sample is contacted or interviewed with traditional methods, other random parts of the sample receive adapted treatments. Adaptive survey designs might be even more relevant in the realm of the present doctoral thesis, because they aim to apply varying methods to specific demographic subgroups in the sample (Tourangeau et al. 2017). For instance, survey researchers could alter their methods for the old-aged population to improve the inclusion of respondents who could potentially drop out due to their incapacity. There is a lot of room for developing innovative new survey methods and test them during fieldwork.

Representation and errors of non-observation

Nearly all countries participating in rounds 5 through 7 of the European Social Survey suffered from underrepresentation of respondents aged between 25 to 34 years, respondents older than 75 years, and respondents from the group of non-nationals (Koch 2018, 12f. cf. Chapter 5).[11] Assuming that this underrepresentation is not caused by coverage or sampling errors, the higher nonresponse among the oldest age group and among non-nationals could be relevant in the context of this doctoral thesis. Those uniform observations indicate problems surveys face with some segments of the population, revealing room for improvements by better adapting to hard-to-survey people living in institutions *and* private households. In nearly all Western countries, the number of elderly people receiving care benefits while living in private households is larger than the respective number of care recipients in institutional settings (Rodrigues et al. 2012, 84). As of now, many of those elderly people are presumably labeled as "incapable" to take part in the survey interview. Spill-over from survey research in institutions could help to raise participation of elderly respondents, especially by increasing flexibility, structuring communication with gatekeepers, or considering the use of proxy interviews.

Survey researchers working on a longitudinal survey among the old-aged population provide a number of solutions to deal with the most common reasons for refusing the further participation in the panel survey (Davies et al. 2014, 9). In a nutshell, they recommend valuing respondents, showing genuine interest in their rationale to refuse participation, and offer flexibility in terms of when and how survey interviews are carried out (ibid.). Being responsive can imply admitting a third person of trust during the interview to assist the respondent, hiring interpreters, or paying for travel expenses for respondents (Kammerer et al. 2019). In addition, Kammerer and colleagues (2019) identified trust, incentives, and identification of individual barriers as important instruments for enabling interviews with the hard-to-reach elderly population.

Thoughts spent on how to convince operators to let interviewers enter institutions might also help to convince gatekeepers in other contexts. This includes authorities limiting the access to the population register, especially in those countries using a decentralized register (cf. Gaertner et al. 2019). It also applies to the rising number of private households in North America and Europe in gated communities or locked apartment buildings with technical or human gatekeepers (Tourangeau 2014). In Round 9 of the European Social Survey, 3,800 sample units could not be interviewed due to the refusal of a proxy within the household (3.8%). Surveys in institutions always plan how to communicate with gatekeepers, develop their arguments and insert those into advance letters for gatekeepers in institutions *and* private households (Neuert et al. 2016).

As another adaptation often used in surveys in institutions, survey programs might consider offering the possibility to run a shorter interview with proxy respondents for those respondents who are not able to participate in the survey interview themselves, either due to language barriers, cognitive impairment, or a bad health condition during the time of fieldwork. Obviously the range of questions that can be posed in a proxy interview is limited (Schanze and Levinson 2019). Proxy

[11] This underrepresentation is independent from the exclusion of institutionalized residents in the ESS, since they were also excluded from the benchmark data used by Koch (2018).

respondents are less reliable when answering questions about the attitudes or the behavior on behalf of respondents. However, depending on the proximity of the proxy respondents with the respondent, they can be a valuable source of information (ibid.). Research on the bias in electoral participation clearly showed a smaller bias when proxy respondents provided information on the turnout of their housemates, probably because social desirability is less influential in those interviews (DeBell et al. 2020). As mentioned in Chapter 4, proxy interviews are a tool to support hard-to-interview respondents and ultimately safeguard the quality of survey data. From a representation perspective, proxy information on sample units can be used to assess bias in a survey if those respondents would be missing entirely.

Concluding with a general observation, previous research has shown that nonresponse rate and nonresponse bias are not perfectly correlated (Brick and Tourangeau 2017; Groves and Peytcheva 2008). Thus, increasing the response rate can help to reduce bias, but does not eliminate it entirely if additional respondents resemble previous respondent. Nonresponse bias, instead of nonresponse rate could be the prevailing quality indicator in future (cf. Kreuter 2013). More attention might be paid to institutionalized residents and other hard-to-survey people as a consequence. At the moment, additional fieldwork effort apparently leads to the recruitment of respondents who do not differ much from previous respondents. To lower bias it might be more promising to focus on hard-to-survey respondents, as they should have a larger marginal impact on the nonresponse bias, even if it probably comes at the price of a lower response rate, e.g., because respondents are not capable to take part (cf. Section 1.4.1). A discussion of trade-offs between standardization and adaptation of survey methods might be a prerequisite to reach the target of lower nonresponse bias.

Measurement and errors of observation

Any effort to extend the coverage by increasing the scope of the target population, improve sampling and reduce nonresponse will also have an impact on the measurement branch of the Total Survey Error framework. The key trade-off for survey researchers consists in increasing participation to alleviate bias without compromising the validity of data due to respondents' insufficient capacity to take part (Sangl et al. 2007). To improve validity, thoughts can be spent on how to adapt survey questionnaires and survey questions, making them easier to understand, easier to answer, or more relevant for subgroups of the target population.

Chapter 4 analyzes respondents' capacity without considering the survey items used by SHARE and how those items influence their response. The literature on measurement error and response behavior provides precise and concrete recommendations on how to design survey items, sometimes also for more challenging respondents, such as residents of retirement or nursing homes. For instance, avoid using long reference periods, avoid using negatively formulated items, think about how many levels scales should have and whether scale points should be labeled, and use routing questions to minimize the response burden (Fox et al. 2007; Sangl et al. 2007). Fox and colleagues (2007) highlight the importance of relevance of survey questions for respondents in the answer process. Commonly asked questions perceived as not relevant for residents of institutions

due to their special living circumstances might decrease the motivation in this group and lead to undesired respondents' behavior such as satisficing (cf. Krosnick 1991). Once the questionnaire is designed, survey methodology already knows various qualitative techniques like cognitive pretesting, focus groups, or pretests and quantitative techniques to test existing or new survey items (cf. Sangl et al. 2007; Yan et al. 2012). A comparison of easy to survey respondents with hard-to-survey respondents would be meaningful, just as analyses of whether a range of survey measures applies similarly in institutions as in private households.

Awareness for the needs and demands of hard-to-survey respondents improves the process of developing the survey. As a good example, the survey *Adults with Learning Difficulties in England* (ALDE) developed their questionnaire by putting together an inclusive research team and integrating members from their actual target population, which required more flexibility in their timeline, but also improved the way the survey and their interviewers communicated with their respondents in the end (Malam et al. 2014). If refugees should be interviewed, cultural awareness is required in all steps of questionnaire development and interviewing. Lyberg (2014, 88) observes that cultural awareness is often stressed by survey programs without defining it further or explaining what follows from it. "*The two main types of challenge are linguistic and socio-demographic/cultural and it is likely that most, if not all, of these challenges would exist in non-refugee populations with the same cultural and linguistic background*" (Stathopoulou et al. 2019, 109). For refugees living in communal accommodations the mental health after potentially living through traumatic experiences are additional layers of complexity (ibid.). This is why questionnaires might require adaptations to meet cultural norms and concepts.

The survey mode used in the data collection is a central element in the survey design, though, the choice of mode(s) does not receive much attention in this doctoral thesis. In recent decades, the prevailing face-to-face survey mode saw declining response rates, increasing fieldwork effort and survey costs, and the emergence of new modes for data collection, namely interviews via telephone or internet (Beullens et al. 2018; De Leeuw et al. 2018; Groves and Heeringa 2006; Kreuter 2013). The trend to combine various survey modes is also relevant in the light of institutionalized residents. Roughly 30% of survey programs listed in the SERISS Survey Inventory combine face-to-face interviewing with other survey modes, namely telephone interviewing (CATI) or self-administered online or paper-based questionnaires in all kinds of institutions. For many hard-to-survey groups and parts of the institutionalized population, the face-to-face mode might still be the best survey mode (cf. Gaertner et al. 2019; Schanze and Levinson 2019). Lyberg and colleagues (2014) emphasized the important role of interviewers on the basis of their literature review of surveys among hard-to-survey populations. It would be interesting to further investigate whether observational errors caused by interviewers (West and Blom 2017) are more prevalent in interviews with hard-to-survey respondents, as those respondents require more support by interviewers (cf. Chapter 4). Moreover, as survey modes are increasingly getting more heterogeneous and tend to move to self-administered modes, institutionalized residents are only one part of a larger group of elderly people potentially having trouble completing questionnaires themselves due to their technological sophistication (Quinn 2010). More research on advantages and disad-

vantages of specific survey modes for surveying institutionalized people, as well as innovations to enable those interviews, are highly welcome.

1.6.3 Fields of substantive research

This doctoral thesis has a strong focus on methodological questions in the field of survey methodology. However, collecting survey data in institutions would ultimately serve for substantive analyses. Those analyses of institutionalized residents are highly interesting for multiple reasons. As mentioned above, institutions serve common interest by their residents and fulfill societal goals (Goffman 1957; United Nations 2008). The self-selection mechanisms of moving into institutions make institutionalized residents interesting for substantive researchers.
Specific types of institutions are relevant for answering research questions in a wide range of scientific disciplines. Numbers of prisoners and prisoners themselves can be interesting for analyses of the judiciary system and how a society deals with deviating behavior and which demographic or socio-economic factors increase the likelihood of being sentenced (cf. Rampey et al. 2016). Prisoners or refugees are potentially relevant for psychological research on how humans cope with situations of crisis and trauma (Blackmore et al. 2020; M. Fazel et al. 2005; S. Fazel and Danesh 2002), while experiences made by refugees also help to shed light on the way a society is dealing with the integration and aspirations of foreigners. Retirement and nursing homes are places where the demographic change will be more and more visible in the next few decades, which is why they allow comparative analyses of the public (health) care system or the sociological rules by which families and the society function (Geerts and Bosch 2012; Laferrère et al. 2012; Rodrigues et al. 2012). Residents or retirement and nursing homes as well as residents in health-care institutions are of interest for medical sciences. It is easy to think of further potential scientific or applied research in institutions. A longitudinal perspective on how the institutionalized population and its composition changed over time would tell us a lot about underlying changes in societies, countries, or policies.

1.7 Conclusion

As the Total Survey Error Framework demonstrates, social surveys are complex systems operating with multiple sources of error and trying to mitigate the influence of those errors with a limited budget. Decisions about the eligibility of institutionalized residents are made within this system. This doctoral thesis introduces the *necessity* and *feasibility* as two overarching questions which need to be answered by survey researchers. Most of the papers included in this doctoral thesis treat questions of necessity, while Chapter 4 also covers aspects of feasibility.

As the first step, survey programs should assess whether the inclusion of institutionalized residents is necessary to avoid coverage bias. This assessment depends heavily on the definition of the target population and the topics covered. It is crucial to estimate the size of the institutionalized population within the target population and its statistical distinctiveness with respect to the topics of interest. As a general conclusion drawn in Chapter 2, social survey programs operating in European countries should be mostly concerned about the relative size of the institutionalized population older than 80 years, unless the survey program is interested in a very specific and narrow target population (e.g., immigrants arriving in a country after 2015 or students). This is the reason why the main focus in this thesis is put on the population living in retirement and nursing homes, despite the more general title. Chapters 2 and 3 show how those elderly institutionalized people differ from their peers living in private households with respect to socio-demographic characteristics, health condition, or physical mobility. Chapter 5 delves into potential differences in political behavior, which appear to be confirmed with administrative data from France (cf. Section 5.8). Even a small share of missing institutionalized residents, e.g., 2.5% in Chapter 3, can have a measurable impact on the distortion of point estimates in a sample for residents older than 50 years. When focusing on a more narrow demographic subgroup, e.g., the elderly population, the impact is even stronger (cf. Chapter 5). Still, survey researchers might come to the conclusion to not cover the institutionalized population after assessing questions of necessity. For instance, Chapter 2 suggests this conclusion for the middle-aged groups between 30 to 70 years in European countries.

If the inclusion is deemed necessary, questions on the feasibility follow. Those questions need to be asked for every step in the survey cycle. The TSE helps to guide survey researchers in this process. Developing answers is heavily influenced (1) by the country the survey program is operating in, e.g., by the sampling frame available or by the availability of qualified interviewers, and (2) by the type of institutions that shall be covered given the strong heterogeneity within the institutionalized population. Both observational and non-observational sources of error need to be taken into account. As a side-note, country-specific particularities are one reason why the inclusion of institutionalized residents is especially challenging - and rare - in cross-national survey programs. As a result of financial or methodological reasons, noncoverage of institutionalized residents is sometimes justified with the fear of covering this part of the population insufficiently. However, as Chapter 3 emphasizes, undercoverage yields better results than noncoverage, especially when post-stratification weights for age and gender correct for undercoverage. If weights are applied, institutionalized groups should be included in the benchmark data used to compile weights. Those can be important lessons drawn from this dissertation for the field of feasibility of representation. Considering the measurement branch of the TSE for the elderly institutionalized population, Chapter 4 confirms the hard-to-interview label for a substantial share of institutionalized people. Still, a majority of institutionalized respondents could be interviewed without many differences in the indicators analyzed in this doctoral thesis. This could be also due to fieldwork instruments SHARE uses, such as proxy interviews. As another important conclusion from this chapter, survey programs need to interview similarly challenging respondents living in private households. This is

a reason why I elaborated on potential spill-overs from surveys with institutionalized people to surveys of hard-to-survey groups living in private households (cf. Section 1.6).

This doctoral thesis generates awareness for specific age groups that are especially critical due to the higher share of institutionalized people, it summarizes topics that should require increased attention, advances ideas how to quantify the peril of bias, and also investigates whether the hard-to-interview label is justified for elderly institutionalized respondents. It does not deal with costs associated with increasing the coverage or conducting interviews in institutions and does not comprehensively discuss and test survey methods used for including or interviewing institutionalized residents. If an existing survey program starts to cover institutions, practices by other survey programs are a valuable resource to develop suitable methods (cf. Section 1.3, Schanze and Levinson 2019). The world of administrative statistics, closely linked to survey research, also tries to improve its knowledge and its methods for covering institutions in the decennial national censuses (UNECE 2020). Changes in survey estimates should be expected when starting to cover institutions. Generally, survey programs with repeated data collections are interested in reducing the impact of methodological changes as much as possible to avoid adding noise or breaks in the time series and disturb interpretations of temporal developments. If institutionalized people are included without further methodological changes between rounds, breaks in the time series can be easily controlled for. This will be more difficult if the methodology is also altered as a consequence of the extension, unless an experimental design is used.

Following the various steps of answering questions of necessity and feasibility, survey researchers working on social surveys could argue that their survey program covers broad topics in the broader population. Any special research interests remain to be investigated by more specialized survey programs. Thus, if a survey program comes to the conclusion to not cover institutions, the easiest way to prevent any biased results consists in stressing repeatedly that the survey does only allow users to draw conclusions for the population living in private households. This strategy comes with three restrictions: Even if they are not covered, institutionalized residents still belong to the general population. A survey focusing on private households risks losing a part of its relevance due to being incomplete (cf. Feskens 2009). Are policy makers interested in health surveys if health-care institutions are excluded? Are data users interested in items on aging if retirement and nursing homes are not covered? Second, even if survey researchers clearly describe their target population in methodological reports and technical papers, users might still not be aware that their analyses might only hold for the population living in private households and not for the general population. Obviously, survey researchers cannot fully control how users will use their data. Third, even sticking to the same definition of a target population of private households could cause bias and breaks in the time series if the underlying target population changes, for instance due to an increasing institutionalization or de-institutionalization of certain parts of the population (cf. Lewinter 2004). This issue is even more severe in surveys comparing multiple countries.

In addition, ethical questions can be asked with respect to the widespread exclusion of institutionalized residents. Does the exclusion in most social surveys, political polls, or even ageing surveys imply that those persons no longer belong to their society? Should they have a right to voice

their opinions and report about their lives, not only in elections or in private, but also in publicly financed surveys? Those are important questions which cannot be finally answered here. Maybe the present thesis and the survey programs and literature cited above and in the following chapters can help survey researchers to come up with evidence-based answers of whether institutionalized residents will have an opportunity to take part in social surveys more often.

To conclude with a general thought, labeling parts of the population as hard-to-survey can be scrutinized as such. Stoop (2014, 225) emphasizes that being hard to reach "*may be related to sociodemographic and socioeconomic characteristics of individuals [...], but also depends greatly on the survey design*", for instance, the appropriate survey mode. Willis and colleagues (2014, 175) suggest "*the need for a subtle but important shift in investigator viewpoint – and perhaps in nomenclature*". Being hard-to-reach "*may mainly reflect the separation between researcher and respondent, rather than some immutable characteristic of the latter*" (ibid.). Thus, in times of decreasing response rates, survey researchers should evaluate their tools and instruments and might come up with innovations for so-called "hard-to-survey" groups that also spill-over to those segments of the populations which are not as hard-to-reach, hard-to-persuade, or hard-to-interview. Institutions offer the perfect training grounds for methodological innovations, even if they are not as gloomy or difficult as depicted by filmmakers, painters, or writers.

References

Alwin, Duane and Jon Krosnick (1991). “The reliability of survey attitude measurement: The Influence of Question and Respondent Attributes”. *Sociological Methods & Research* 20(1): 139–181.

Bacher, Johann, Johannes Lemcke, Andreas Quatember, and Patrick Schmich (2019). “Probability and Nonprobability Sampling: Representative Surveys of hard-to-reach and hard-to-ask populations. Current surveys between the poles of theory and practice”. *Survey Methods: Insights from the Field*. DOI: 10.13094/SMIF-2019-00018.

Bajekal, Madhavi, Levin Wheller, and David Dix (2006). *Estimating residents and staff in communal establishments from the 2001 Census*. Tech. rep. Office for National Statistics: 42–50.

Bergmann, Michael, Annette Scherpenzeel, and Axel Börsch-Supan (2019). *SHARE Wave 7 Methodology: Panel innovations and life histories*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Beukenhorst, Dirkjan, Neville Calleja, Päivikki Koponen, Christian Maggiori, Giulia Milan, Linda Vike, and Albane Gourdol (2011). *European Health Interview survey Task Force IV - report on institutionalised people*. Tech. rep. Eurostat, European Commission.

Beullens, Koen, Geert Loosveldt, Caroline Vandenplas, and Ineke Stoop (2018). “Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts?” *Survey Methods: Insights from the Field*: 1–12. URL: https://surveyinsights.org/?p=9673.

Biemer, Paul P. (2010). "Total survey error: Design, implementation, and evaluation". *Public Opinion Quarterly* 74(5): 817–848.

Blackmore, Rebecca, Jacqueline A. Boyle, Mina Fazel, Sanjeeva Ranasinha, Kylie M. Gray, Grace Fitzgerald, Marie Misso, and Melanie Gibson-Helm (2020). "The prevalence of mental illness in refugees and asylum seekers: A systematic review and meta-analysis". *PLOS Medicine* 17(9): 1–24.

Brick, J. Michael and Roger Tourangeau (2017). "Responsive survey designs for reducing nonresponse bias". *Journal of Official Statistics* 33(3): 735–752.

Brodie, Mollyann, Claudia Deane, Elizabeth Hamel, Melissa Herrmann, and Eran Ben-Porath (2014). "Interviewing in disaster-affected areas: lessons learned from post-Katrina surveys of New Orleans residents". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 157–179.

Brücker, Herbert, Nina Rother, and Jürgen Schupp (2017). *IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen*. Bundesamt für Migration und Flüchtlinge (BAMF): 1–76.

Cebulla, A., M. Daniel, C. Devine, S. Tipping, Megan Daniel, Claire Devine, and Sarah Tipping (2010). *Survey of New Refugees*. Tech. rep. Home Office. UK Border Agency. Analysis, Research and Knowledge Management.

Davies, Karen et al. (2014). "Improving retention of very old participants in longitudinal research: Experiences from the Newcastle 85+ study". *PLoS ONE* 9(10): 1–10.

De Leeuw, Edith, Joop Hox, and Mark Huisman (2003). "Prevention and Treatment of Item Nonresponse". *Journal of Official Statistics* 19(2): 153–176.

De Leeuw, Edith, Joop Hox, and Annemieke Luiten (2018). "International Nonresponse Trends across Countries and Years : An analysis of 36 years of Labour Force Survey data". *Survey Insights: Methods from the Field*: 1–11. URL: https://surveyinsights.org/?p=10452.

De Luca, Giuseppe, Claudio Rossetto, and Frederic Malter (2015). "Sample design and weighting strategies in SHARE Wave 5". In: *SHARE Wave 5: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy: 75–84.

DeBell, Matthew, Jon Krosnick, Katie Gera, David Yeager, and Michael McDonald (2020). "The Turnout Gap in Surveys". *Sociological Methods & Research* 49(4): 1133–1162.

Dormagen, Jean-Yves and Laura Michel (2018). "Do surveys accurately report voters over 80 years old?" *French Politics* 16(2): 195–212.

– (2019). "Probability-based versus quota-based surveys? Selection and self-selection biases persist among the oldest old". *French Politics* 17(1): 45–49.

Eckman, Stephanie and Achim Koch (2019). "Interviewer Involvement in Sample Selection Shapes the Relationship Between Response Rates and Data Quality". *Public Opinion Quarterly*: 1–25.

European Union (2008). *Regulation (EC) No 763/2008 of the European Parliament and the Council of 9 July 2008 on population and housing censuses*. Tech. rep. Official Journal of the European Union: 14–20.

Eurostat (2011). *EU legislation on the 2011 Population and Housing Censuses: Explanatory Notes*. Tech. rep. Luxembourg: European Union.

Fazel, Mina, Jeremy Wheeler, and John Danesh (2005). "Prevalence of serious mental disorder in 7000 refugees resettled in western countries: a systematic review." *Lancet* 365: 1309–1314.

Fazel, Seena and John Danesh (2002). "Serious mental disorder in 23 000 prisoners: A systematic review of 62 surveys". *The Lancet* 359: 545–550.

Feskens, Remco (2009). "Difficult Groups in Survey Research and the Development of Tailormade Approach Strategies". PhD thesis. Universiteit Utrecht.

Fox, Mary, Souraya Diani, and David Streiner (2007). "Using Standardized Survey Items With Older Adults Hospitalized for Chronic Illness". *Research in nursing & health* 30: 468–481.

Gaertner, Beate, Carmen Koschollek, Maike Grube, Denise Lüdtke, Judith Fuchs, Christa Scheidt-Nave, Antje Gößwald, and Matthias Wetzstein (2019). "Including nursing home residents in a general population health survey in Germany". *Survey Insights: Methods from the Field*. DOI: 10.13094/SMIF-2019-00010.

Geerts, Joanna and Karel van den Bosch (2012). "Transitions in formal and informal care utilisation amongst older Europeans: The impact of national contexts". *European Journal of Ageing* 9(1): 27–37.

Glasser, Irene, Eric Hirsch, and Anna Chan (2014). "Reaching and enumerating homeless populations". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 180–200.

Goffman, Erving (1957). *Characteristics of Total Institutions*. Tech. rep. Washington DC, USA: Symposium on Preventive and Social Psychiatry.

Groom, Genevieve, Louise Morris, and Maria Tortoriello (2009). *Communal Establishments Survey - Findings of the Pilot Stage: Summary Report*. Tech. rep. Office for National Statistics (UK): 1–97.

Groves, Robert, Floyd Fowler, Mick Couper, James Lepkowski, Eleanor Singer, and Roger Tourangeau (2009). *Survey Methodology*. Second edi. New York: John Wiley & Sons, Inc.

Groves, Robert and Steven Heeringa (2006). "Responsive design for household surveys: Tools for actively controlling survey errors and costs". *Journal of the Royal Statistical Society. Series A: Statistics in Society* 169(3): 439–457.

Groves, Robert and Lars Lyberg (2010). "Total survey error: Past, present, and future". *Public Opinion Quarterly* 74(5): 849–879.

Groves, Robert and Emilia Peytcheva (2008). "The impact of nonresponse rates on nonresponse bias: A meta-analysis". *Public Opinion Quarterly* 72(2): 167–189.

Kalton, Graham (2009). "Methods for oversampling rare subpopulations in social surveys". *Survey Methodology* 35(2): 125–141.

Kammerer, Kerstin, Katrin Falk, Anna Herzog, and Judith Fuchs (2019). "How to reach 'hard-to-reach' older people for research: The TIBaR model of recruitment". *Survey Methods: Insights from the Field*. URL: https://surveyinsights.org/?p=11822.

Koch, Achim (2018). *Assessment of socio-demographic sample composition in ESS Round 7*. Tech. rep. Mannheim (DE): GESIS - Leibniz Institute for the Social Sciences.

Koch, Achim and Michael Blohm (2009). "Item Nonresponse in the European Social Survey". *Research & Methods* 18(1): 45–65.

Kölln, Ann-Kristin, Yfke Ongena, and Kees Aarts (2018). "The Effects of Sampling Frame Designs on Nonresponse and Coverage Error: Evidence From the Netherlands". *Journal of Survey Statistics and Methodology* 7(3): 422–439.

Kreuter, Frauke (2013). "Facing the Nonresponse Challenge". *Annals of the American Academy of Political and Social Science* 645(1): 23–35.

Krosnick, Jon (1991). "Response strategies for coping with the cognitive demands of attitude measures in surveys". *Applied Cognitive Psychology* 5(3): 213–236.

Kühne, Simon, Jannes Jacobsen, and Martin Kroh (2019). "Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees". *Survey Methods: Insights from the Field*. URL: https://surveyinsights.org/?p=11416.

Kutschar, Patrick and Martin Weichbold (2019). "Interviewing elderly in nursing homes – Respondent and survey characteristics as predictors of item nonresponse". *Survey Methods: Insights from the Field*. DOI: 10.13094/SMIF-2019-00015.

Laferrère, Anne, Aaron Van Den Heede, and Karel Van Den Bosch (2012). "Entry into institutional care: predictors and alternatives". In: *Active ageing and solidarity between generations in Europe*. Ed. by Axel Börsch-Supan, Martina Brandt, Howard Litwin, and Guglielmo Weber: 253–264.

Lewinter, Myra (2004). "Developments in home help for elderly people in Denmark: The changing concept of home and institution". *International Journal of Social Welfare* 13(1): 89–96.

Lohr, Sharon (2010). *Sampling: Design and Analysis*. Boston (MA), USA: Brooks/Cole.

Lugtig, Peter (2014). "Panel attrition: separating stayers, fast attriters, gradual attriters, and lurkers". *Sociological Methods & Research* 43(4): 699–723.

Lyberg, Lars (2012). "Survey Quality". *Survey Methodology* 38(2): 107–130.

Lyberg, Lars, Mathew Stange, Janet Harkness, Peter Mohler, Beth-Ellen Pennel, and Lilli Japec (2014). "A review of quality issues associated with studying hard-to-survey populations". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 82–107.

Malam, Sally, Eric Emerson, and Ian Davies (2014). "Challenges in the first ever national survey of people with intellectual disabilities". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy Johnson, Kirk Wolter, and Nancy Bates. Cambridge University Press: 347–367.

Neuert, Cornelia, Otto Wanda, Angelika Stiegler, Clara Beitz, Robin Schmidt, Katharina Meitinger, and Natalja Menold (2016). "SHARE Gatekeeper-Project: Cognitive Pretest". Mannheim.

Nicaise, Ides and Ingrid Schockaert (2014). "The hard to reach among the poor in Europe: lessons from Eurostat's EU-SILC survey in Belgium". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 541–554.

OECD (2006). *Glossary of statistical terms: Institutional households - Eurostat*. URL: https://stats.oecd.org/glossary/detail.asp?ID=1372 (visited on 08/02/2018).

Pennell, Beth-Ellen, Yashwant Deshmukh, Jennifer Kelley, Patty Maher, James Wagner, and Dan Tomlin (2014). "Disaster research: surveying discplaced populations". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 111–133.

Pickering, Kevin, Patten Smith, Caroline Bryson, and Christine Farmer (2008). *Home Office*.

Poulain, Michel and Anne Herm (2013). "Central Population Registers as a Source of Demographic Statistics in Europe". *Population* 68(2): 183–212.

Quinn, Kelly (2010). "Methodological considerations in surveys of older adults: Technology matters". *Australian Journal of Emerging Technologies and Society* 8(2): 114–133.

Rampey, Bobby, Shelley Keiper, Leyla Mohadjer, Tom Krenzke, Jianzhu Li, Nina Thornton, and Jacqui Hogan (2016). *Highlights from the U.S. PIAAC Survey of Incarcerated Adults: Their Skills, Work Experience, Education, and Training*. Washington DC, USA: U.S. Department of Education.

Rodrigues, Ricardo, Manfred Huber, and Giovanni Lamura (2012). *Facts and Figures on Healthy Ageng and Long-term Care: Europe and North America*. Vienna: European Centre for Social Welfare Policy and Research.

Sangl, Judith et al. (2007). "The Development of a CAHPS® Instrument for Nursing Home Residents (NHCAHPS)". *Journal of Aging & Social Policy* 19(2): 63–82.

Schanze, Jan-Lucas (2017). *Report on sampling practices for the institutionalized population in social*. Tech. rep. Deliverable 2.16 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. URL: https://seriss.eu/resources/deliverables/.

Schanze, Jan-Lucas and Ma'ayan Levinson (2019). *Report on what persons live in institutions and the most relevant institution types they live in, the availability of data to select them and their ability to be reached in practice*. Tech. rep. Deliverable 2.17 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. URL: https://seriss.eu/resources/deliverables/.

Schwarz, Norbert (2007). "Cognitive Aspects of Survey Methodology". *Applied Cognitive Psychology* 21: 277–287.

SHARE ERIC (2020). *English generic questionnaire Wave 7: CAPI coverscreen*. URL: http://www.share-project.org/fileadmin/pdf_documentation/Wave_7_Updates___Documents/SHAREw7_English_Coverscreen_PDF.pdf (visited on 11/17/2020).

Singleton, Nicola, Howard Meltzer, and Rebecca Gatward (1998). *Psychiatric morbidity among prisoners*. Tech. rep. UK Office for National Statistics.

Smith, Tom (2014). "Hard-to-survey populations in comparative perspective". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 21–36.

Stathopoulou, Theoni and Terje Andreas Eikemo (2019). "New Perspectives on the European Refugee Crisis. An Empirical Review. Introduction to the special issue". *Journal of Refugee Studies* 32(1): 1–11.

Stathopoulou, Theoni, Elica Krajčeva, Natalja Menold, and Steve Dept (2019). "Questionnaire Design and Translation for Refugee Populations: Lessons Learned from the REHEAL Study". *Journal of Refugee Studies* 32(1): 105–121.

Stoop, Ineke (2014). "Representing the populations: what general social surveys can learn from surveys among specific groups". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 225–244.

Tourangeau, Roger (2014). "Defining hard-to-survey populations". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy Johnson, Kirk Wolter, and Nancy Bates. Cambridge University Press: 3–20.

Tourangeau, Roger, J. Michael Brick, Sharon Lohr, and Jane Li (2017). "Adaptive and responsive survey designs: a review and assessment". *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180(1): 203–223. URL: http:https://dx.doi.org/10.1111/rssa.12186.

Tourangeau, Roger, Brad Edwards, Timothy Johnson, Kirk Wolter, and Nancy Bates (2014). *Hard-to-Survey Populations*. Cambridge (UK): Cambridge University Press.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski (2000). *The Psychology of Survey Response*. Cambridge University Press.

UNECE (2020). *Recommendations for Measuring Older Populations in Institutions*. ECE/CES/ST. 1. Geneva (CH): United Nations.

United Nations (2008). "Principles Recommendations Population Housing Censuses: Revision 2". New York.

U.S. Census Bureau (2012). "Technical documentation: 2010 Census Summary File 1". Washington DC, USA.

Wagner, Michael, Matthias Kuppler, Christian Rietz, and Roman Kaspar (2018). "Non-response in surveys of very old people". *European Journal of Ageing* 16(2): 249–258.

West, Brady and Annelies Blom (2017). "Explaining interviewer effects: A research synthesis". *Journal of Survey Statistics and Methodology* 5(2): 175–211.

Willis, Gordon, Tom Smith, Salma Shariff-Marco, and Ned English (2014). "Overview of the Special Issue on Surveying the Hard-to-Reach". *Journal of Official Statistics* 30(2): 171–176.

Yan, Ting, Frauke Kreuter, and Roger Tourangeau (2012). "Evaluating survey questions: A comparison of methods". *Journal of Official Statistics* 28(4): 503–529.

Chapter 2

Hard-to-survey and negligible? The institutionalized population in Europe

A shortened and slightly modified version of this chapter was submitted to Survey Research Methods in January 2021

2.1 Introduction

Most social surveys are designed to allow users to draw inferences for the population living in a given territory at a certain point in time. This claim usually ignores the fact that most social surveys exclude certain parts of the population on purpose. In 1991, Schnell described and quantified those excluded groups for the case of Germany: Apart from foreign residents without command of the survey language(s), illiterates, mobile populations, and elites, a large fraction of the excluded group consisted of the institutionalized population (Schnell 1991). Among others, this last group comprises residents living in prisons, student dorms, refugee accommodations, monasteries, health care institutions, or retirement and nursing homes (cf. OECD 2006).
Persons living in institutions can be labeled as hard-to-survey (cf. Tourangeau 2014). Survey researchers with experiences in surveying institutionalized residents confirmed that they are hard-to-reach due to gatekeepers in institutions and hard-to-interview due to health impairment, frailty, or language barriers (Schanze and Levinson 2019, 65, Cambois et al. 2016; Feskens 2009). In addition, their small relative size of less than 1.5% in Europe might not justify additional financial effort to extend the definition of the target population. Thus, matters have not changed dramatically since 1991. A majority of social surveys like the ESS, EVS, GGP, or the German and American social surveys (ALLBUS and GSS) and even some health surveys and aging surveys still exclude institutionalized citizens (Cambois et al. 2016; Schanze and Levinson 2019).
On the other hand, Schanze and Levinson (2019) list more than 150 survey programs which cover

residents living in institutions. Nearly half of those survey programs focus on nursing and retirement homes, like SHARE, the U.S. *Health and Retirement Study* (HRS), or the German Ageing Survey (DEAS). Other survey programs cover prisoners, especially in the USA (e.g., PIAAC). Recently, the increasing number of refugees led established German surveys like the German *Socio-economic Panel* (SOEP) to cover refugees living in refugee accommodations and private households. Another group of survey programs, like the *European Health Interview Survey* (EHIS), *Labour Force Survey* (LFS), or the British UKHLS cover institutions generally without specifying certain types of institutions as part of their target population.

Still, comprehensive empirical knowledge about the institutionalized population is limited in most European countries, even in National Statistical Offices, as a task force of the UNECE (2020) concluded. This paper aims to address to survey researchers and users of both groups of survey programs, notwithstanding whether they currently cover institutionalized residents or not. It uses data from the 2011 European census to provide a quantitative overview of *all* institutionalized citizens in Europe and examines the likelihood of coverage bias if those groups are left out in social surveys. The paper aims to inform decisions about how to deal with the institutionalized population in the phase of defining the target population in social surveys. I only take into account the *necessity* of an inclusion due to the peril of coverage bias and will not consider questions on the *feasibility* of an extension of coverage in the light of the hard-to-survey nature of institutionalized residents. The structure of the paper is as follows: The next section elaborates on the theory of coverage bias and explains why the exclusion of institutionalized residents might increase the peril of coverage bias. The third section describes the underlying data and methodology of the present paper. As census data play a vital role in the analysis of a coverage bias, the paper also mentions potential sources of error within register-based and traditional censuses with respect to collecting data in institutions. The fourth section presents the results of the data analysis. It is split into two parts: The first part quantifies the size of the institutionalized population, while the second part elaborates on its statistical distinctiveness, comparing its basic demographic and socio-economic characteristics to the population living in private households. The paper concludes by discussing the implications of the results for social survey programs.

2.2 Theory and background

Most of the literature on coverage error deals with mismatches between the definition of a target population and the coverage in a sampling frame (Groves et al. 2009; Lessler and Kalsbeek 1992; Lohr 2010). Undercoverage occurs whenever a part of the population is defined as belonging to the target population but is not included in the sampling frame. Contrary to the previous literature, the present paper does not deal with undercoverage as a result of incomplete sampling frames but examines coverage error because of a deliberate exclusion of a specific subgroup of the population.

The following simple formula grasps coverage bias (cf. Groves et al. 2009, 55):

$$\overline{Y}_p - \overline{Y} = \frac{N_i}{N}(\overline{Y}_p - \overline{Y}_i) \tag{2.1}$$

Following the adapted formula, the difference in a mean or any other point estimate drawn from a sample of private households ($\overline{Y}_p$) compared to the respective value in the total population ($\overline{Y}$) is determined by the multiplication of the share of the institutionalized population ($\frac{N_i}{N}$) and additional groups excluded from the survey with the aggregate differences between the excluded institutionalized population ($\overline{Y}_i$) and the included population living in private households ($\overline{Y}_p$). Thus, coverage bias gets stronger with the relative size of the institutionalized population and increasing differences between this population and the population living in private households. Both, the size and statistical distinctiveness are necessary conditions for coverage bias (cf. UN-ECE 2020) and will be considered in the following analysis.

Why could institutionalized residents differ from residents living in private households? Leaving a private household and moving to an institution - voluntarily or under duress - implies a very strong (self-)selection mechanism. Institutionalized residents might be interested in pursuing a higher education, have to work in another place for economic reasons, are too frail or sick to live on their own, have fled their native countries to escape war, or committed a crime and end up being sentenced to prison. The plural of the term *institutionalized populations* used in this paper expresses the large heterogeneity within the population living in institutions.

All institutionalized inhabitants live under the same authority by the institution, which might have an additional transformational influence on behaviors, attitudes, and aspirations by institutionalized residents (cf. Goffman 1957). Even though few of the modern institutions might exert such a strong impact as the total institutions described by Goffman in 1957, the (self-)selection and the transformation still increase the likelihood of differences when comparing institutionalized and community-dwelling citizens. Institutionalized residents usually lose a part of their autonomy when moving to institutions (Schanze and Levinson 2019).

As a consequence of the (self-)selection and the transformation influence of institutions, drawing inferential conclusions for the entire population from a sample of private households might result in coverage bias for some topics. The following paragraphs take a closer look at how institutionalized residents might differ from residents living in private households with respect to demographic and socio-economic characteristics. Before that, I briefly summarize information on the various types of institutions and how age and gender are distributed within those institutions.

2.2.1 Types of institutions

Previous detailed analyses of the institutionalized populations are restricted to single countries, since they were published by National Statistical Offices. In 2011, 1.7% of the total population

lived in institutions in England and Wales (ONS 2015). Within this group, 41% lived in educational institutions, 38% lived in retirement and nursing homes, 7% in prisons, and 5% live in military institutions (ibid.). In Sweden, 1.8% of the population lived in "special housing" in 2015. Those numbers are based on analyses of the population register available upon request to Statistics Sweden. Nearly all institutionalized residents in Sweden either live in student dorms (47.9% of the institutionalized population) or in institutions for elderly and disabled persons (47.7%). In France, the 2009 rolling census estimated a share of 2.5% for the population living in institutions (Pirou et al. 2013). 30% live in retirement homes, 26% live in boarding schools, 19% live in health-related institutions, and 9% live in worker dormitories (ibid.). Comprehensive cross-national figures on the types of institutions are not available from a single source at the European level.
For specific institutionalized groups, such as refugees (Eurostat 2020a) or prisoners (Eurostat 2020b), cross-national numbers are available. The European statistics recorded 320,000 people applying for asylum in 30 European countries in 2011 (Eurostat 2020a). Until 2016, this number has quadrupled as a consequence of various international armed conflicts (ibid.). Information about the types of housing is not available from this source, but it can be assumed that oftentimes refugees live in refugee accommodations shortly after arriving in Europe. Prisoners certainly belong to the institutionalized population. A lower proportion of the European population lives in prisons than in many other countries worldwide (Eurostat 2020b; Walmsley 2018). In Europe, 450,000 people were in prison in 2011, roughly 10% less than in 2019 (Eurostat 2020b). The highest rate of prisoners was reported in Baltic countries, Poland, Czechia, and Slovakia with more than 200 prisoners per 100,000 people, while the European median rate is 81 prisoners per 100,000 residents (Eurostat 2020b; Walmsley 2018, cf. Schanze and Levinson 2019, 18). Institutions for elderly people focus on the provision of care and medical services. Rodrigues and colleagues (2012) present comparative numbers on the distribution of care for the population older than 65 years in Europe and North America. According to those numbers, roughly 5% of the population in Northern and Western European countries receive institutional care, while this proportion is smaller in Southern and Eastern Europe. Despite the population aging in many European countries, the share of residents receiving institutionalized care remained stable since the mid-1990s due to a large increase of residents receiving home care (ibid., p. 88).

2.2.2 Age and gender distribution

Age and gender are decisive in distinguishing types of institutions, mainly because they are highly correlated with other explanatory factors for institutionalization. A number of studies found a clear impact of higher age on the probability of moving to a retirement or nursing home in Europe (Angelini and Laferrère 2012; Einio et al. 2012; Geerts and Bosch 2012; Laferrère et al. 2012; Luppa et al. 2010; Martikainen et al. 2009; McCann et al. 2012; UNECE 2020). The share of institutionalized residents within the population older than 85 years reaches more than 15% in a lot of European regions (Eurostat 2015, 148). In France, the average age in retirement homes

is 84 years, while residents living in religious institutions are 66 years on average (Pirou et al. 2013). The mean age of residents in French boarding schools (18 years), student dorms (22 years) and military barracks (26 years) is much younger. In between, prisoners are 34 years on average, workers living in workers dormitories are 45 years on average (ibid.). This age pattern is reflected in equivalent numbers from England and Wales (ONS 2015). When looking at the gender distribution in institutions in France and England and Wales, it is nearly perfectly balanced, ranging from 49% of female institutionalized residents in France to 51% in England and Wales (ONS 2015; Pirou et al. 2013). The aggregate level is hiding differences in gender distributions across different types of institutions: According to the 2009 French census, worker dormitories are predominantly inhabited by men, only 15% of all residents are female, just as in military institutions (12%), and prisons (4%), whereas a majority of residents in religious institutions (71%) and in retirement and nursing homes (74%) are female (Pirou et al. 2013; Rodrigues et al. 2012; Walmsley 2017). Those gender distributions bear a great resemblance with distributions in England and Wales (ONS 2015). In boarding schools, other social and health-related institutions, and student dorms the gender distribution is nearly balanced in France, England and Wales. I expect this age and gender distribution to be mirrored in most other European countries.

2.2.3 The distinctiveness of institutionalized residents

In this paper, I examine several demographic and socio-economic variables that are expected to differ between private households and institutions. Table 2.1 lists different types of institutions. It also advances a number of hypotheses on the expected predominant characteristics of the population living in different types of institutions. Those stylized facts are based on previous literature and my own expectations. If cells are empty, it is not possible to postulate any *predominant* characteristic with the given information. This could be due to inconclusive empirical evidence (e.g., level of education in institutions for elderly persons) or due to absent empirical knowledge (e.g., citizenship in religious institutions).

Marital status was identified as a predictor of institutionalization for the elderly population. Widows have a higher likelihood of moving to institutions for elderly persons (Angelini and Laferrère 2012; Einio et al. 2012), similar to persons who become or are living on their own without family (Geerts and Bosch 2012; Laferrère et al. 2012; Luppa et al. 2010; Martikainen et al. 2009; McCann et al. 2012; Pimouguet et al. 2016), e.g., because they are divorced or have never been married. Those people are less likely to receive informal care at home, which is why they are more often living in retirement and nursing homes. The marital status in other types of institutions is much harder to predict. It can be assumed that residents in religious institutions are mostly never married, while the same holds true for young adolescents living in educational institutions and military barracks. Presumably differences in marital status between private households and institutions are small in young age groups.

Considering the citizenship of institutionalized residents in various types of institutions, it is save

to expect a high degree of non-EU citizenship and stateless persons in refugee accommodations, while a large majority of residents living in military institutions probably possesses citizenship by their respective country. The Office for National Statistics (2015, 22) reports a higher diversity of citizenship in educational institutions than in retirement homes in England and Wales. For the USA, studies found a higher likelihood of living in a retirement or nursing home for people with a white ethnicity (Castora-Binkley et al. 2014; Luppa et al. 2010; Thomeer et al. 2015). To my knowledge a similar finding was not yet made for a European country, which is why the cell for institutions for the elderly is left empty.

Table 2.1: Stylized facts about the institutionalized populations

Type of institution	**Age**	**Gender**	**Marital status**	**Citizenship**	**Education**	**Economic activity**
Educational institutions	Young (10-30 yrs.)	Mixed	Never married		Medium/ high	Inactive
Military institutions	Young (20-40 yrs.)	Men	Never married	Nationals		Employed
Prison	Young/ middle-aged	Men			Low/ medium	
Refugee accommodation	Young/ middle-aged	Mixed		Non-European/ stateless		Inactive
Worker's dormitories	Middle-aged	Men			Low	Employed
Health care institutions	Middle-aged/old	Mixed				
Religious institution	Middle-aged/old	Mixed	Never married		High	Employed
Institutions for elderly persons	Old (70+)	Mixed (more women)	Widowed, never married			Inactive

Turning to education, studies on the institutionalization of elderly residents found a negative impact of higher education on the probability of moving to a retirement or nursing home in Finland (Einio et al. 2012), while two other studies using survey data from the USA came to the opposite conclusion (Castora-Binkley et al. 2014; Thomeer et al. 2015). Cross-national differences in the distribution of education across types of housing were also found in another study: A larger share of institutionalized residents between the age of 50 to 69 had a lower level of education than their counterparts living in private households (UNECE 2020). In contrast, the differences across types of housing disappeared for the elderly population (ibid.). As far as educational institutions are concerned, undergraduate or graduate students living in student dorms will have a post-secondary or tertiary level of education, presumably higher than their community-dwelling counterparts on average. Prisoners will potentially have a lower level of education than the average population.
Socio-economic factors like income or house ownership are important explanatory variables in a number of European studies on the institutionalization of elderly residents (Angelini and Laferrère 2012; Einio et al. 2012; Luppa et al. 2010; Martikainen et al. 2009; McCann et al. 2012).

In this paper, the socio-economic status of residents is only captured by the formal level of education and the economic activity. The last column of Table 2.1 deals with the economic activity of institutionalized populations. The 2010 U.S. census distinguished between "institutional group quarters" and "noninstitutional group quarters" with the residents in noninstitutional group quarters being "eligible, able, or likely to participate in the labor force while residents" (U.S. Census Bureau 2012, B-16). Residents living in student dorms, refugee accommodations, and retirement and nursing homes are usually economically inactive, whereas residents living in military institutions, monasteries, and worker dormitories are economically active and belong to the labor force. The economic activity is more uncertain for residents living in health care institutions and prisons. Both groups might be either economically active (e.g., if they are only certified sick for a certain period or employed in the prison), unemployed or economically inactive. The latter is true if institutionalized residents in health care institutions are not able to work due to their health condition or if prisoners are not allowed to work in prison or are not counted as active members of the labor force irrespective of prison labor. This is why the cells for economic activity in health care institutions and prisons are left blank in Table 2.1.
I do not postulate any expectations regarding the size of locality. For residential mobility, it can be assumed that institutionalized residents generally have a higher mobility than community-dwelling residents, since they at least changed their usual place of residence when moving to the institution. The shorter the average duration of their stay in the institution, the higher will be the reported mobility in the census and the more difficulties will potentially occur during the sampling for a survey.

2.3 Data and analysis

Calculating the effect of undercoverage or non-coverage on statistical estimates is impossible with most social survey data, because institutionalized residents are excluded from the sample and relevant information are missing by definition. This paper uses administrative census data from 30 European countries to learn more about the population living in institutions. In addition to the 27 EU member states, the following analyses also convey information on Iceland, Norway, and the United Kingdom. This section describes how various National Statistical Offices (NSO) collected data in institutions in 2011 and which variables will be used in this paper.
The European Union aims to publish comparable data from national censuses every 10 years. A regulation by the European Parliament and the Council of the EU provided definitions and topics to be covered in the 2011 European census (Eurostat 2011). Apart from the operationalization of variables and certain guidelines on the transmission of data and metadata, the EU applied an output harmonization in the 2011 cross-national census. Member states themselves were required to apply the most appropriate census methods (ibid.). This leads to three different groups of countries in Europe: A first group uses the population register as well as additional registers to count the population and collect all the relevant data, a second group runs a traditional census and sends

interviewers to each and every household, and a third group of countries combines the two approaches (Poulain and Herm 2013, 197, Valente 2010, 3). The Scandinavian countries and Austria belong to the first group and fully relied on their population registers and other administrative sources for their 2011 census. Many South-Eastern European countries, Portugal, Ireland, the UK, and Cyprus ran a traditional census, contacting every single household and letting them reply to the census form. A last group of countries consisting of Southern European countries, the Baltic countries, Germany, Belgium, and the Netherlands combined traditional census methods with their registers to a greater or lesser extent. In Germany, private households were counted on the basis of the population register, which was combined with a very large survey of approximately 10% of the population. Interestingly, the population living in German institutions was covered with a traditional census using full coverage (Geiger and Styhler 2012).

When considering the representation branch of the Total Survey Error Framework (Groves et al. 2009, 42), census data suffers less from coverage error than any social survey. This is due to the fact that censuses are commissioned by the governments and National Statistical Offices and have access to the best administrative sources. Censuses operate with very large sample sizes; traditionally they even cover every single member of the target population. Usual survey samples with 1,000 to 5,000 respondents are not sufficient to allow separate analyses of this subgroup, unless oversampling is used. Moreover, participation in censuses is mandatory in contrast to a voluntary participation in social surveys.

Nevertheless, census data are still prone to some errors, especially in hard-to-survey populations (Abbott and Compton 2014; Mulry 2014). Institutionalization implies being in an atypical living situation (Sweet and Alberti 1994, 324): College students living in a student dorm, prisoners, soldiers, or spouses who live in a retirement home risk being not counted or double counted. For instance, the instructions of the Norwegian population register specify that *"spouses should normally be registered at the same address in the [register] even if one of them actually is living in an institution such as homes for the elderly. Persons who are not married are normally registered as residents in the institution"* (Andersen and Utne 2011, 5). As a consequence, this part of the population is probably underestimated in a census fully based on registers. The higher residential mobility of institutionalized residents leading to outdated entries in the register motivated the German Statistical Office to conduct a traditional census in institutions in 2011 (Geiger and Styhler 2012). Proxy replies can be another source of inconsistencies and error. In the 2001 English and Welsh census, residents of retirement and nursing homes were falsely recorded as staff of their institutions, because some employees filled in their own position when answering the census form on behalf of residents (Bajekal et al. 2006).

Thus, even a census with its full coverage of the population might have problems to cover all groups equally. As another limitation, the census only collects hard-facts every ten years. The amount of information available is limited to demographic and socio-economic facts. Any information about attitudes or behaviors of citizens is missing, which is why they are missing in the present paper. Still, basic variables like age, gender, marital status, or education are important structural variables and very often used as explanatory variables in many empirical studies on a

wide variety of topics. Any empirical differences for those variables will most probably have an impact on bias in other attitudinal or behavioral variables. Regarding the institutionalized populations, it is very important to mention that Eurostat does not publish information on the types of institutions. Those information are only available at the national level for some countries (e.g., ONS 2015; Pirou et al. 2013).
When comparing census data on institutions from various countries, differences in the definition of institutions are another potential source of errors. In a recent working paper published by the UNECE (2020) the difficulty to distinguish various types of institutions across countries is highlighted. In the present paper, the difficulties rather arise from blurred lines between institutions and private households. This is especially the case for the elderly population, whenever new forms of assisted living facilities replace traditional nursing homes in a number of mostly Northern European countries (Lewinter 2004; UNECE 2020).

2.3.1 Variables used in the analysis

In a so-called census hub, Eurostat offers a number of variables and their cross-tabulations (hypercubes) that can be used to examine institutionalized populations and private households cross-nationally (Eurostat 2016). This paper compares national data on the two types of housing with respect to age, gender, marital status, nationality, education, and economic activity. Moreover, it considers the size of localities where residents live (rural versus urban) and the residential mobility of citizens, notably whether citizens lived somewhere else one year before the census took place.
In the 2011 European census, participating countries were required to apply the "housekeeping concept" to identify private households (Eurostat 2011, 92): According to this concept, private households can accommodate a single person living in a separate room or rooms without being connected to other potential occupants of the housing unit or a group of related persons who *"provide themselves with food and possibly other essentials for living"* and who *"their incomes to a greater or lesser extent"*. Institutions or collective housing are not explicitly defined by the European census. Following from the definition of private households, institutions host large groups of mostly unrelated inhabitants without any economic connection behind a single door. Institutionalized residents cannot fully care for their own housekeeping, which is why operators of institutions provide essentials for living, meals, medical services or assistance with activities of daily living (Schanze and Levinson 2019; U.S. Census Bureau 2012). Moreover, short-term residents of institutions, like inpatients in hospitals or persons in pretrial imprisonment are excluded from the institutionalized population. The European census applied a temporal qualifier, specifying that an *"institution shall be taken as the place of usual residence of all its residents who at the time of the census have spent, or are likely to spend, 12 months or more living there"* (Eurostat 2011, 66).
The published census data contains two variables on the type of housing. A first variable is called *household status* and contains information on "persons in an institutional household", while a second variable named *housing arrangements* provides information on residents living in a "col-

lective living quarter". In 15 out of 30 European countries, the number of people not living in private households is exactly the same in the two variables. In additional 6 countries, the numbers are very close with a difference of less than 0.1 percentage points in the ratio of institutionalized residents to the total population. Those first two groups add up to 70% of all 30 countries analyzed in this paper. In another 8 countries, larger differences between the two variables can be observed (27%), 5 of those countries show strong differences larger than 0.5 percentage points. Finland did not report any numbers for the variable measuring household status. In the following analyses, I mostly rely on the variable measuring housing arrangements, because the share of citizens not living in a private household as measured by household arrangements is usually larger than the share as measured by household status, except in Iceland and Greece. For analyses on the marital status and education, I used the second variable, household status, because hypercubes for those variables were not available in combination with housing arrangements.

Table 2.2: Operationalization of variables used in the empirical analysis

Variable	**Operationalization**
Gender	Male; Female
Age	(1) Five-year age brackets (from 0-5 until 95-100+) (2) 6 broader age groups (<15; 15-29; 30-49; 50-64; 65-84; >85)
Marital status	Never married and never in a registered partnership; Married or in a registered partnership; Divorced or registered partnership legally dissolved; Widowed or registered partnership ended with the death of partner
Citizenship	Citizenship of reporting country; Citizenship by another EU member state; Citizenship by another country outside the EU; Stateless
Education (ISCED-1997)	Low education (ISCED 0-2); Medium education (ISCED 3 and 4); High education (ISCED 5 and 6)
Economic activity	Employed; Unemployed; Currently not economically active
Size of locality	Villages (<1,999 inhabitants); Very small towns (2,000-5,000 inh.); Small towns (5,000-20,000 inh.); Middle-sized towns (20,000-50,000 inh.); Large towns (50,000-200,000 inh.); Metropolises (>200,000 inh.)
Residential mobility	Usual residence did not change in the previous year; Moved within the reporting country; Moved from outside the reporting country

This paper uses further variables provided by the census hub (Eurostat 2016). The operationalization of those variables is summarized in Table 2.2. For citizenship, the EU membership is applied as of 2011.[1] For marital status, citizenship, education, economic activity and changes of residence, missing values are more prevalent in some European countries ("no information stated"). As a matter of transparency, missing values were not suppressed from graphs or analyses of dissimilarity. In cases of excessive missing values, single countries were dropped. Those decisions are always conveyed in the description of results.

[1] At this point, Croatia was not EU member state yet, while the United Kingdom was still part of the European Union.

2.3.2 Course of analysis

The analysis starts by providing an overview of the share of the population living in institutions, taking into account age and gender of residents not living in private households. This overview is a revision and extension of work presented in an earlier working paper (Schanze and Levinson 2019). It helps to determine which countries, age and gender cohorts are especially critical regarding the quantitative size of the institutionalized populations. Age and gender are essential variables in the light of missing aggregate data on the types of institutions in European countries. The second part of the analysis examines marital status, citizenship, education, and economic activity. Eurostat reports census data for most of the European countries, allowing a comparison of those variables for the types of housing, age, and gender. As first step, I compare the distributions of the variables within the population living in private households and in institutions in different age and gender groups. This analysis does not take into account the relative size of the institutionalized population. It tells us something about the second part of the bias formula, the institutionalized population and its statistical distinctiveness. As second step, Duncan's Index of Dissimilarity is calculated for each variable to get a country-specific overview about the joint impact of relative size and statistical distinctiveness. This index was developed by Duncan and Duncan (1955) and provides aggregate information on a deviation of proportions within a variable between two groups. In this paper, the following formula is used to calculate Duncan's Index of Dissimilarity:

$$D = \frac{1}{2} \sum_{c=1}^{n} |p_{pc} - p_c| \tag{2.2}$$

For each variable, $|p_{pc} - p_c|$ gives the absolute difference between the proportion within the population living in private households p_p and the respective proportion within the total population p for every category (c) of this variable. The absolute differences are added up and divided by 2 for each variable. This calculation is conducted separately for the six age groups within the 30 European countries and the European aggregate. The index of dissimilarity is non-directional, meaning it does not provide any information about whether a specific category of variable is over- or underestimated if institutionalized residents are not covered. The scale of the resulting index reaches from 0 to 100, with 0 indicating the absence of dissimilarity and 100 indicating complete dissimilarity. The resulting numbers indicate what percentage of community-dwelling residents would need to switch categories to yield perfect congruence with the total population. Finally, mosaic plots are used as another tool to visualize the peril of bias. The Appendix of this paper contains tables showing the distributions by age, gender, and type of housing for each variable at the European level.

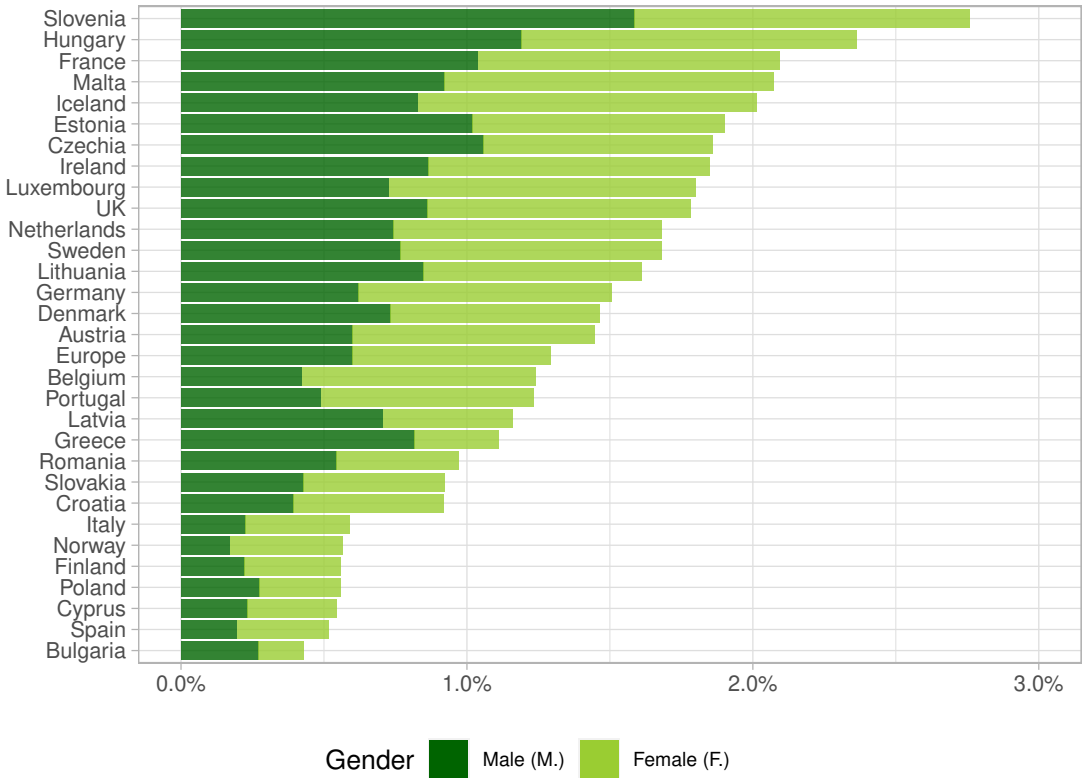
2.4 Results

The first part of the following section presents information on the quantitative size of the institutionalized population, while the second section presents additional insights into characteristics of the institutionalized populations in European countries and the likelihood of bias if they are excluded from social surveys.

2.4.1 Relative size

In the 2011 European census, about 1.3% of the European population did not live in a private household but in an institution. This population amounts to 6.6 million people, a group larger than the respective numbers of citizens living in 13 European countries, among those Lithuania, Ireland, Norway, or Denmark. Figure 2.1 shows the share of the population living in institutions within countries and the gender distribution within those groups. Within the aggregate institutionalized European population, 46% are male.

Figure 2.1: Share of the population living in institutions in Europe



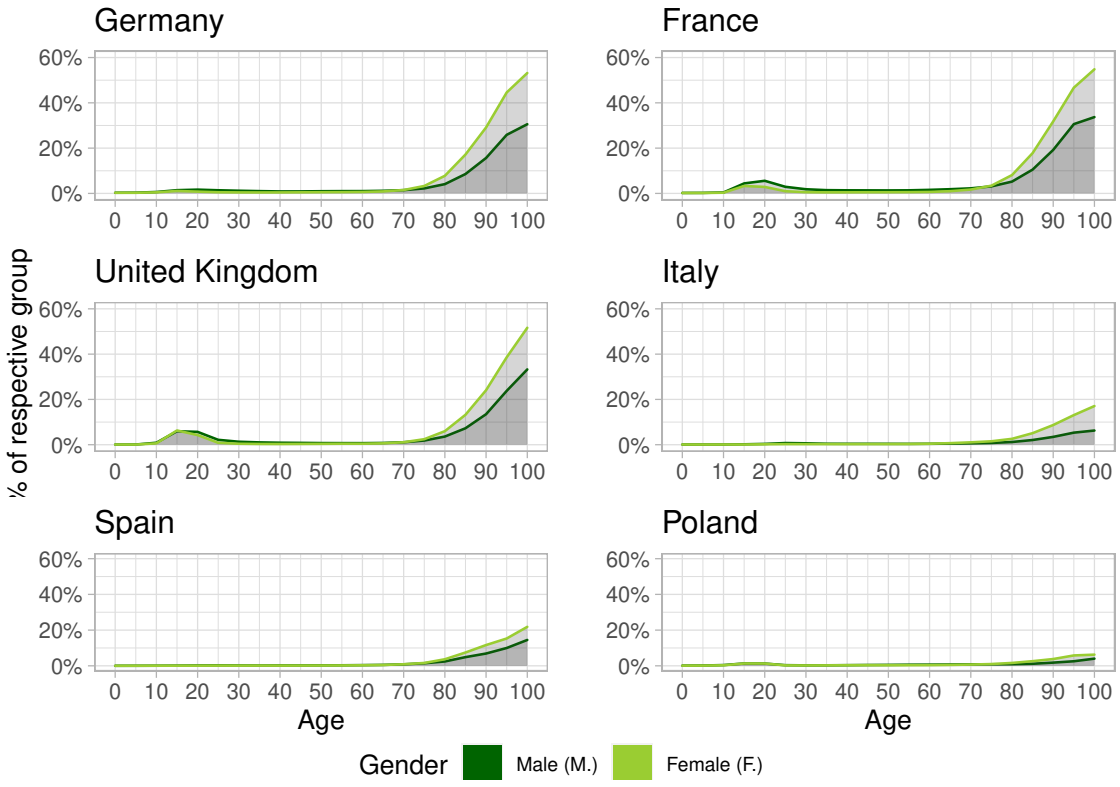
At the national level, roughly two thirds of all European countries reported a share of institutions above 1%. A small number of countries counted a very small share of less than 0.6% of the entire

population. This group of countries located in various parts of Europe as well as the remaining list of countries presented in Figure 2.1 allow the conclusion that there is no clear geographical pattern, such as a "*North-South or West-East divide*" (Schanze 2017, 9).

The nearly equal gender distribution within the institutionalized population in Europe is replicated in most European countries. In two thirds of all countries, the share of female citizens living in institutions ranges between 42.5% to 58.9%. In most countries, the share of institutionalized women is higher than the share of institutionalized men. In Belgium and Norway, more than 65% of the population in institutions are women, whereas in Greece only 26.6% of this population is female. It can be assumed that the gender distribution reflects distinct types of institutions used to a varying degree in European countries.

Figure 2.2 shows the share of institutionalized populations within the respective age and gender groups in the 6 most populated European countries. The proportion is calculated within 21 5-year intervals. Table 2.3 in the Appendix provides similar numbers for broader age groups for all 30 countries and Europe.

Figure 2.2: Share of the population living in institutions within respective age and gender groups



Nearly all European countries follow one of the patterns shown in Figure 2.2. Two main groups of countries can be distinguished. In the first group, the share of the institutionalized population is very minor in the young and middle-aged age and gender groups, reaching not more than 2%. The share steadily increases among the population aged 70 to 79 years and reaches its maximum

in the oldest age groups. Germany, Italy, and Spain are three examples, Austria, Belgium, Croatia, Cyprus, Finland, Iceland, Ireland, Norway, Portugal, and Slovakia are also part of this first group of countries labeled as *one peak countries*.

The pattern in the second group of countries looks very similar for the middle-aged and elderly population with one important difference in younger age groups: A second, smaller peak can be observed for the population between 10 to 29 years. In Figure 2.2, France and the United Kingdom are two examples of *two peaks countries*; Czechia, Denmark, Estonia, Hungary, Luxembourg, Malta, the Netherlands, Slovenia, and Sweden also belong to this group. Based on earlier findings and expectations presented above, it can be assumed that those younger institutionalized age groups live predominantly in educational institutions, like student dorms or boarding schools, or in military barracks and refugee accommodations (cf. Schanze and Levinson 2019).

As a third, residual group of countries, Bulgaria, Greece, Latvia, Lithuania, Poland, and Romania do not follow either of the two patterns described above (cf. Table 2.3). In Greece and Romania, the share of the population living in institutions is larger in the younger age groups than in the older age groups. The overall share of the institutionalized Latvian population is not much smaller than the European average. However, the two peaks in the young and oldest age groups are not very distinct. The distribution looks similar in Lithuania, with a slightly higher share of institutionalized young adults and less institutionalized elderly residents than in Latvia. In Poland, an even smaller fraction of the population lives in institutions (cf. Figure 2.2). The highest share of the institutionalized population is reached in the female population older than 100 years (6.3%). All other age and gender groups live predominantly in private households in Poland. The distribution in Bulgaria resembles the Polish pattern.

Figure 2.2 also allows a comparison of gender distributions. Schanze and Levinson (2019, 16) pointed out that *"[e]ven more than for age, the gender distribution bears a great resemblance across the European countries"*. From the birth to the age of 69 years, the share of institutionalized citizens among women is equal or lower than the respective share among men in almost all age groups in Europe. Only in 4.3% of those age groups a higher share of women is institutionalized in 30 European countries. During childhood up to the age of 15, many countries record equal proportions among the two genders. For adolescents and young adults (15 to 29 years), the differences in proportions get a little bigger across genders. Usually a higher share of men is institutionalized in this age group. Considering the middle-aged age groups between 30 to 60 years, not a single country recorded a higher share of institutionalized citizens among women than among men. Only Norway and Poland reported equal shares for most of the middle-aged groups. In Europe, approximately 0.8% of the men are institutionalized between the age of 35 to 60, while this share is 0.4 percentage points lower among women. This widespread gender pattern is consistent with earlier findings from France and the UK: Institutions addressed to young and middle-aged groups, such as military barracks, prisons, or worker's dormitories, are usually inhabited by men rather than by women (cf. Table 2.1).

The pattern of higher proportions of institutionalized residents among men is completely reversed for the population older than 70 years. In 91.4% of all European age groups, the share of insti-

tutionalized citizens among women is higher than the respective share among men. Above the age of 80, only Bulgaria reported a higher share for men in a single age group. In Europe, 15.4% of women older than 85 years are institutionalized, whereas this fraction is half as large for men (7.9%). As can be seen from the country plots presented in Figure 2.2, the share of institutionalized residents constantly increases with age, as does the difference between the two genders. Again, the uniform gender distribution in old age groups corroborates numbers form France and the UK for other European countries: More women than men live in retirement and nursing homes. To summarize the distribution of relative size, institutions for elderly people are most relevant for a potential bias following the exclusion of institutions in surveys.

2.4.2 Statistical distinctiveness

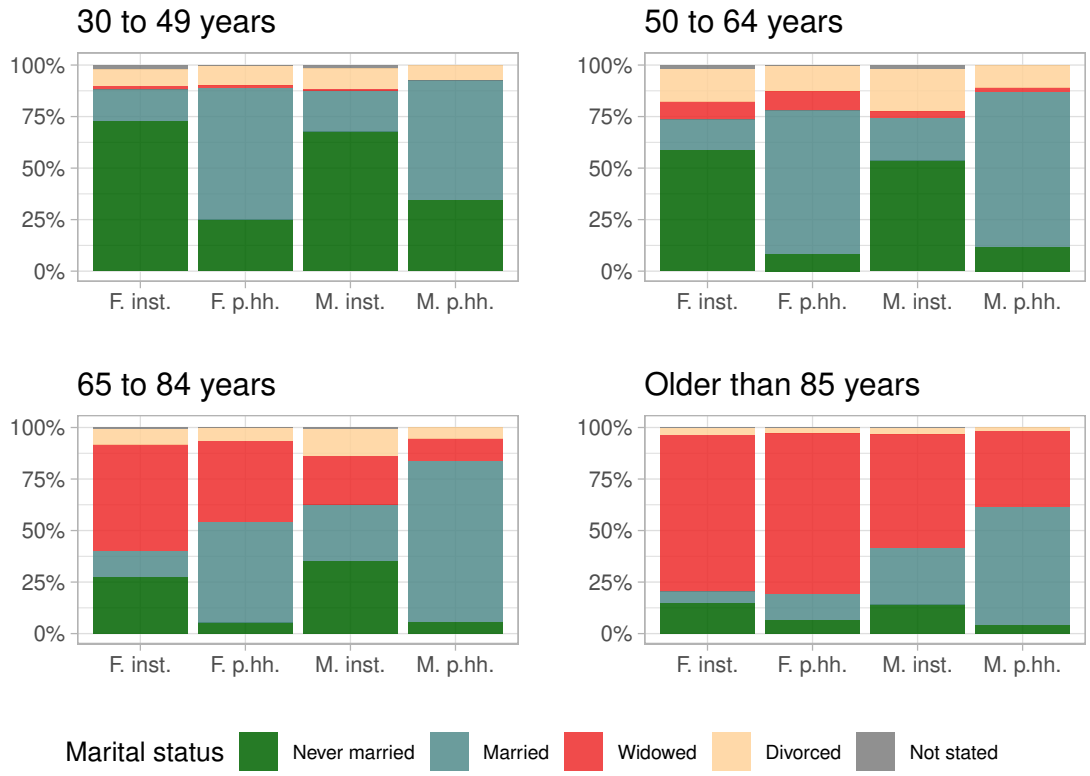
This section considers the second element for determining the peril coverage bias, namely demographic and socio-economic differences between institutionalized and community-dwelling residents that might cause bias in samples limited to private households.

Marital status

Figure 2.3 shows the distribution of marital status within age and gender groups in Europe, comparing institutions and private households. The first two age groups up to the age of 30 are not shown in this plot, because the variance in those groups is very small: Nearly all citizens are classified as never married in those age groups, except for 15.5% of community-dwelling citizens being married between the age of 15 to 30 (cf. Table 2.4 in the Appendix).

As can be seen in Figure 2.3 for Europe, a larger share of institutionalized men and women has never been married than their counterparts living in private households up to the age of 64. In addition, a larger share of institutionalized men and women is divorced in this age group compared to the equivalent population living in private households. For the age group from 65 to 84 years, the share of institutionalized residents who have never been married is smaller. A stronger difference between the two genders emerge in this group: Women aged between 65 to 84 years are more often widowed than men in this group, while institutionalized residents are more often widowed than residents in private households for both genders. For the population older than 85 years, European men living in institutions are more often widowed than their community-dwelling counterparts, while the share of widows for female citizens is nearly equal across the types of housing (see Table 2.4). Still, a larger share of institutionalized populations in the oldest age group has never been married than the respective share in private households. Those results are in line with the assumptions made in Table 2.1 regarding the marital status of residents living in retirement and nursing homes.

Figure 2.3: Marital status in European countries within the institutionalized and private household populations



The overwhelming majority of large and small European countries closely follows the patterns presented in Figure 2.3 for the European aggregate. Stronger deviations can be observed in Bulgaria and Croatia, where the census did not report any widows in some age and gender groups. In Italy and Malta, divorces are very rare compared to all other European countries. The share of elderly widowed women living in institutions is clearly lower than the respective share among community-dwelling women in those two countries (cf. Figure 2.5).

The differences with respect to marital status across the two types of housing make bias more likely when institutionalized residents are not considered in social surveys. However, the relative size of institutionalized populations needs to be taken into account. As shown above, the proportion of institutionalized residents is smaller in middle-aged groups than in the oldest age groups.[2] Duncan's Index of Dissimilarity for marital status presented in Figure 2.4 confirms this assumption. The dissimilarity indexes are rather small for the younger age groups, ranging between 0% to 1% in all countries. The index and the variance across countries only increase for the two oldest age groups. In Iceland, with its large share of institutionalized elderly well above 20% (cf. Table 2.3), a little more than 7% of the community-dwelling population older than 85 years would need to change categories of marital status to yield statistical equality between private households and

[2]Estonia, Germany, Lithuania, and Slovenia did not report any information on the marital status of a relevant fraction of their institutionalized population. This could potentially lead to an underestimation of the dissimilarity index due to a lower relative size of the institutionalized population.

the total population. Belgium, France, Ireland, Luxembourg, and Malta are further countries with a comparably large dissimilarity index in this age group. The European aggregate shown in Figure 2.3 reaches a dissimilarity index of 0.2% to 0.3% in the age groups between 15 to 64 years. For the subsequent age group between 65 to 84 years, the dissimilarity index for Europe rises to 0.8%, and finally reaches its maximum in the oldest age group (2.2%).

Figure 2.4: Duncan's Index of Dissimilarity comparing the marital status of all citizens in private households to the total population

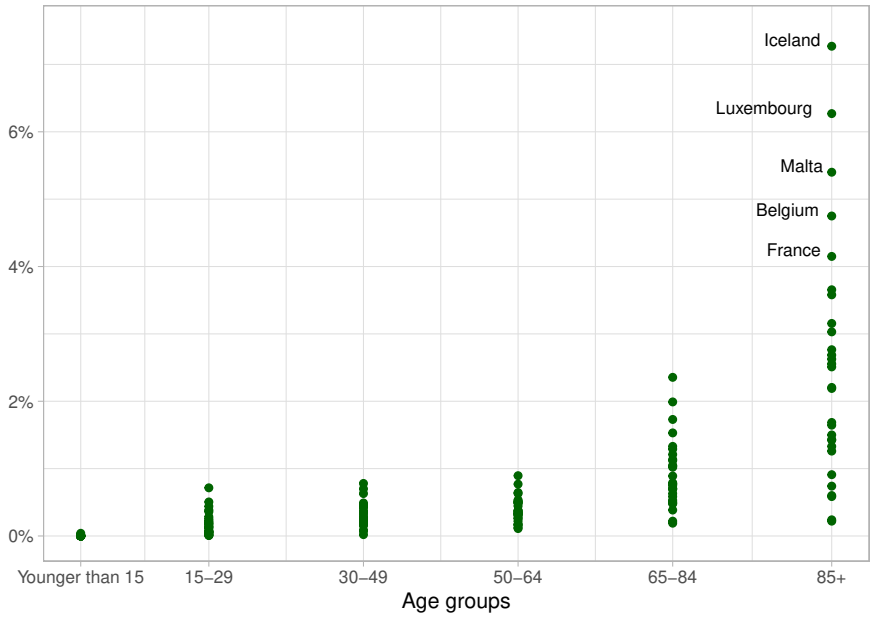
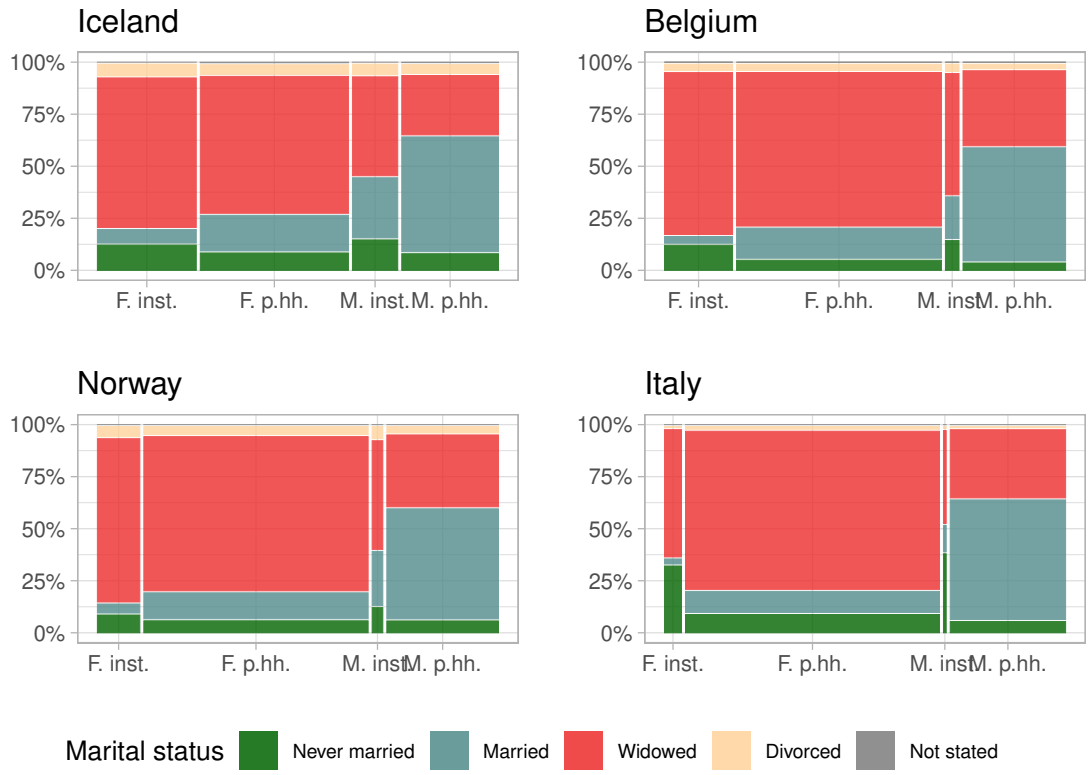


Figure 2.5 visualizes the peril of bias by showing the distribution of marital status while adjusting the size of the bars to the relative size of the specific subgroup. On the basis of Figure 2.4, four country examples for the oldest age group are shown. Iceland and Belgium have a comparably large Duncan Index of Dissimilarity, 7.3% and 4.8% respectively. In contrast, Norway (2.5%) and Italy (1.3%) have small indexes of dissimilarity for marital status in this age group.
In most countries, the distribution of marital status for the population older than 85 years resembles the European aggregate (cf. Figure 2.3). The cross-country comparison underlines the importance of considering statistical distinctiveness *and* relative size. While the statistical distinctiveness does not differ much across countries, the share of institutionalized resident helps to explain differences in Duncan's Index of Dissimilarity (cf. Figure 2.4). The statistical distinctiveness of the institutionalized population is a little more pronounced in Italy than in Iceland or Belgium. In Italy, a larger share of residents of both genders living in institutions has never been married. This might lead to the general hypothesis of a stronger statistical distinctiveness of the institutionalized population in countries with a small population living in institutions (cf. Cambois et al. 2016; UNECE 2020). This could be explained with a more extreme (self-)selection mechanism in those countries, in contrast to countries where a larger fraction of residents live in institutions.

Figure 2.5: Marital status for the population older than 85 years within the institutionalized and private household populations



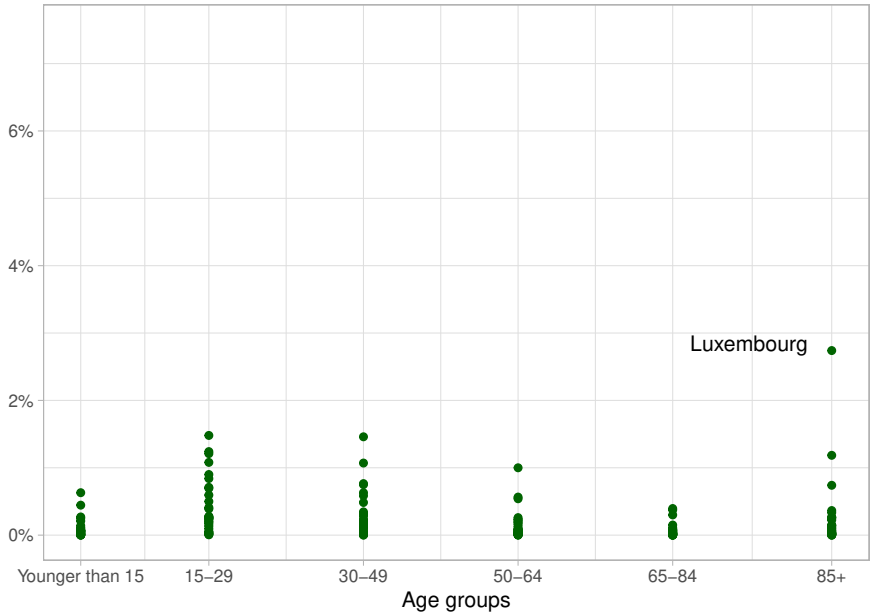
Citizenship

In contrast to marital status, the variance of the dissimilarity indexes for citizenship is larger for young and middle-aged groups and smaller for the elderly population (cf. Figure 2.6). In the youngest age group, 94% of the boys and girls living in private households hold citizenship by their native country. The equivalent share in institutions is lower by about 10 percentage points. 11% of the institutionalized boys and girls younger than 15 years hold a non-EU citizenship.

The larger diversity in institutions persists for most age groups: Between the age of 15 to 49 years, more than 90% of the community-dwelling population possesses the citizenship of the respective country. This proportion converges towards 100% with increasing age. Less than 2% of the community-dwelling European population older than 85 years does not hold the citizenship of their respective country. As can be seen in Table 2.5 in the Appendix, the group of institutionalized men and women sees a similar convergence towards 100% of citizenship of the reporting countries in the oldest age group. In the two oldest age groups, the differences across the two types of housing are rather small, which is why the larger relative size of the institutionalized population in those two age groups does not translate into larger dissimilarity indexes in most countries (cf. Figure 2.6). Indeed, this demographic pattern is spread homogeneously across nearly all European countries. Only Estonia and Latvia report higher shares of non-EU residents in private households and institutions for both genders in the two oldest age groups. As other exceptions for the el-

derly population, 16.5% of institutionalized men between the age of 65 to 84 years has a non-EU citizenship in France. The comparably high dissimilarity index in Luxembourg in the oldest age cohort can be explained with the lower share of EU foreigners in institutions (7%) compared to private households (15%).

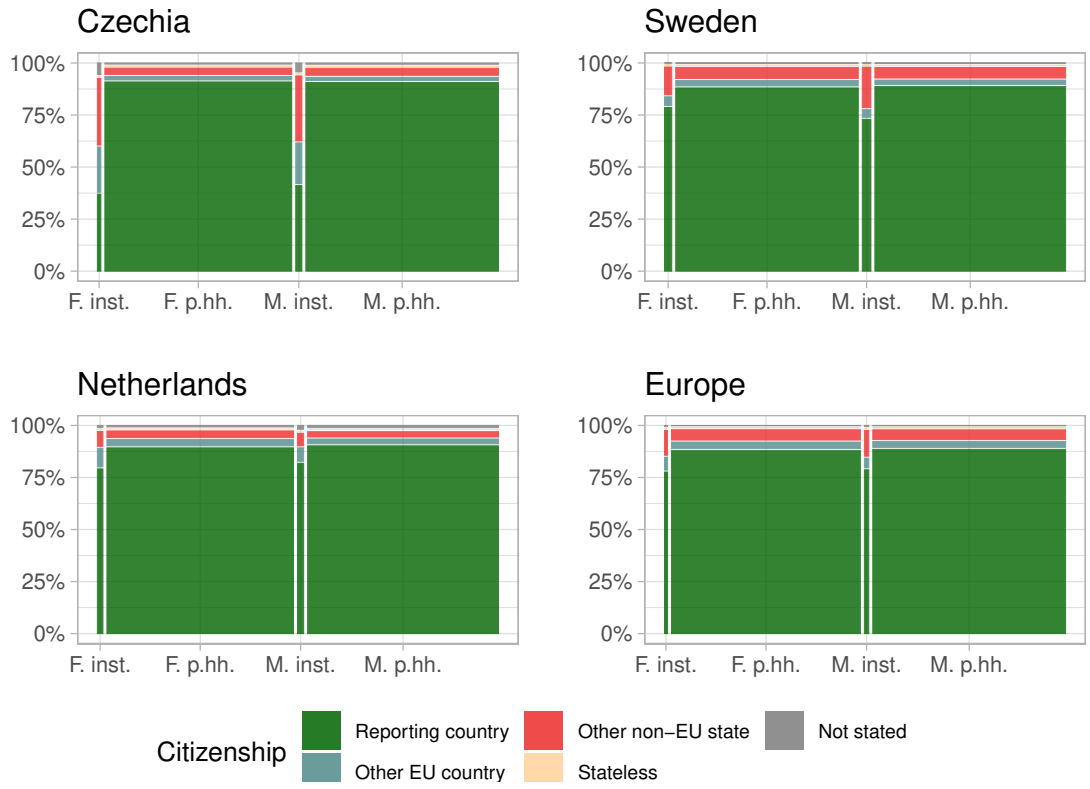
Figure 2.6: Duncan's Index of Dissimilarity comparing the citizenship of all citizens in private households to the total population



Many countries like Denmark, Finland, France, Germany, Portugal or the UK closely follow the pattern of the European aggregate. A second group of countries mainly follows the European pattern with particularities in some age and gender groups: Czechia reported a higher share of institutionalized foreigners from EU and non-EU countries in the middle-aged groups (cf. Figure 2.7), while Norway also counted a high share of institutionalized EU and non-EU foreigners in the two youngest age groups. In Sweden, half of all institutionalized children are citizens of mostly non-EU countries, and in Austria, Belgium, Italy, Luxembourg, and Malta a majority of institutionalized residents up to 49 years does not hold the respective national citizenship but possess a non-EU citizenship. In Luxembourg, a larger share of institutionalized residents possesses a citizenship by another EU member state compared to other countries. However, only Cyprus and Iceland reported more citizens of other EU countries than non-EU citizens as residents in their institutions. Finally, a last group of countries consists of mostly South-Eastern European countries: In Bulgaria, Croatia, Hungary, Latvia, Poland, and Romania, the overall share of non-national citizens is close to zero in private households and institutions.

It is somewhat difficult to infer from a high proportion of (non-EU) foreigners in institutions on the type of institution, since those residents might live in refugee accommodations, educational institutions, worker's dormitories, or prisons. The group of stateless persons is less ambiguous; it

Figure 2.7: Citizenship for the population between 15 to 29 years within the institutionalized and private household populations



can be assumed that this group mainly consist of persons living in refugee accommodations. The share of the stateless persons is minor in Europe, reaching a maximum of 0.1% in the institutionalized population up to 15 years. In Austria, Greece, Ireland, Luxembourg, Spain, and Sweden, this share is a little higher within the young institutionalized population, reaching a proportion of 0.5% to 2.3%. In Lithuania and Iceland, a similar share of institutionalized residents is stateless within the middle-aged and older age groups. Luxembourg reaches the highest share of stateless residents in institutions: Between 2.5% to 5.3% of the institutionalized men and women younger than 49 years did not hold any citizenship when the census took place in 2011. It can be expected that the number of stateless persons living in institutions increased in the meantime in many European countries as a consequence of the increasing immigration of refugees since 2015 (Eurostat 2020a).

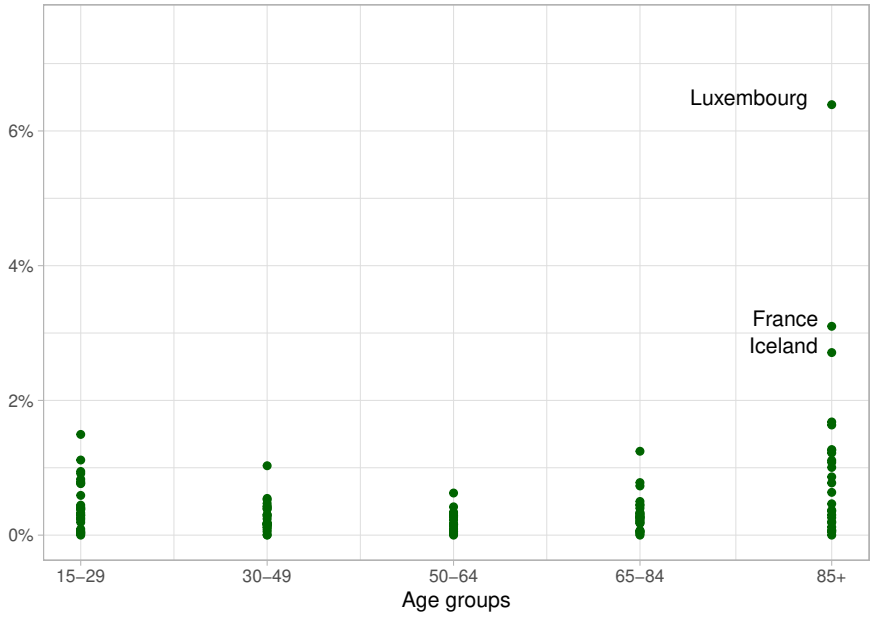
Education

The dissimilarity indexes for education in Figure 2.8 provide an overview of potentially critical age groups. The majority of countries does not suffer from coverage bias in the middle-aged part of their population. Only for some countries, the peril of bias is a little higher for the youngest and oldest age group, whenever institutionalized populations are excluded. Romania and Sweden obtain a dissimilarity index larger than 1% in the age group from 15 to 29 years. In both countries, the proportion of residents with a low level of education is much smaller in institutions than in

private households. This finding could be a hint that a majority of institutionalized residents lives in boarding schools or student dorms in those countries. It is reflected at the European level with a less pronounced difference between the two types of housing (cf. Table 2.6 in the Appendix). In Romania, most institutionalized adolescents have a medium level of education (secondary and post-secondary education), while the share of institutionalized residents with a high level of tertiary education is smaller than in private households. In Sweden, the share of residents living in institutions with a tertiary education is higher than the equivalent share in private households (cf. Figure 2.9).

Missing information is an important factor while interpreting country-specific results. Denmark, Finland, Ireland, Italy, and Poland did not report any numbers on education for institutions or private households to Eurostat, while additional countries only provided information for a fraction of their institutionalized populations, especially Estonia, Germany, Lithuania, and Slovenia. In those countries, the relative size of institutionalized residents is underestimated like for marital status. Due to this exclusion, the importance of any potential coverage bias caused by their exclusion will be underestimated. A little less than 10 countries suffer from a rather high share of residents without any substantive information on the level of education (e.g., Belgium, Cyprus, Estonia, Luxembourg Norway, Sweden).

Figure 2.8: Duncan's Index of Dissimilarity comparing the formal level of education of all citizens in private households to the total population

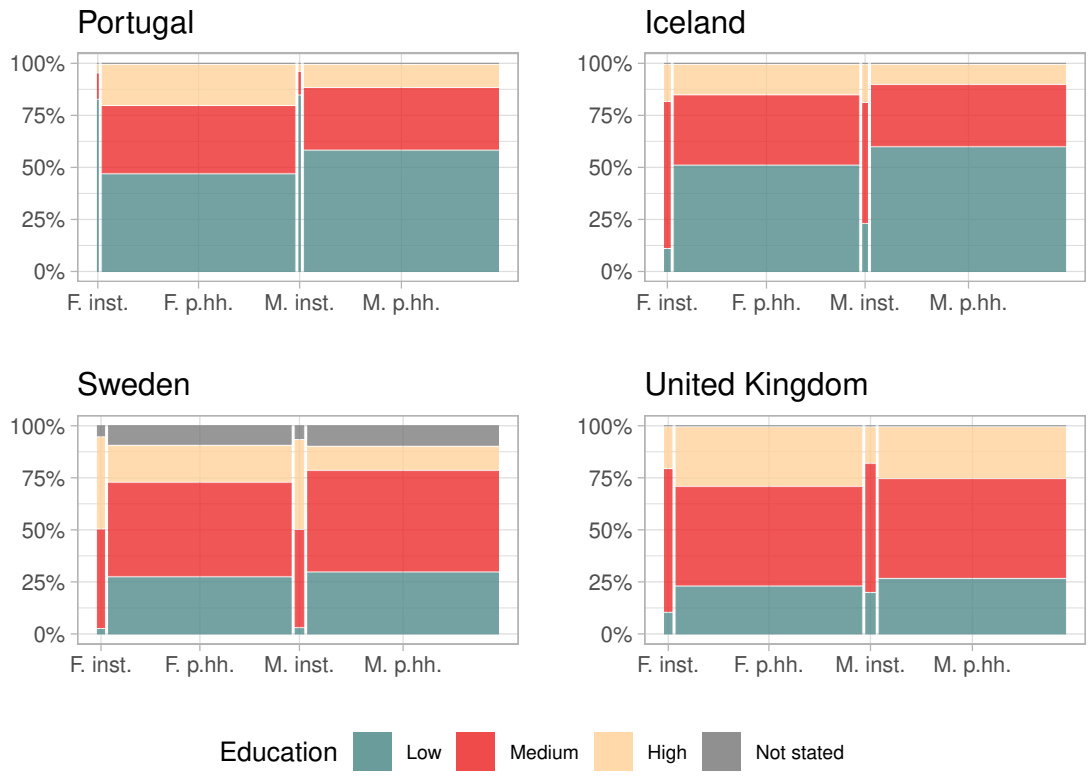


In addition to Sweden and the United Kingdom, Figure 2.9 also shows educational differences for young adults in Portugal and Iceland, which were both classified as countries with only one peak in the relative size of institutionalized residents earlier in this paper. In Portugal, the relative size of institutionalized residents is very small in this age group (roughly 0.5% in both genders) and

the respective educational distribution is nearly invisible in the mosaic plot. The comparably high share of institutionalized residents with a low level of formal education does not lead to a high dissimilarity index in Portugal.

Investigating the oldest age group, 11 countries reported numbers that lead to a dissimilarity index higher than 1%. Luxembourg (6.4%), France (3.1%), and Iceland (2.7%) are the three countries with the highest dissimilarity indexes. In Luxembourg, this is partly due to a rather high share of missing values in institutions and private households. In France, the average level of education is higher in institutions for elderly people than in private households. In contrast to the French numbers, the share of highly educated residents in Icelandic institutions is lower than in private households, while the share of elderly residents with a low level of education is higher in institutions. The contradictory findings from Iceland and France mirror earlier inconclusive results on the impact of education on institutionalization cited above.

Figure 2.9: Level of formal education for the population between 15 to 29 years within the institutionalized and private household populations



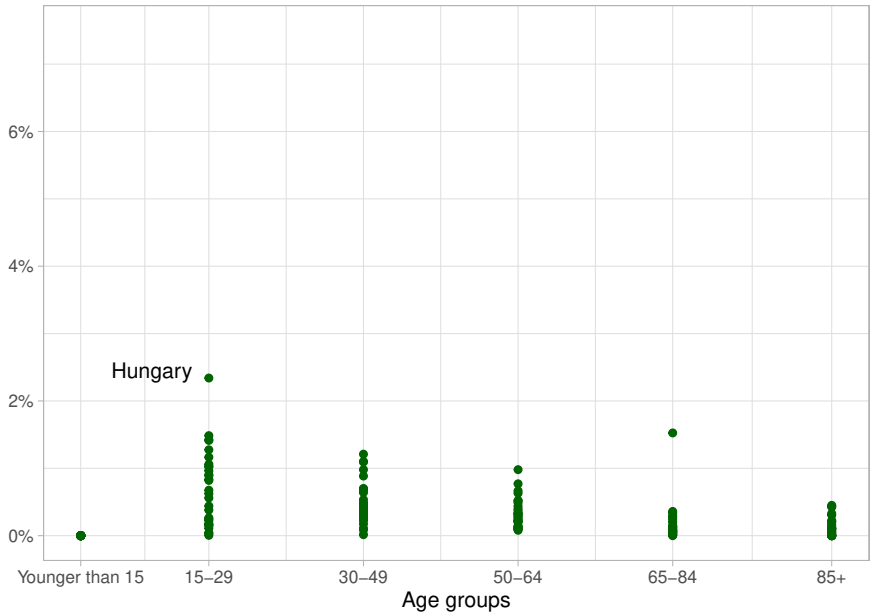
Economic Activity

The economic activity is expected to differ markedly across the two types of housing. In educational institutions, refugee accommodations, prisons, or health care institutions, residents are either not willing, not able or not allowed to work. This is why a larger share of economically inactive and unemployed persons can be expected in institutions.

The most interesting groups are young adults and middle-aged groups. Children and the very old

population are mostly economically inactive irrespective of their type of housing and will not be considered in the following analysis. All patterns in the age groups from 15 to 64 years are hinting in the same direction: A higher fraction of community-dwelling residents is economically active compared to the institutionalized populations (cf. Table 2.7 in the Appendix). The unemployment rates are quite similar across the two types of housing with a higher share of unemployed resident living in private households than in institutions. Again, several countries suffer from a large share of missing information in institutions (Cyprus, Czechia, Luxembourg, and Slovakia).

Figure 2.10: Duncan's Index of Dissimilarity comparing the economic activity of all citizens in private households to the total population

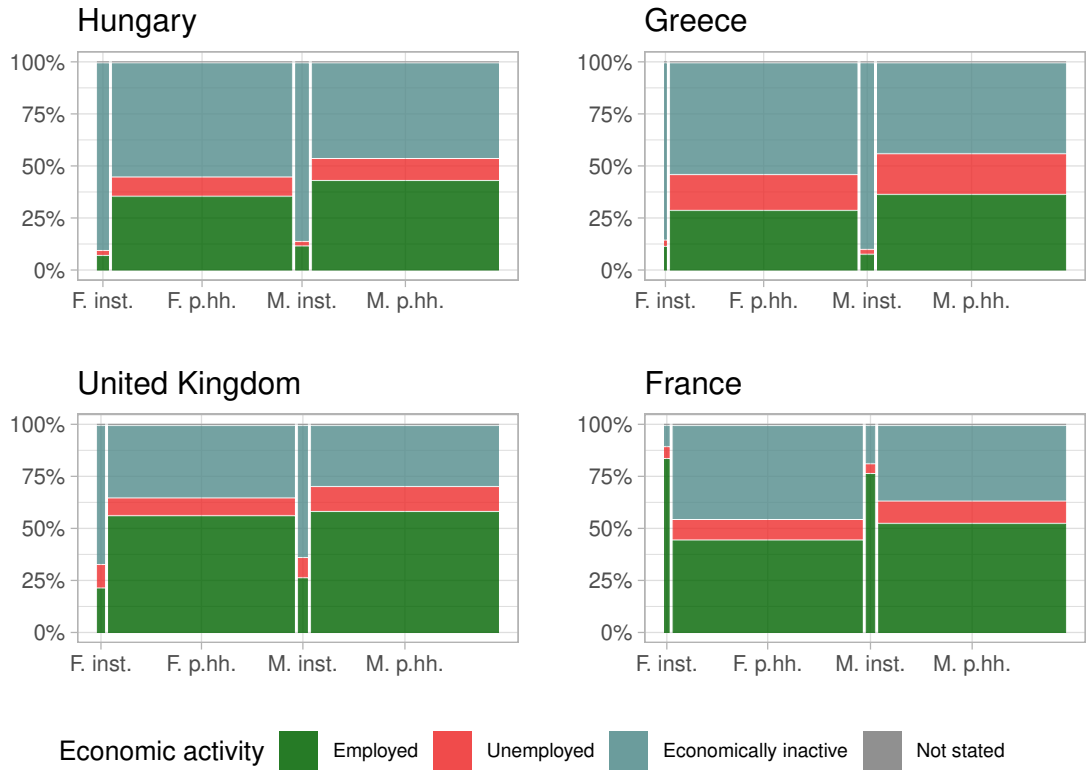


The strongest differences between institutions and private households can be observed in the two age groups from 30 to 64 years. Approximately 55% of the institutionalized men and women are economically inactive in the range from 30 to 49 years. This is the case for only 8% of the community-dwelling men and 19% of the community-dwelling women. In institutions, the share of economically inactive persons increases to more than 70% for the population aged between 50 to 60 years. The respective fraction of inactive persons living in private households is approximately half as large as in institutions in this age group. Due to a larger relative size of the institutionalized population in the age group from 15 to 29 years in many countries, the overall peril of coverage bias is a little larger in this age group than in the two subsequent age groups (cf. Figure 2.10). Hungary, the UK, Romania, Estonia, Sweden, and France have comparably high dissimilarity indexes. Those countries are classified as two peaks countries with a higher share of institutionalized young adults.

Regarding the economic activity of institutionalized residents, countries can be split in two groups. In the first group of countries, the large majority of institutionalized residents is economically in-

active. Figure 2.11 shows Hungary, Greece, and the United Kingdom as three examples from this group of 20 countries. In a second group of countries, the majority of institutionalized residents is economically active. The numbers for France are presented in Figure 2.11 as one example for this group of countries: 85% of the institutionalized French women aged between 15 to 29 years are employed, additional 5% are economically active but unemployed. The respective proportions among men are 78% of employed and 4% of unemployed institutionalized men. Further countries with a high share of economically active persons living in institutions are Cyprus and Latvia (especially female residents), as well as Germany, Iceland, Ireland, Slovenia, Spain, and Malta. In Malta, a large fraction of institutionalized residents is unemployed. Relying on the assumptions made in Table 2.1, institutionalized residents from the second group of countries could live in military barracks, worker's dormitories, or religious institutions. Differences in the economic activity of institutionalized residents across the two groups of countries could be also explained by different definitions. For instance, in some countries, prisoners might be defined as economically active or unemployed because they are working in the prison, while other countries do not count prison labor as an economic activity.

Figure 2.11: Economic activity for the population between 15 to 29 years within the institutionalized and private household populations



Size of locality

The sizes of localities provide information on the degree of urbanity of institutionalized populations. It can also be relevant during the sampling for surveys, if institutionalized residents are

either clustered in cities or in rural areas of a country, increasing the difficulty to cover them adequately with simple random sampling.

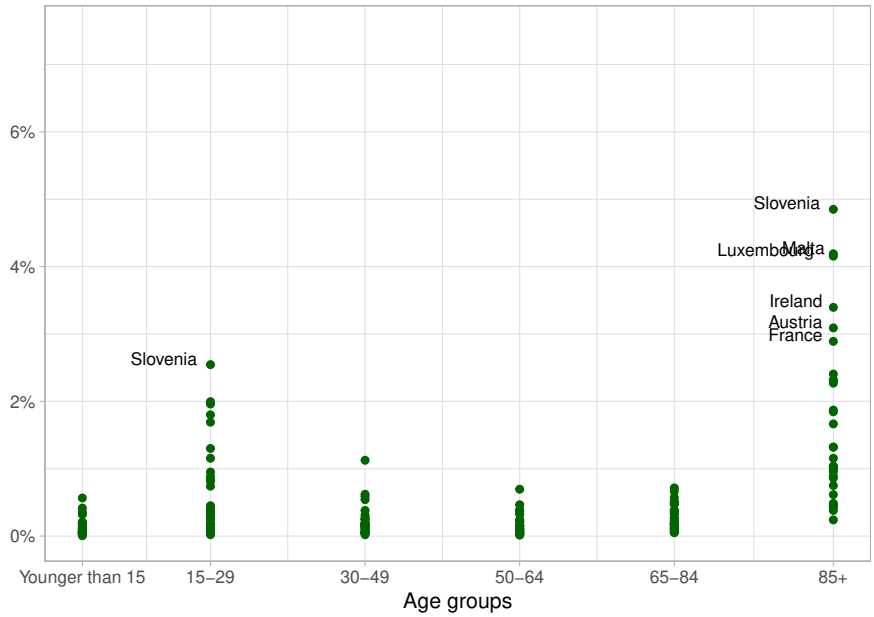
Figure 2.12 shows the distribution of the size of locality for the aggregate of all 30 European countries. The strongest difference across the two types of housing can be observed in the age group from 15 to 29 years. 35% of the institutionalized men and more than 46% of the institutionalized women live in metropolises with more than 200,000 inhabitants. The equivalent share is lower in private households: Roughly every fourth community-dwelling resident lives in very large cities in this age group. At the other end of the urbanity scale, institutionalized men and women live less often in villages, very small towns, or small towns than their peers living in private households.

For the European aggregate, the distribution is very similar in the age group younger than 15 years, and in the middle-aged groups from 30 to 64 years. For those age groups, the Duncan Indexes of Dissimilarity do not reach values higher than 1% in nearly all countries (cf. Figure 2.13), except in Finland due to a larger share of institutionalized men living in very large cities more often and less often in smaller localities than community-dwelling citizens. The dissimilarity indexes are higher for the age group above 85 years. In Europe, a smaller share of the institutionalized residents live in villages in this age group, whereas more institutions for elderly people are located in small and middle-sized towns.

Figure 2.12: Size of the locality residents live in in European countries within the institutionalized and private household populations

The small differences at the European level (cf. Table 2.8 in the Appendix) are driven by the rather

Figure 2.13: Duncan's Index of Dissimilarity comparing the size of locality of all citizens in private households to the total population



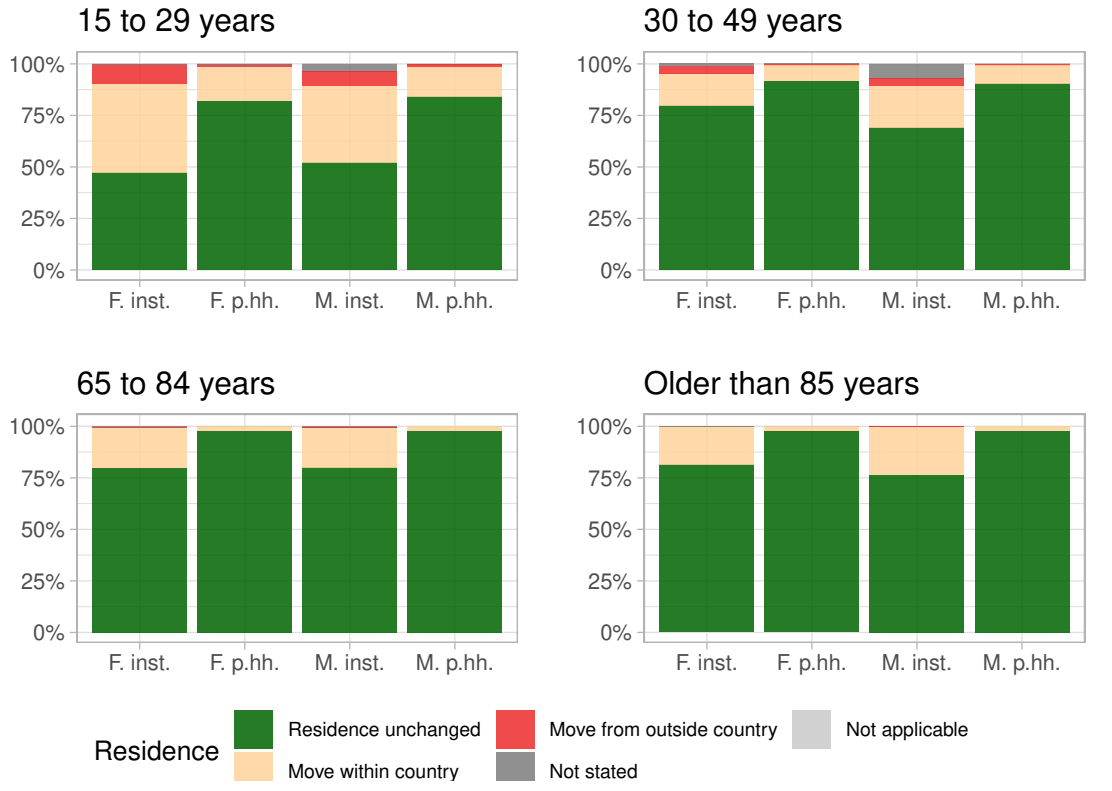
balanced numbers in a couple of large countries, namely France, Germany, Spain, and the UK. More countries like Austria, Czechia, the Netherlands, or Poland follow the European aggregate. A few particularities can be observed in other countries. While the institutionalized population is more urban than the community-dwelling population in most European countries, a higher share among institutionalized residents lives in smaller cities or villages compared to the population living in private households in Belgium, Estonia, Finland, Iceland, Spain, and the UK. In the UK, this is especially the case for younger age groups, but also in middle-aged groups. The urbanity is more balanced among the elderly population in the UK. The same observation holds for Belgium, while Estonia counted a higher share of institutionalized men and women living in rural parts of the country. In a third group of countries, the average urbanity of the institutionalized population changes with age. While the institutionalized populations live more rurally or in small cities in younger age groups in Bulgaria, Greece, the Netherlands, and Norway, the elderly institutionalized population lives more urban in those countries. In Hungary, Latvia, and Lithuania, it is the reverse. Those national patterns might be taken into account when developing a country-specific sampling design

Residential mobility

The residential mobility could be of interest to survey researchers, because more mobile groups of the population are harder-to-survey. For those residents, addresses retrieved from sampling frames could be outdated with a higher probability compared to residents with a high residential stability. This variable can be also interpreted as as a proxy for the average duration of residence in institutions. The higher the mobility during the last year, the shorter the average duration in a specific

institution. A number of countries did not report any information for this variable to Eurostat.[3] The assumption of a higher mobility of institutionalized populations is definitely confirmed for the 19 European countries for which sufficient data is available: In 228 different age and gender groups, only once community-dwelling residents changed their residence more often than their institutionalized counterparts.[4] Looking at the European aggregate as depicted in Figure 2.14, mobility is highest for the institutionalized population younger than 30 years. While more than 80% of the population living in private households did not change their place of residence in the year leading to the census, this was only the case for every second citizen who lived in institutions during the census. A majority of institutionalized residents moved in from another place within the respective country, while roughly 5 to 10% of institutionalized residents came from another country. This finding reflects numbers on citizenship presented above.

Figure 2.14: Changing residence within previous year in European countries within the institutionalized and private household populations



A few European countries reported higher proportions of institutionalized residents who moved in from abroad during the previous year. This is especially the case in Austria, Belgium, Malta, Norway, and in the UK, to a lesser extent. In Austria, 36% of institutionalized men aged between

[3]This groups contains Denmark, Finland, Germany, Iceland, Italy, Lithuania, Spain, and Sweden. In addition, the census in Czechia, Cyprus, Estonia, Luxembourg, and Romania reported a very high share of missing information for multiple age and gender cohorts. Czechia, Romania, and Slovakia were dropped from the analysis due to this reason.

[4]In France, 75.3% of the community-dwelling women aged between 15 to 29 years did not change residence compared to 75.8% of institutionalized women.

15 to 29 years came from abroad during the previous year, which is the case for 38% of female institutionalized residents in Malta. Those foreign institutionalized residents might live in educational institutions, refugee accommodations, or worker's dormitories. The published census data does not provide more information on the origin of residents moving in from abroad.
The residential mobility of institutionalized populations is much lower in the older age groups. Moreover, institutionalized residents rather moved to institutions within their respective countries instead of immigrating (cf. Figure 2.14). While more than 95% of the community-dwelling population older than 50 years did not change residence in the previous year, this was the case for approximately 80% of the institutionalized men and women. Still, a majority of institutionalized residents stays in the institution for more than one year.

2.5 Conclusion

The present paper analyzes census data on the institutionalized population living in Europe. This part of the population is usually excluded from social surveys, which might cause bias in estimates derived from those surveys. Coverage bias is driven (1) by the relative size of the excluded population and (2) by its statistical distinctiveness compared to the covered population. Both factors are analyzed in this paper.
Considering the relative size of institutionalized populations within age and gender groups, three groups of countries can be identified. The first group of 13 European countries are labeled as *one peak countries*, because institutionalized populations usually amount to less than 1% of the age and gender groups up to the age of 70 years. Only in older age cohorts the share of institutionalized residents increases constantly, up to 10%, 20% or even more than 30%, especially in the very old population. A second group of 11 European countries is labeled as *two peaks countries*. In addition to a similar increase of institutionalized residents in older age groups, this group of countries records a second, smaller peak among younger residents. Between the age of 20 to 29 years, 3% to 8% of their population is institutionalized. In a third group of six remaining countries from Eastern and South-Eastern Europe, only a minority of less than 10% of the elderly population lives in institutions in all age groups.
The size of institutionalized populations serves as a first indicator for the peril of bias in respective age and gender groups. Social survey programs operating in one peak countries, as well as in Latvia and Poland, should not worry too much about institutionalized citizens younger than 70 years, since their negligible share will not lead to a significant bias even though those citizens might be statistically quite distinctive when compared to the community-dwelling population. This is different for the group of two peaks countries: The comparably larger share of young residents living in institutions between the age of 10 to 29 years should be a reason of concern for survey researchers. The institutionalized populations aged between 30 to 69 years are not relevant for social surveys in any European country when considering their minor relative size. This conclu-

sion is different for the elderly institutionalized population. All social survey programs need to be aware about the proportion of old-aged institutionalized residents in nearly all European countries. They amount to a substantial share within their age cohort (cf. Eurostat 2015, 148).

Turning to the statistical distinctiveness of institutionalized populations, it can be assumed that moving to a refugee accommodation, prison, or nursing home implies a very strong selection mechanism, leading to various differences across the two types of housing. Depending on the topic examined, data users might suffer from an underestimation or overestimation of certain behaviors, attitudes, or characteristics. Given the availability of data, this paper analyzes a few basic variables like marital status, education, or economic activity and how institutionalized populations in different age and gender groups differ from the population living in private households. Institutionalized residents between 15 to 29 years do not differ too much from their community-dwelling counterparts with respect to marital status and their formal level of education. Smaller differences emerge when investigating citizenship, economic activity, and the size of locality. Young institutionalized residents are more often from non-EU countries, economically inactive, and live more often in large cities and metropolises than the respective age group living in private households. It can be assumed that young institutionalized residents live predominantly in educational institutions, and in military institutions, refugee accommodations, or prisons. Roughly every second institutionalized resident in this age group changed their residence in the previous year compared to approximately 15% percent in private households. This can be a reason why this part of the population is more difficult to reach than persons of the same age living in private households.

Looking at the population older than 85 years, small to no differences between institutions and private households can be observed regarding citizenship, the level of education, and economic activity. A higher share of institutionalized elderly was never married than the population living in private households, institutionalized men of this age are more often widowed than community-dwelling men. Moreover, the old-aged institutionalized population lives more often in small and middle-sized towns and less often in villages. It can be assumed that the population older than 85 years lives predominantly in retirement and nursing homes. This populations stays in their institutions for a while: Approximately 75% of those people did not change their residence in the year prior to the census. Still, they are a little more mobile than the population living in private households, however, potentially less difficult to reach than the young institutionalized population.

The comparative analysis of 30 European countries reveals an astonishing homogeneity of many age- and gender-related patterns within Europe. Nevertheless, I try to shed light on multiple national particularities, which might be of interest for researchers working on the demography of those countries. To conclude, institutionalized populations in young and middle-aged age cohorts are rather of interest for researcher with more specific research interests like educational aspirations of young adolescents, acceptance and violation of the legal system, the integration of refugees, or the social inclusion of disabled citizens. This conclusion is different for institutionalized populations in old age groups. This group reaches a critical size above 10% of the European population and it is relevant for any social survey program in nearly all European countries. Even basic variables like marital status, citizenship, or size of locality differ slightly between institu-

tions and private households. The small differences in those demographic variables found in this paper might easily cause further bias in additional behavioral or attitudinal variables in addition to the impact of institutionalization as such. In the near future, the proportion of very old people will continue to increase in many European countries (Rodrigues et al. 2012), which will most probably also increase the absolute size of the institutionalized population and its relevance for social surveys.

References

Abbott, Owen and Garnett Compton (2014). "Counting and estimating hard-to-survey populations in the 2011 census". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 58–81.

Andersen, Espen and Harald Utne (2011). "Censuses in a register-based statistical system : Norwegian experiences". In: *Int. Statistical Inst.: Proc. 58th World Statistical Congress*. Dublin (IE): 1162–1171.

Angelini, Viola and Anne Laferrère (2012). "Residential mobility of the European elderly". *CESifo Economic Studies* 58(3): 544–569.

Bajekal, Madhavi, Levin Wheller, and David Dix (2006). *Estimating residents and staff in communal establishments from the 2001 Census*. Tech. rep. Office for National Statistics: 42–50.

Cambois, Emmanuelle, Carol Jagger, Wilma Nussleder, Herman Van Oyen, and Jean-Marie Robine (2016). "International Comparisons of Disability Prevalence Estimates: Impact of Accounting or Not Accounting for the Institutionalized Population". In: *International Measurement of Disability*. Ed. by Barbara Altmann: 207–229.

Castora-Binkley, Melissa, Hongdao Meng, and Kathryn Hyer (2014). "Predictors of long-term nursing home placement under competing risk: Evidence from the health and retirement study". *Journal of the American Geriatrics Society* 62(5): 913–918.

Duncan, Otis and Beverly Duncan (1955). "A Methodological Analysis of Segregation". *American Sociological Review* 20(2): 210–217.

Einio, Elina, Christine Guilbault, Pekka Martikainen, and Michel Poulain (2012). "Gender Differences in Care Home Use Among Older Finns and Belgians". *Population* 67(1): 75–101.

Eurostat (2011). *EU legislation on the 2011 Population and Housing Censuses: Explanatory Notes*. Tech. rep. Luxembourg: European Union.

– (2015). *People in the EU: who are we and how do we live?* Tech. rep. Luxembourg.

– (2016). *European Statistical System - Census Hub*. URL: https://ec.europa.eu/CensusHub2.

– (2020a). *Database on Asylum and Managed Migration*. URL: https://ec.europa.eu/eurostat/web/asylum-and-managed-migration/data/database.

– (2020b). *Prison statistics*. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Prison%7B%5C_%7Dstatistics.

Feskens, Remco (2009). "Difficult Groups in Survey Research and the Development of Tailor-made Approach Strategies". PhD thesis. Universiteit Utrecht.

Geerts, Joanna and Karel van den Bosch (2012). "Transitions in formal and informal care utilisation amongst older Europeans: The impact of national contexts". *European Journal of Ageing* 9(1): 27–37.

Geiger, Marion and Doris Styhler (2012). *Zensus 2011 : Erhebungsteil Sonderbereiche*. Tech. rep. Bayern in Zahlen (5): 280–285.

Goffman, Erving (1957). *Characteristics of Total Institutions*. Tech. rep. Washington DC, USA: Symposium on Preventive and Social Psychiatry.

Groves, Robert, Floyd Fowler, Mick Couper, James Lepkowski, Eleanor Singer, and Roger Tourangeau (2009). *Survey Methodology*. Second edi. New York: John Wiley & Sons, Inc.

Laferrère, Anne, Aaron Van Den Heede, and Karel Van Den Bosch (2012). "Entry into institutional care: predictors and alternatives". In: *Active ageing and solidarity between generations in Europe*. Ed. by Axel Börsch-Supan, Martina Brandt, Howard Litwin, and Guglielmo Weber: 253–264.

Lessler, Judith T. and William D. Kalsbeek (1992). *Nonsampling Error in Surveys*. New York: John Wiley & Sons, Inc.

Lewinter, Myra (2004). "Developments in home help for elderly people in Denmark: The changing concept of home and institution". *International Journal of Social Welfare* 13(1): 89–96.

Lohr, Sharon (2010). *Sampling: Design and Analysis*. Boston (MA), USA: Brooks/Cole.

Luppa, Melanie, Tobias Luck, Herbert Matschinger, Hans-Helmut König, and Steffi Riedel-Heller (2010). "Predictors of nursing home admission of individuals without a dementia diagnosis before admission - results from the Leipzig Longitudinal Study of the Aged (LEILA 75+)." *BMC health services research* 10: 186.

Martikainen, Pekka, Heta Moustgaard, Michael Murphy, Elina Einiö, Seppo Koskinen, Tuija Martelin, and Anja Noro (2009). "Gender, Living Arrangements, and Social Circumstances as Determinants of Entry into and Exit from Long-Term Institutional Care at Older Ages: A 6-Year Follow-Up Study of Older Finns". *Gerontologist* 49(1): 34–45.

McCann, Mark, Emily Grundy, and Dermot O'Reilly (2012). "Why is housing tenure associated with a lower risk of admission to a nursing or residential home? Wealth, health and the incentive to keep 'my home'". *Journal of Epidemiology & Community Health* 66(2): 166–169.

Mulry, Mary (2014). "Measuring undercounts for hard-to-survey groups". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press: 37–57.

OECD (2006). *Glossary of statistical terms: Institutional households - Eurostat*. URL: https://stats.oecd.org/glossary/detail.asp?ID=1372 (visited on 08/02/2018).

ONS (2015). "2011 census analysis: What does the 2011 census tell us about people living in communal establishments?" *Office for National Statistics*: 1–32.

Pimouguet, Clément, Debora Rizzuto, Pär Schön, Behnaz Shakersain, Sara Angleman, Marten Lagergren, Laura Fratiglioni, and Weili Xu (2016). "Impact of living alone on institutionaliza-

tion and mortality: A population-based longitudinal study". *European Journal of Public Health* 26(1): 182–187.

Pirou, David, Nadine Poullain, and Sandrine Rochelle (2013). "La vie en communauté". *INSEE Première*(1434): 6–9.

Poulain, Michel and Anne Herm (2013). "Central Population Registers as a Source of Demographic Statistics in Europe". *Population* 68(2): 183–212.

Rodrigues, Ricardo, Manfred Huber, and Giovanni Lamura (2012). *Facts and Figures on Healthy Ageng and Long-term Care: Europe and North America*. Vienna: European Centre for Social Welfare Policy and Research.

Schanze, Jan-Lucas (2017). *Report on sampling practices for the institutionalized population in social*. Tech. rep. Deliverable 2.16 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. URL: https://seriss.eu/resources/deliverables/.

Schanze, Jan-Lucas and Ma'ayan Levinson (2019). *Report on what persons live in institutions and the most relevant institution types they live in, the availability of data to select them and their ability to be reached in practice*. Tech. rep. Deliverable 2.17 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. URL: https://seriss.eu/resources/deliverables/.

Schnell, Rainer (1991). "Wer ist das Volk? Zur faktischen Grundgesamtheit bei „allgemeinen Bevöllkerungsfragen": Undercoverage, Schwererreichbare und Nichtbefragbare". *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 1(43): 106–137.

Sweet, Elizabeth M. and Nicholas S. Alberti (1994). *Coverage Implications of Using the Term 'Usual Residence'*. Tech. rep. Washington DC, USA: Bureau of the Census: 319–324.

Thomeer, Mieke Beth, Stipica Mudrazija, and Jacqueline Angel (2015). "How do race and hispanic ethnicity affect nursing home admission? Evidence from the health and retirement study". *Journals of Gerontology, Series B* 70(4): 628–638.

Tourangeau, Roger (2014). "Defining hard-to-survey populations". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy Johnson, Kirk Wolter, and Nancy Bates. Cambridge University Press: 3–20.

UNECE (2020). *Recommendations for Measuring Older Populations in Institutions*. ECE/CES/ST. 1. Geneva (CH): United Nations.

U.S. Census Bureau (2012). "Technical documentation: 2010 Census Summary File 1". Washington DC, USA.

Valente, Paolo (2010). "Census taking in Europe: how are populations counted in 2010?" *Population & Societies*(467): 1–4.

Walmsley, Roy (2017). "World female imprisonment list: Fourth edition, women and girls in penal institutions, including pre-trial detainees/remand prisoners". *World Prison Brief*: 1–13.

– (2018). "World Prison Population List (twelfth edition)". *Institute for Criminal Policy Research* 12: 1–19.

Appendix

Table 2.3: Relative share of institutionalized populations in age and gender groups

Country	% in pop.	% in male population						% in female population					
		0-15	15-29	30-49	50-64	65-84	>85	0-15	15-29	30-49	50-64	65-84	>85
Austria	1.4	0.5	1.6	1.0	0.9	1.7	9.0	0.5	1.1	0.5	0.6	2.9	19.6
Belgium	1.2	0.1	0.6	0.6	0.7	1.9	11.7	0.1	0.3	0.2	0.5	3.7	24.9
Bulgaria	0.4	0.7	0.9	0.6	0.3	0.4	1.5	0.5	0.3	0.1	0.1	0.5	1.7
Croatia	0.9	0.2	0.7	0.7	0.8	1.6	8.0	0.2	0.4	0.3	0.6	2.6	11.5
Cyprus	0.5	0.1	0.3	0.5	0.4	0.9	7.6	0.1	0.2	0.3	0.4	1.7	14.1
Czechia	1.9	0.6	3.0	2.7	1.8	1.7	7.3	0.6	1.8	1.1	0.9	2.4	13.9
Denmark	1.5	0.3	4.5	0.7	0.6	1.2	7.8	0.2	3.7	0.4	0.4	1.6	12.1
Estonia	1.9	0.7	3.6	1.9	2.1	2.6	5.7	0.6	2.9	0.7	1.0	2.1	9.0
Finland	0.6	0.1	0.6	0.2	0.3	0.9	5.1	0.1	0.5	0.1	0.2	1.3	8.5
France	2.1	0.3	4.3	1.5	1.4	2.8	13.3	0.2	2.4	0.5	0.6	3.3	24.0
Germany	1.5	0.4	1.5	1.0	1.0	1.9	10.7	0.4	0.8	0.5	0.6	2.9	22.1
Greece	1.1	0.2	6.1	1.0	0.6	0.6	1.5	0.2	1.0	0.3	0.4	0.8	3.3
Hungary	2.4	1.2	6.3	1.5	1.5	2.1	6.4	1.1	5.8	0.6	0.7	2.8	12.0
Iceland	2.0	0.2	0.9	0.9	0.9	5.7	35.3	0.1	0.8	0.5	0.7	8.3	46.7
Ireland	1.9	0.4	1.7	1.4	1.6	4.8	19.6	0.4	1.1	1.0	1.4	5.8	28.8
Italy	0.6	0.1	0.4	0.5	0.4	0.7	2.5	0.1	0.2	0.3	0.3	1.4	6.7
Latvia	1.2	0.5	2.1	1.7	1.3	1.8	2.5	0.4	0.9	0.4	0.6	1.3	4.3
Lithuania	1.6	1.2	3.5	1.6	1.2	1.2	1.8	1.1	3.4	0.9	0.6	1.0	3.4
Luxembourg	1.8	0.7	2.2	0.9	0.7	2.8	19.6	0.6	0.9	0.4	0.6	5.7	37.2
Malta	2.1	0.8	2.3	1.8	0.9	2.7	19.1	0.7	0.9	0.6	0.8	6.0	35.1
Netherlands	1.7	0.3	3.1	1.1	0.9	1.7	13.2	0.3	2.9	0.6	0.6	2.8	23.2
Norway	0.6	0.0	0.1	0.1	0.2	1.1	8.7	0.0	0.1	0.0	0.1	1.7	15.9
Poland	0.6	0.2	0.9	0.3	0.6	0.8	1.3	0.2	0.9	0.3	0.3	0.9	3.0
Portugal	1.2	0.4	0.8	0.7	0.6	2.0	11.8	0.4	0.5	0.3	0.5	3.2	18.7
Romania	1.0	0.4	3.6	0.7	0.4	0.4	0.9	0.3	3.2	0.3	0.2	0.4	1.4
Slovakia	0.9	0.6	0.6	0.7	1.0	1.9	5.5	0.6	0.5	0.5	0.7	2.5	8.5
Slovenia	2.8	0.6	5.0	3.5	2.9	2.6	11.7	0.6	4.2	0.6	0.6	3.8	21.2
Spain	0.5	0.0	0.2	0.2	0.3	1.1	5.6	0.0	0.1	0.1	0.2	1.5	9.2
Sweden	1.7	0.2	4.1	0.8	0.5	1.4	9.6	0.1	3.7	0.5	0.3	1.9	16.9
UK	1.8	0.3	4.6	1.0	0.7	1.5	9.6	0.3	3.7	0.4	0.4	2.1	18.7
Europe	1.3	0.3	2.4	0.9	0.8	1.5	7.9	0.3	1.7	0.4	0.5	2.2	15.4

Table 2.4: The distribution of marital status in European countries in the total population (Tot.), private households (P.hh.) and institutions (Inst.)

		0-15 years			15-29 years			30-49 years			50-64 years			65-84 years			> 85 years		
Gender	**Marital status**	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.
Female	Never married	100.0	100.0	99.5	83.6	83.4	95.9	25.2	25.0	72.8	8.5	8.3	58.6	5.8	5.4	27.6	8.0	6.8	14.8
	Married	0.0	0.0	0.0	15.3	15.5	2.0	63.5	63.7	15.3	69.3	69.5	14.9	48.1	48.8	12.6	11.4	12.4	5.5
	Widowed	0.0	0.0	0.0	0.1	0.1	0.1	1.3	1.3	1.6	9.2	9.2	8.8	39.5	39.2	51.6	77.6	77.9	76.2
	Divorced	0.0	0.0	0.0	0.9	0.9	0.3	9.9	9.9	8.2	12.9	12.9	15.8	6.6	6.6	7.7	2.9	2.8	3.3
	Not stated	0.0	0.0	0.5	0.2	0.2	1.7	0.1	0.1	2.2	0.0	0.0	1.9	0.0	0.0	0.6	0.1	0.0	0.2
Male	Never married	99.9	99.9	99.5	91.3	91.2	95.9	34.6	34.4	67.6	12.1	11.8	53.6	6.2	5.7	35.2	5.1	4.3	14.2
	Married	0.0	0.0	0.0	8.1	8.2	2.5	57.8	58.1	20.1	74.7	75.1	20.6	77.2	78.0	27.4	54.9	57.2	27.4
	Widowed	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.4	0.6	2.3	2.3	3.2	10.9	10.8	23.7	38.1	36.7	54.9
	Divorced	0.0	0.0	0.0	0.4	0.4	0.4	7.0	7.0	10.1	10.9	10.8	20.8	5.6	5.5	13.1	1.9	1.8	3.4
	Not stated	0.0	0.0	0.4	0.2	0.2	1.1	0.1	0.1	1.5	0.1	0.0	1.8	0.0	0.0	0.7	0.0	0.0	0.1

Table 2.5: The distribution of citizenship in European countries in the total population (Tot.), private households (P.hh.) and institutions (Inst.)

		0-15 years			15-29 years			30-49 years			50-64 years			65-84 years			> 85 years		
Gender	**Citizenship**	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.
Female	Reporting country	94.1	94.1	84.6	91.2	91.4	80.7	91.1	91.2	76.5	95.3	95.4	91.0	97.4	97.4	97.4	98.4	98.4	98.2
	Other EU state	2.1	2.1	2.3	3.4	3.3	6.4	3.5	3.5	5.9	2.1	2.1	2.8	1.3	1.3	1.4	0.9	0.9	1.1
	Non-EU state	3.6	3.6	11.3	5.3	5.1	12.3	5.2	5.1	15.7	2.4	2.4	5.2	1.2	1.2	0.9	0.6	0.6	0.5
	Stateless	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Not stated	0.2	0.2	1.6	0.2	0.2	0.6	0.1	0.1	2.0	0.1	0.1	0.9	0.1	0.1	0.3	0.1	0.1	0.1
Male	Reporting country	93.9	93.9	85.6	91.7	92.0	81.8	91.0	91.2	75.6	95.3	95.4	83.8	96.7	96.7	93.0	98.0	98.0	97.8
	Other EU state	2.2	2.2	2.2	2.9	2.9	4.9	3.5	3.5	6.0	2.3	2.2	3.5	1.8	1.8	1.6	1.2	1.1	1.3
	Non-EU state	3.8	3.7	10.6	5.1	4.9	12.8	5.3	5.2	17.1	2.3	2.2	11.8	1.5	1.4	4.8	0.7	0.7	0.8
	Stateless	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Not stated	0.2	0.2	1.4	0.2	0.2	0.6	0.2	0.2	1.3	0.1	0.1	0.9	0.1	0.0	0.5	0.1	0.1	0.2

Table 2.6: The distribution of the formal level of education in European countries in the total population (Tot.), private households (P.hh.) and institutions (Inst.)

		15-29 years			30-49 years			50-64 years			65-84 years			> 85 years		
Gender	**Education**	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.
Female	Low	32,7	32,9	21,3	22,7	22,6	49,3	38,1	38,0	57,8	61,9	61,9	65,8	72,8	73,7	67,9
	Medium	42,6	42,3	59,5	42,5	42,6	24,6	39,6	39,6	24,4	26,7	26,8	21,9	19,3	18,9	21,2
	High	23,0	23,1	17,4	34,2	34,2	19,7	22,0	22,0	13,8	10,9	10,9	10,7	7,2	6,7	9,6
	Not stated	1,7	1,7	1,9	0,6	0,6	6,5	0,4	0,3	4,0	0,5	0,4	1,6	0,8	0,7	1,3
Male	Low	37,3	37,5	29,2	23,7	23,5	47,4	30,2	30,0	54,0	45,4	45,2	57,6	57,7	57,7	57,7
	Medium	44,6	44,4	53,8	45,1	45,2	31,7	44,0	44,1	29,8	34,2	34,3	27,4	25,6	25,6	25,5
	High	16,4	16,5	14,8	30,4	30,6	14,6	25,4	25,5	11,8	20,1	20,2	13,4	16,2	16,2	15,7
	Not stated	1,6	1,6	2,2	0,7	0,7	6,2	0,4	0,4	4,3	0,4	0,4	1,6	0,6	0,5	1,1

Table 2.7: The distribution of economic activity in European countries in the total population (Tot.), private households (P.hh.) and institutions (Inst.)

		0-15 years			15-29 years			30-49 years			50-64 years			65-84 years			> 85 years		
Gender	**Economic activity**	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.
Female	Employed	0.0	0.0	0.0	46.1	46.3	34.8	72.4	72.6	34.2	51.1	51.3	20.2	4.8	4.8	1.8	1.2	1.3	0.8
	Unemployed	0.0	0.0	0.0	9.6	9.7	5.8	8.6	8.6	5.0	5.1	5.1	2.4	0.2	0.2	0.0	0.2	0.2	0.0
	Inactive	100.0	100.0	100.0	43.9	43.7	58.3	18.7	18.6	57.2	43.6	43.5	75.6	95.0	94.9	98.0	98.6	98.5	99.2
	Not stated	0.0	0.0	0.0	0.4	0.3	1.2	0.2	0.2	3.7	0.2	0.1	1.8	0.1	0.0	0.1	0.0	0.0	0.0
Male	Employed	0.0	0.0	0.0	50.9	51.2	37.7	83.2	83.7	35.7	64.1	64.4	21.8	9.3	9.4	2.9	2.2	2.3	1.3
	Unemployed	0.0	0.0	0.0	10.9	11.0	5.0	8.2	8.2	6.1	6.1	6.1	4.5	0.3	0.3	0.1	0.2	0.3	0.0
	Inactive	100.0	100.0	100.0	37.8	37.4	55.9	8.2	7.8	54.1	29.6	29.3	71.3	90.3	90.2	96.8	97.4	97.3	98.6
	Not stated	0.0	0.0	0.0	0.4	0.4	1.4	0.3	0.3	4.1	0.2	0.2	2.4	0.1	0.1	0.3	0.1	0.1	0.1

Table 2.8: The distribution of size of locality in European countries in the total population (Tot.), private households (P.hh.) and institutions (Inst.)

		0-15 years			15-29 years			30-49 years			50-64 years			65-84 years			> 85 years		
Gender	**Size of locality**	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.
Female	Village	20.3	20.3	20.0	17.7	17.9	7.9	18.9	18.9	19.5	20.2	20.2	21.4	20.6	20.6	15.5	19.5	20.6	13.8
	Very small town	10.4	10.4	8.0	9.1	9.1	4.5	9.8	9.8	10.0	10.1	10.1	11.3	10.6	10.5	12.6	10.4	10.1	12.5
	Small town	17.6	17.6	16.9	16.2	16.4	9.5	17.2	17.2	17.1	17.4	17.4	18.6	17.5	17.4	21.2	17.6	16.9	21.5
	Middle-sized town	12.3	12.3	12.2	12.1	12.2	9.1	12.4	12.4	11.3	12.4	12.4	11.4	12.0	12.0	12.5	11.9	11.8	12.6
	Large town	15.0	15.0	18.4	16.5	16.4	22.8	15.5	15.5	14.8	15.4	15.4	13.8	14.6	14.6	14.4	14.7	14.7	14.6
	Metropolis	24.4	24.4	24.6	28.4	28.1	46.3	26.3	26.3	27.3	24.6	24.6	23.5	24.8	24.8	23.8	25.9	26.0	25.0
Male	Village	20.4	20.3	22.3	18.8	18.9	12.3	19.7	19.7	18.6	22.3	22.2	20.3	22.1	22.1	17.2	21.7	22.2	15.7
	Very small town	10.4	10.4	8.5	9.4	9.5	6.7	9.9	10.0	8.5	10.4	10.4	9.6	10.7	10.7	11.9	10.5	10.3	13.0
	Small town	17.6	17.6	17.0	16.5	16.6	13.9	17.0	17.0	16.2	17.3	17.3	17.0	17.4	17.4	19.9	17.2	16.8	21.4
	Middle-sized town	12.3	12.3	12.2	12.1	12.1	11.2	12.1	12.1	12.8	11.9	11.9	12.7	11.8	11.8	13.0	11.5	11.4	12.4
	Large town	15.0	15.0	16.8	16.1	16.0	20.8	15.2	15.2	17.4	14.6	14.6	16.1	14.1	14.1	14.9	14.4	14.4	14.3
	Metropolis	24.3	24.3	23.2	27.2	27.0	35.2	26.0	26.0	26.6	23.5	23.5	24.3	23.8	23.8	23.0	24.7	24.8	23.2

Table 2.9: The distribution of changes of residence in European countries in the total population (Tot.), private households (P.hh.) and institutions (Inst.)

		0-15 years			15-29 years			30-49 years			50-64 years			65-84 years			> 85 years		
Gender	**Residence**	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.	Tot.	P.hh.	Inst.
Female	Residence unchanged	84.6	84.7	57.8	81.2	82.0	47.1	91.4	91.5	79.6	96.4	96.5	86.2	97.3	97.8	79.7	94.6	97.5	81.5
	Move within country	8.0	7.9	29.4	17.0	16.4	43.0	7.8	7.8	15.7	3.3	3.2	12.4	2.5	2.1	19.8	5.3	2.4	18.3
	Move from outside country	0.5	0.5	5.1	1.6	1.5	9.6	0.7	0.6	3.6	0.2	0.2	0.7	0.1	0.1	0.1	0.0	0.0	0.0
	Not stated	0.1	0.1	0.5	0.2	0.1	0.3	0.1	0.1	1.0	0.1	0.1	0.6	0.1	0.0	0.4	0.1	0.1	0.2
	Not applicable	6.8	6.8	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Male	Residence unchanged	84.6	84.8	59.5	83.3	84.4	52.2	90.1	90.3	68.9	96.0	96.2	81.6	97.4	97.7	80.2	95.7	97.8	76.5
	Move within country	7.9	7.9	29.1	14.8	14.1	37.2	8.9	8.8	20.3	3.6	3.5	14.4	2.4	2.1	19.0	4.2	2.1	23.2
	Move from outside country	0.5	0.5	4.8	1.6	1.4	6.9	0.8	0.8	3.6	0.3	0.3	0.9	0.2	0.2	0.2	0.1	0.1	0.1
	Not stated	0.1	0.1	0.6	0.3	0.2	3.7	0.2	0.1	7.2	0.1	0.1	3.1	0.1	0.0	0.6	0.1	0.0	0.2
	Not applicable	6.8	6.8	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Chapter 3

Undercoverage of the elderly institutionalized population: The risk of biased estimates and the potentials of weighting

This chapter was published as follows: Schanze, Jan-Lucas and Stefan Zins. 2019. "Undercoverage of the elderly institutionalized population: The risk of biased estimates and the potentials of weighting" In: *Survey Methods: Insights from the Field*. Retrieved from https://doi.org/10.13094/SMIF-2019-00017

3.1 Introduction

Large national and cross-national survey programs aim to enable their users in science and the wider public to draw conclusions for the so-called *general population*, which comprises all residents living in a certain country or region. However, due to practical and financial reasons, surveys usually do not cover the entire population and deliberately exclude certain groups. In round 8, the *European Social Survey* (ESS) defined its *target population* as "[a]ll persons aged 15 and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language" (ESS Sampling Expert Panel 2016, 5). This definition excludes children, the homeless population, and the population living in institutions, such as prisons, retirement and nursing homes, or refugee accommodations. As another example, the *Health Survey for England* (HSE) "was designed to be representative of the population living in private households in England. Those living in institutions were outside the scope of the survey" (Craig et al. 2015, 13).[1]

[1] See Schnell for a description of excluded groups in German surveys (1991).

The deliberate exclusion of institutionalized residents can be explained with pragmatic reasons on the basis of cost-benefit analyses. Following Tourangeau's classification of hard-to-survey populations (2014), elderly persons living in institutions can be classified as hard-to-reach and potentially hard-to-interview (Feskens 2009; Schanze 2017). Gatekeepers, such as staff working in institutions or relatives, might prevent interviewers from getting to the respondents because they aim to protect them (see Neuert et al. 2016). Moreover, as the summary of research findings in this paper shows elderly institutionalized residents are more likely to be older and suffer from various non-cognitive and cognitive impairments, which renders them more difficult to be interviewed with standard questionnaires (see Sangl et al. 2007).

As long as it can be assumed that excluded parts of the population do not differ to a significant extent from the included survey population, the restricted definition of the target population would not have any negative impact on the generalizability of survey results. This study tests this assumption with regard to the elderly institutionalized population. It is motivated by the following overarching research questions: To avoid coverage bias, do survey programs need to make additional efforts to extend their coverage to the institutionalized population? And secondly, to what extent do survey weights help to counterbalance potential bias due to noncoverage and undercoverage?

Bias in survey results can be driven by two factors: the size of the institutionalized population and the distinctiveness of this population with respect to any variable of interest (Groves et al. 2009; Lessler and Kalsbeek 1992). Regarding the size of the institutionalized population, the latest European census in 2011 counted a cross-national share of 1.3% of institutionalized residents (Eurostat 2015, 45).[2] This proportion gets larger in the age cohorts older than 50 years (1.9%), older than 65 years (3.3%), and in the oldest age cohorts older than 80 years (8.5%).[3] The majority of the institutionalized elderly lives either in health care institutions or retirement and nursing homes (Eurostat 2015). In 2011, this group comprised nearly 2.7 million people and was by far the largest group within the European institutionalized population (ibid.).

As the second factor of coverage bias, the statistical distinctiveness of the institutionalized population comes into play. Moving to a retirement or nursing home implies a strong (self-)selection mechanism. Previous research in the fields of gerontology, medicine, and public health has identified a number of variables that differed significantly between the community-dwelling population living in private households and the institutionalized population (see Section 3.3). These variables measured socio-demographic, socio-economic, and medical characteristics and are potentially sensitive to coverage bias when institutionalized residents are not covered or undercovered. These differences limit the capacity to make an inference to the whole population, if estimates are based on data that excludes the institutionalized population.

Coverage bias is difficult to quantify, since information about the noncovered population usually is missing (Lessler and Kalsbeek 1992). Our approach uses the data from the *Survey of Health,*

[2]This figure also contains a small number of homeless residents. Excluding them from the Eurostat data with the publicly available census hub tool does not change the overall share of 1.3% (see Eurostat 2016).

[3]Own calculations using the census hub homepage (Eurostat 2016). We used the category "Occupants living in a collective living quarter" of a variable measuring housing arrangements (HC39).

Ageing and Retirement in Europe (SHARE) to run a Monte Carlo simulation. In Europe, SHARE is the only large cross-national survey that has covered the elderly institutionalized population in a comprehensive way (see Schanze 2017). In our simulation, we do not manipulate the statistical distinctiveness of institutionalized residents compared to community-dwelling residents, since this information is contained in the SHARE data. Instead, we alter the coverage rates of institutionalized residents and simulate noncoverage and different degrees of undercoverage. We use the Monte Carlo simulation to detect a possible bias in sample estimates. Moreover, we assess the possibilities and limitations of using survey weights to compensate for the exclusion and inadequate inclusion of the institutionalized population. Our research could be of interest to survey researchers and researchers working with survey data, especially on health or aging.
This paper is structured as follows. In the next section we introduce the concept of coverage bias and also elaborate on the idea of counterbalancing the bias with sample weights. The third section summarizes the previous research on the statistical distinctiveness of the population living in institutions for the elderly. Following this summary, we advance our hypotheses in the fourth section. The fifth section describes the survey data we used and how we processed the data to compile an empirical population for our simulation. In the sixth section, we explain our Monte Carlo simulation approach. This section also describes the weights we applied to the samples. In the seventh section, we present and discuss the results of the Monte Carlo simulation for different conditions of coverage and different weighting schemes. Finally, we draw our conclusion in the last section.

3.2 The concept of coverage bias

Analyses of the impact of undercoverage or noncoverage in given survey data are confronted with uncertainty because data on the undercovered parts of the target population is missing by definition (Lessler and Kalsbeek 1992). Very few studies have examined the possible bias if elderly institutionalized respondents are excluded from social surveys. Using a Swedish panel of the population aged 77 years or older, Kelfve and colleagues concluded that limitations in ADL, restriction of mobility, and psychological problems would be significantly underestimated if 22% of the institutionalized sample units were left out (2013). A Dutch pilot survey of elderly homes and nursing homes found that institutionalized residents had more physical limitations and were in poorer health (Feskens 2009). Since the results were based on a small sample of less than 300 institutionalized respondents, the author concluded cautiously that "excluding residents of elderly and nursing homes may bias survey estimates about the elderly on important population characteristics" (ibid., 95).
The literature on undercoverage mainly covers the issue of frame imperfections, when elements of the target population are excluded from the sampling frame. In the case of institutionalized populations, noncoverage does not arise from frame imperfections. Institutionalized residents are excluded deliberately from most social surveys, which in general leads to biased results. An exclusion from a survey can be organized in two ways. Either the institutionalized residents are cut-off

from the sampling frame (see Särndal et al. 1992, 531), or they remain in the sampling frame and are exempt from measurement if sampled. The two procedures have different implications for statistical inference, in particular for the sampling variance, but their estimation bias can be the same. In the following, we only focus on the case of cut-off sampling.

Suppose we have a population of N residents who are fully enumerated, so we can create a set of unique indices in which each indice belongs to one element of the population and one element only. This set of indices $\mathcal{U} = \{1, \ldots, i, \ldots, N\}$ is our sampling frame, where i is the indice associated with the i-th person in the population. We denote the set of indices of the institutionalized population as $\mathcal{U}_I$. The set of indices of the non-institutionalized population is then given by $\mathcal{U}_p = \mathcal{U} \setminus \mathcal{U}_I$.

If our variable of interest is a real valued and positive variable y, in cut-off sampling its total for the private population $\tau_{p,\,y} = \sum_{i \in \mathcal{U}_p} y_i$ can be estimated by $\hat{\tau}_{p,\,y} = \sum_{i \in s_p} \frac{y_i}{\pi_i}$, where s_p is our sample from $\mathcal{U}_p$, i.e $s_p \subset \mathcal{U}_p$, y_i is the observation of variable y for the i-th element in the frame, and π_i is the probability for including the i-th person in the sample from $\mathcal{U}_p$. If we are interest in estimating the total of the entire population $\tau_y = \sum_{i \in \mathcal{U}} y_i$, then $\hat{\tau}_{p,\,y}$ is biased for τ_y if $\hat{\tau}_{I,\,y} = \sum_{i \in \mathcal{U}_I} y_i \neq 0$. To reduce a potential bias, a ratio estimator can be used instead of $\hat{\tau}_{p,\,y}$. If we know the population total τ_x of a real valued and positive auxiliary variable x, that is also measured in the survey, we are able to construct the following estimator:

$$\hat{\tau}_{p,\,y,\,\text{ra}} = \tau_x \frac{\hat{\tau}_{p,\,y}}{\hat{\tau}_{p,\,x}} \;,$$

where $\hat{\tau}_{p,\,x} = \sum_{i \in s_p} \frac{x_i}{\pi_i}$. Estimator $\frac{\hat{\tau}_{p,\,y}}{\hat{\tau}_{p,\,x}}$ is consistent for $R_p = \frac{\tau_{p,\,y}}{\tau_{p,\,x}}$ with $\tau_{p,\,x} = \sum_{i \in \mathcal{U}_p} x_i$. Thus, an approximation to the expected value of $\hat{\tau}_{p,\,y,\,\text{ra}}$ is $\tau_x R_p$.

The bias of estimator $\hat{\tau}_{p,\,y,\text{ra}}$ is given by:

$$B(\hat{\tau}_{p,\,y,\text{ra}}) = \tau_x R_p - \tau_y = (\tau_{x_p} R_p - \tau_{y_p}) + (\tau_{x_I} R_p - \tau_{y_I})$$

and $B(\hat{\tau}_{y,\text{ra}}) = 0$ if $R = \frac{\tau_x}{\tau_y} = R_p$, if the ratio between auxiliary variable x and y is the same in the private population and the entire population (see Särndal et al. 1992, 532).

Another class of estimator that could be used to reduce the potential bias of $\hat{\tau}_{p,\,y}$ would be the regression estimator. In this case, we use known totals of J auxiliary variables, given by $\sum_{i \in \mathcal{U}} x_{ij} = \tau_{x_j}\ j = 1, \ldots, J$, where x_{ij} is the value of the j-th auxiliary variable of the i-th person in the sampling frame. We denote the vector of the auxiliary variables for the i-th person as $\mathbf{x}_i = (x_{i1}, \ldots, x_{iJ})'$. The general regression estimator for τ_y based on s_p can be written as:

$$\hat{\tau}_{p,\,y,\,\text{re}} = \hat{\tau}_{p,\,y} + \sum_{j=1}^{J} \left(\tau_{x_j} - \hat{\tau}_{p,\,x_j}\right) \hat{B}_{p,\,j}$$

where $\hat{\tau}_{p,\,x_j} = \sum_{i \in s_p} \frac{x_{ij}}{\pi_i}$, and $\hat{B}_{p,\,j}$ is the j-th component of the vector of regression coefficients

(see Särndal et al. 1992, 225):

$$\hat{\mathbf{B}}_p = \left(\hat{B}_{p,1},\, \ldots,\, \hat{B}_{p,J}\right)' = \left(\sum_{i \in s_p} \frac{\mathbf{x}_i \mathbf{x}_i'}{\pi_i}\right)^{-1} \sum_{i \in s_p} \frac{\mathbf{x}_i y_i}{\pi_i}$$

If $\hat{\mathbf{B}}_p$ is unbiased for $\mathbf{B} = \left(\hat{B}_1,\, \ldots,\, \hat{B}_J\right)' = \left(\sum_{i \in \mathcal{U}} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \sum_{i \in \mathcal{U}} \mathbf{x}_i y_i$, the regression coefficients of the entire population, then any bias in $\hat{\tau}_{p,y}$ would be corrected by the second term in $\hat{\tau}_{p,y,\text{re}}$.

The regression estimator $\hat{\tau}_{p,y,\text{re}}$ also can be written as:

$$\hat{\tau}_{p,y,\text{re}} = \sum_{i \in s_p} \left[1 + \left(\vec{\tau}_x - \hat{\vec{\tau}}_{p,x}\right) \hat{\mathbf{T}}^{-1} \mathbf{x}_i\right] y_i$$

where $\vec{\tau}_x = (\tau_{x_1},\, \ldots,\, \tau_{x_J})'$, $\hat{\vec{\tau}}_{p,x} = (\hat{\tau}_{p,x_1},\, \ldots,\, \hat{\tau}_{p,x_J})'$, and $\hat{\mathbf{T}} = \sum_{i \in s_p} \frac{\mathbf{x}_i \mathbf{x}_i'}{\pi_i}$. If we define $g_i = 1 + \left(\vec{\tau}_x - \hat{\vec{\tau}}_{p,x}\right) \hat{\mathbf{T}}^{-1} \mathbf{x}_i$, we can express $\hat{\tau}_{p,y,\text{re}}$ as:

$$\hat{\tau}_{p,y,\text{re}} = \sum_{i \in s_p} g_i \pi_i^{-1} y_i = \sum_{i \in s_p} w_i y_i$$

where $w_i = g_i \pi_i^{-1}$. This expression of the regression estimator highlights its similarity to calibration estimators. With respect to calibration estimators, the so-called design weights, given by π_i^{-1}, are adjusted by a g_i such that $\sum_{i \in s_p} = g_i \pi_i^{-1} \mathbf{x}_i = \vec{\tau}_x$. This adjustment is done so g_i's are as small as possible, while still fulfilling the calibration condition $\sum_{i \in s_p} = g_i \pi_i^{-1} \mathbf{x}_i = \vec{\tau}_x$. If as a distance measure between $g_i \pi_i^{-1}$ and π_i^{-1} the chi-square distance is used, the calibration method leads to the regression estimator (see Särndal 2007, 106). Thus, the condition that ensures that $\hat{\tau}_{p,y,\text{re}}$ is unbiased, i.e., the expected value of $\hat{\mathbf{B}}_p$ is equal to $\mathbf{B}$, is the same that ensures that the calibration estimator is unbiased.

3.3 Research findings: How do the elderly in institutions differ?

For more than two decades, a growing body of scientific literature in the fields of gerontology and public health have examined the elderly institutionalized population. Various variables have been identified as predictors of transitions from private households to retirement homes and nursing homes. Moreover, comparative analyses of community-dwelling elderly and elderly living in institutions have shown significant differences in various respects. All these studies offer first insights into the statistical distinctiveness of the institutionalized population. The following paragraphs provide a summary of previous research findings. In this section and in Table 3.2 and Table 3.3 in the Appendix, we list only the variables that explain institutionalization in multivariate models

that control for different confounding variables. If a certain variable explains institutionalization or differs between the community-dwelling population and the institutionalized population, it possibly is subject to coverage bias in social surveys whenever institutionalized residents are excluded from these surveys.

Demographic variables

Demographic and socio-economic variables are very stable, unchanging characteristics of individuals, which is why they belong to the group of predisposing factors for the need of health care (Andersen 1995). They are a first group of strong explanatory factors for the institutionalization of the elderly (see Table 3.2 for an overview of variables and studies). Across countries and survey designs, nearly all studies found a positive effect of increasing *age* on institutionalization (e.g., Angelini and Laferrère 2012; Castora-Binkley et al. 2014; Einio et al. 2012; Gaugler et al. 2007; Laferrère et al. 2012; Luppa et al. 2010b; Maxwell et al. 2013; McCann et al. 2012; Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015). The effect of *gender* on the likelihood of institutionalization is inconclusive in literature. While some studies conclude that *female gender* increases the likelihood of institutionalization (e.g., Bravell et al. 2009; Kasper et al. 2010; McCann et al. 2012), other studies come to the opposite conclusion, and found a positive effect of *male gender* (e.g., Gaugler et al. 2007; Luppa et al. 2010b; Martikainen et al. 2009; Pot et al. 2009). The contradictory results can be explained due to the influence of cross-national particularities, different study designs, different times of data collection (see Einio et al. 2012; Himes et al. 2000), and also the strong multicollinearity of gender with other explanatory variables (Einio et al. 2012; Nöel-Miller 2010).

Socio-economic variables

In several studies, socio-economic variables also belong to the group of predisposing factors with significant explanatory power. Residents who *own and live in their own houses* have a lower probability of moving to institutions for the elderly (e.g., Einio et al. 2012; Gaugler et al. 2007; Luppa et al. 2010b; McCann et al. 2012; Thomeer et al. 2015). Using regional panel data from Sweden, Bravell and colleagues detected a negative correlation between a rising *socio-economic status* and the likelihood of institutionalization (2009). Some studies have found that both a higher *education* (Asakawa et al. 2009; Einio et al. 2012) and a higher *income* (e.g., Angelini and Laferrère 2012; Gaugler et al. 2007; Laferrère et al. 2012; Martikainen et al. 2009; Thomeer et al. 2015 are protective factors against institutionalization. However, two other studies have found the opposite direction of relationship for education and income in other studies. These recent studies using U.S. panel data observed a growing probability of institutionalization with a better education (Castora-Binkley et al. 2014; Thomeer et al. 2015). Concerning income, Rodríguez-Sánchez and colleagues observed a lower probability of institutionalization for individuals with a low household income (2017). Differences at the macro-level with respect to the long-term care system and the care culture could help to explain some of the contradictory results with respect to socio-economic status

(Geerts and Bosch 2012; Laferrère et al. 2012; Suanet et al. 2012).

Social networks and informal care

Marital status and family networks are highly significant independent variables in many studies that aim to explain the institutionalization of the elderly. These variables work as enabling resources, which can protect residents from institutionalization (1995). Indeed, *being married* protects community-dwelling residents from transition to institutions (e.g., Castora-Binkley et al. 2014; Gaugler et al. 2007; Luppa et al. 2010b; Rodríguez-Sánchez et al. 2017) compared to larger odds of institutionalization for residents who are *not married* (e.g., Asakawa et al. 2009; Bravell et al. 2009; Thomeer et al. 2015), *widowed* (Angelini and Laferrère 2012; Einio et al. 2012; Nöel-Miller 2010; Thomeer et al. 2015), or *divorced* (Thomeer et al. 2015). As a consequence, *living alone* (e.g., Gaugler et al. 2007; McCann et al. 2012; Pimouguet et al. 2016) without a *partner in the household* (Désesquelles and Brouard 2003; Laferrère et al. 2012) increases the likelihood of institutionalization, whereas *co-residence with a partner and/or children* decreases the likelihood of institutionalization (Angelini and Laferrère 2012; Kasper et al. 2010; McCann et al. 2012; Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015). Apart from a partner, some studies also observed a negative effect of *having (grand-)children* (e.g., Kasper et al. 2010; Nöel-Miller 2010; Rodríguez-Sánchez et al. 2017) on institutionalization. Marital status and family networks, as well as additional social networks are closely linked to availability of informal care in private households. Smaller *social networks* (Hays et al. 2003; Luppa et al. 2010b; Maxwell et al. 2013) increased the probability of institutionalization.

Health-related variables

Health-related variables are strong predictors of institutionalization across all kinds of study designs within different countries (see Table 3.3 for an overview of significant explanatory variables). First, *cognitive impairments* (e.g., Castora-Binkley et al. 2014; Einio et al. 2012; Gaugler et al. 2007; Luppa et al. 2010b; Maxwell et al. 2013; Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015; Toot et al. 2017) increase the probability that community-dwelling residents have to move to institutions for the elderly, for example, due to the prevalence of *dementia* (e.g., Einio et al. 2012; Laferrère et al. 2012; Luppa et al. 2010b; Nihtilä et al. 2008; Toot et al. 2017. In addition to dementia, a number of *medical conditions*, especially if they occur simultaneously, also are associated with a higher likelihood of institutionalization (e.g., Angelini and Laferrère 2012; Einio et al. 2012; Gaugler et al. 2007; Luppa et al. 2010b; Maxwell et al. 2013; Rodríguez-Sánchez et al. 2017; Toot et al. 2017). Turning from the objective measures of health to a measurement of subjective health, a bad *self-rated health* also increases the likelihood of institutionalization according to some studies (Castora-Binkley et al. 2014; Einio et al. 2012; Hancock et al. 2002; Luppa et al. 2010b; McCann et al. 2012; Nöel-Miller 2010).

Autonomy and mobility

Often, following a critical state of health, an elderly person's ability to cope with autonomous daily living is diminished. With respect to the likelihood of institutionalization, studies have found a positive influence of so-called *limiting long-term illnesses* (Grundy and Jitlal 2007; McCann et al. 2012), *functional impairments* (Luppa et al. 2010b; Maxwell et al. 2013; Pot et al. 2009), *mobility difficulties* (Hays et al. 2003; Thomeer et al. 2015; Toot et al. 2017; Von Bonsdorff et al. 2006), *motor limitations* (Angelini and Laferrère 2012), and *physical dependency* (Désesquelles and Brouard 2003). Closely related to the concept of physical dependency, many studies have examined limitations in basic activities of daily living (ADL) and instrumental activities of daily living (IADL). ADL includes limitations in walking, dressing, eating, bathing, going to the toilet, and getting in and out of bed. IADL measures limitations with respect to managing money, preparing meals, getting groceries, using the telephone, and taking medications. Studies have found that both *ADL* (e.g., Bravell et al. 2009; Gaugler et al. 2007; Laferrère et al. 2012; Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015; Toot et al. 2017) and *IADL* (e.g., Castora-Binkley et al. 2014; Laferrère et al. 2012; Thomeer et al. 2015; Toot et al. 2017) often are very strong predictors of institutionalization.

Previous research suffers from two limitations. First, many studies relied on regional data in the absence of national data (see Agüero-Torres et al. 2001; Bravell et al. 2009; Hancock et al. 2002; Hays et al. 2003; Maxwell et al. 2013; Pimouguet et al. 2016; Riedel-Heller et al. 2000; Von Bonsdorff et al. 2006). Second, analyses are more difficult to carry out due to the very small number of institutionalized residents. The small number of respondents might also serve as an explanation for some of the contradictory results.

Our study contributes to the current research by testing the impact of noncoverage and undercoverage with a larger dataset of institutionalized respondents. Moreover, we distinguish different degrees of undercoverage and also assess whether survey weights can counterbalance a potential coverage bias. To our knowledge, it is the first simulation analysis of the bias caused by the exclusion of the elderly institutionalized population in social surveys.

3.4 Hypotheses

Following from the previous section, the hard-to-survey population living in institutions for the elderly differs in many respects from their community-dwelling counterparts. They have a worse objective and subjective state of health and are more often suffering from cognitive and functional impairments. They are older, have a different marital status, and have lived alone more frequently prior to institutionalization. Considering the statistical distinctiveness of the institutionalized population described in the previous section, it is our overarching hypothesis that bias arises when institutionalized respondents are excluded from the sample. We tested two different variables be-

cause bias affects different variables to a different extent. On the basis of the literature, we chose two health-related variables.
If institutionalized residents are not covered adequately in survey samples, we expect an underestimation of a variable measuring the limitations in ADL. This underestimation is expected to have a visible impact on the estimated mean and on the estimated share. For our second dependent variable, self-rated health, we assume an overestimation of the mean in our samples compared to the overall statistical population. The share of respondents with poor self-rated health is assumed to be underestimated.
The bias might be reduced by applying calibration weights when analyzing the sample data drawn from sampling frames with noncoverage or undercoverage, as described in Section 3.2. Kelfve and colleagues weighted the samples without institutionalized respondents for age and gender, and concluded that the weighting did not improve the estimation substantially since the bias persisted (2013). We also tested whether weights for age and gender improve the point estimates of samples that suffer from noncoverage and undercoverage. In addition to traditional survey weights, we tested weights calibrated on age, gender, and institutionalization. Those weights can be used by surveys that try to cover the hard-to-survey population living in institutions but might suffer from undercoverage. In case of noncoverage, the latter weights cannot be used. When no institutionalized respondents are in the sample, it is not possible to calibrate on the variable measuring institutionalization.
For both types of weights, we expect a decreasing bias. We assume that the marginal positive effect of the survey weights calibrated on age and gender will be much smaller than that for the extended survey weights, which also account for the undercovered population. Our focus for the analysis is on the interplay of weights with an decreasing undercoverage of institutionalized respondents. We assume that weights cannot fully eliminate the bias under the condition of noncoverage, but they will yield better results with an increasing coverage rate.

3.5 Data

For our simulation-based analysis, we did not compute a synthetic population; instead, we used the survey data from the *Survey of Health, Ageing and Retirement in Europe* (SHARE) to construct our empirical population. SHARE is a cross-national panel survey, which collects its data in the face-to-face mode and uses strict quality standards and a harmonization across the participating countries preceding the data collection (Börsch-Supan et al. 2013). Moreover, to our best knowledge, SHARE is the only cross-national social survey that includes institutionalized respondents to a significant extent.
With the exception of the third wave, we used all six waves of SHARE data collected in a number of European countries and Israel, approximately every 2 years between 2004 and 2015 (Börsch-Supan 2017a,b,c,d,e). In the first wave in 2004, the SHARE target population was defined as all households with "at least one member born in 1954 or earlier, speaking the official language of

the country and not living abroad or in an institution such as a prison during the duration of field work" (Klevmarken et al. 2005, 30). However, in addition to this minimum definition of the target population, residents of institutions for the elderly also are officially part of the SHARE target population and shall be sampled and interviewed if possible (De Luca et al. 2015; Klevmarken et al. 2005; Lynn et al. 2013). SHARE also uses proxy interviews (Börsch-Supan et al. 2013), which is especially important with respect to hard-to-interview groups like institutionalized residents.

Before describing our methods, we need to mention one limitation of the data we used. According to the documentation of SHARE, institutionalized residents suffer from undercoverage. A number of countries reported in their sampling design forms that it was not possible to sample institutionalized respondents in the first wave and in refreshment samples because they were excluded from the sampling frames (De Luca et al. 2015; Klevmarken et al. 2005; Lynn et al. 2013). In the first wave, this was the case in Austria, France, Greece, Italy, and Switzerland (Klevmarken et al. 2005). In contrast, an equal number of countries reported that their target population included institutionalized residents in the first wave (ibid.).[4] Due to these frame imperfections, the first group of countries does not cover the entire institutionalized population, since it includes only the community-dwelling respondents who moved to institutions between two survey waves. As a consequence, the sample of institutionalized respondents suffers from undercoverage and could be biased.[5] However, even within those countries that reported a noncoverage of institutionalized residents in the baseline wave we identified institutionalized respondents (Table 3.4 in Appendix B).[6] Nevertheless, the undercoverage diminishes the potential to generalize our simulation results and our conclusions are restricted to our empirical population.

Given the small share of institutionalized respondents in the separate waves, we cumulated five of the six waves of SHARE and only excluded the third wave of the panel because its content differs largely from all the other waves.[7] For every panel respondent, we only retained the most recent interview in our dataset. However, if a respondent lived in an institution and moved back to a private household in a subsequent wave, we dropped the more recent interview conducted in the private household.[8] We include every respondent who was interviewed between 2004 and 2015 by SHARE with only one observation in our dataset. Moreover, in the pooled dataset we dropped six countries with less than 50 interviews in institutions[9], but still cover most parts of Europe and Israel (see Table 3.4 in Appendix B) We also dropped all respondents younger than 50 years at the point of data collection. These respondents are only eligible as a partner of a SHARE respondent

[4] Denmark, Germany, Netherlands, Spain, and Sweden

[5] As a matter of fact it will be representative of the institutionalized population in the very long run, at the moment when the population does not include any institutionalized citizens that have been institutionalized *before* the first wave of the panel was recruited (Lynn 2011). This assumption only holds true if the panel survey does not draw any biased refreshment samples in the meantime, and if panel attrition does not lead to a stronger decrease of panel members from the institutionalized populations.

[6] See also De Luca et al. for a similar observation of a mismatch of country reports about noncoverage and the de facto presence of nursing home residents in the SHARE data (2015: 78).

[7] SHARELIFE collected retrospectively information on life events of the SHARE respondents (see Börsch-Supan and Schröder 2011). Many variables used in this analysis were not part of the questionnaire in SHARELIFE.

[8] This rule only affected a small number of 508 respondents, adding up to 0.2% of the entire pooled sample before dropping duplicate observations across waves.

[9] Croatia, Hungary, Ireland, Poland, Portugal, and Slovenia

and are not a representative sample of the population younger than 50 years.
In total, we obtained a dataset with 100,595 observations, among these 2,514 nursing home interviews, which amounted to 2.5% of our empirical population. Due to missing values in some of the variables we used in our analysis, we dropped another 1,929 cases (1.9% of the pooled sample) and obtained our final statistical population of 98,666 cases, among these 2.47% institutionalized residents (N_i = 2,441).
A comparison of the complete cases with the excluded cases shows that item nonresponse occurs more often in interviews with institutionalized respondents (3.8% of the excluded respondents) and with proxy respondents (10.6% compared to 6.6% of proxy interviews without item nonresponse). Moreover, the excluded population is older than the included population (69.3 years compared to 67.9 years), is more often widowed (19% compared to 15.7% of the included units), has more limitations in activities of daily living (0.67 compared to 0.46), and feels less healthy (2.66 compared to 2.81). Very small differences between included and excluded respondents occur in terms of income, the number of children, and education. In addition, the two populations do not differ significantly in the distribution of gender.

Variables of interest

Since we used several waves of a panel survey in a cross-sectional analysis, we had to make sure that all the variables we used in our analysis have been measured in the same way for all the five waves (see Appendix B for the main variables). To assess the statistical impact of an exclusion of institutionalized residents, we analyzed two different variables that have been salient for the institutionalized population in previous research: an index of the limitations in activities of daily living (ADL) and self-rated health.
Our first dependent variable measures the dependency of respondents and evaluates their ability to run a household and cope with daily living. Respondents were provided with a showcard that listed various activities of daily living (ADL). They answered which activities they could not execute because of "a physical, mental, emotional or memory problem." For our analysis, we used six items classified by SHARE as ADL items and created a 9-level additive index with three additional dummy variables that belong to the group of items that measure the limitations in instrumental activities of daily living (IADL). These nine items displayed a high tau-equivalent reliability (Cronbach's α = 0.89) and loaded clearly on a single factor in an exploratory factor analysis. Our first dependent variable is a count variable with a very large number of respondents with an outcome of 0 (zero-inflated, see Table 3.1).
As our second dependent variable we analyzed self-rated health. This variable measured the subjective health of respondents, which implicitly encompasses physical and mental health. SHARE used the U.S. version of this variable to measure self-rated health on a 5-level scale in all five waves.

Institutionalization

Long-term institutionalization was the primary variable of interest in our analysis. We relied on three different variables to generate a dummy variable that captured whether a respondent lived in a private household or in an institution during the interview. Since the second wave, SHARE interviewers registered whether a sampled address was a private household or a nursing home when they arrived at the address. According to the SHARE codebook[10], a *nursing home* "provides all of the following services for its residents: dispensing of medication, available 24-hour personal assistance and supervision (not necessarily a nurse), and room & meals".[11] In addition to this process-generated variable, we also coded all the respondents as institutionalized residents, who reported to have lived permanently in an institution during the last year.[12] As a third variable, we used the type of building in which a respondent lives. If respondents either answered this question with category 7 "*a housing complex with services for elderly*" or 8 "*special housing for elderly (24 hours attention)*", we coded these cases as being institutionalized.[13]

3.6 Monte Carlo Simulation

In this section, we introduce our simulation-based analytical approach and the different coverage conditions and weights that we applied in the Monte Carlo simulation.

To evaluate the bias of the different estimation strategies of different coverage scenarios of the institutionalized population, we repeatedly selected samples from our empirical population. The sampling design of the simulations study is a stratified sample, in which the 15 countries in our empirical population serve as the strata. Within the strata, we selected individuals using a simple random sample. The sampling fraction is always 3% if the sampling frame and the allocation of the sample size is done proportionally to the number of persons within the strata. Thus, all persons in the sampling frame have an almost equal inclusion probability of 0.03 (i.e., ignoring the rounding problem of the actual allocation).

We considered five different coverage scenarios, and we constructed five different sampling frames to include 0%, 25%, 50%, 75%, and 100% of the institutionalized population. Thus, as our sampling frames were $\mathcal{U}_{p\,I_\alpha} = \mathcal{U}_p \cup \mathcal{U}_{I_\alpha}$, with $\alpha = 0,\ 0.25,\ 0.5,\ 0.75,$ and 1, where $\mathcal{U}_{I_\alpha}$ is a simple random sample from $\mathcal{U}_I$ with sampling fraction α. To construct $\mathcal{U}_{I_\alpha}$, $\mathcal{U}_I$ is portioned into four disjoint random subsets, each of size $N_I * 0.25$. $\mathcal{U}_{I_\alpha} = \bigcup_{a=(0,0.25,\ldots,\alpha)} \mathcal{U}_{I_a}$, with $\mathcal{U}_{I_0} = \emptyset$. From each of the five sampling frames, we selected 1000 samples with the sampling design described above. In the following, we denote α as the coverage rate of the institutionalized population in the

[10] Generic CAPI coverscreen, can be downloaded at http://www.share-project.org/data-documentation/questionnaires.html

[11] 1,424 institutionalized respondents were identified as institutionalized residents by the interviewers in our final empirical population.

[12] We added 326 respondents to the group of institutionalized residents based on their reply to this variable. These respondents were classified as community-dwelling respondents by the interviewers. From the second wave onward, respondents were usually not asked this question if interviewers identified their place of living as an institution.

[13] Additional 691 respondents were classified as institutionalized according to their replies to this questions.

sampling frame.

Estimators

Our statistics of interest are the means and proportions of the categories of our self-constructed ADL variable and self-rated health. To estimate these means, we employed three different estimators. The first estimator is:

$$\overline{y}_{I_\alpha, d} = \sum_{i \in s_{p,\, I_\alpha}} \frac{y_i d_i}{\sum_{i \in s_{p,\, I_\alpha}} d_i},$$

where $s_{p,\, I_\alpha}$ is a sample from frame $\mathcal{U}_{p\, I_\alpha}$, d_i is the design weight given by $d_i = \pi_i^{-1}$, and y_i is the measurement or category indicator for one of the variables of interest for the i-th person in the sampling frame.

For our second estimator, we used a regression estimator as described in Section 3.2 to estimate the total of our variable of interest, which we divided by the sum of the regression estimator or calibration weights. Thus, we produced the following estimator:

$$\overline{y}_{I_\alpha, \mathrm{re}_1} = \sum_{i \in s_{p,\, I_\alpha}} \frac{y_i w_{i,\,1}}{\sum_{i \in s_{p,\, I_\alpha}} w_i},$$

where $w_i = g_i d_i$, and g_i is constructed as described in Section 3.2. Like many other social surveys, SHARE uses mainly region, age, and gender as auxiliary variables for constructing cross-sectional calibration weights at the level of individuals (De Luca et al. 2015). For $\overline{y}_{p\alpha, \mathrm{re}_1}$, we also used age in 5 categories and gender and their known totals over all element in frame $\mathcal{U}$ as auxiliary variables. This estimation strategy reflects the situation where a survey with non or partial coverage of the institutionalized population uses aggregate data for weighting purposes, that covered the entire population including institutionalized residents (e.g., census data).

For our third estimator $\overline{y}_{I_\alpha, \mathrm{re}_2}$, we added institutionalization to the survey weights described above. By weighting for the hard-to-survey domain that suffers from undercoverage, we aim to examine whether this approach is a feasible option for surveys to counterbalance their insufficient coverage. However, this approach does not work for survey samples without any institutionalized residents.

For each of the 1,000 samples from each of the five coverage scenarios, we estimated for our two variables of interest the proportions and the overall mean using all three estimators. Additionally, we calculated the 95% confidence intervals (CI) of all the point estimates for each of the samples. The confidence intervals are constructed as follows:

$$\hat{0} \pm 1.96\sqrt{\hat{V}(\hat{0})}$$

where $\hat{0}$ is any of our point estimates under consideration and $\hat{V}(\hat{0})$ is its variance estimate. The confidence intervals enabled us to check whether the deviations between a sample estimate and the true share or true mean in our empirical population were significant enough to be defined as

biased. Every confidence interval that missed the true value was marked as rejected. If there was no bias at all, the expected proportion of rejected samples would be 5%.

3.7 Results

Table 3.1 provides a description of our empirical population for some variables, and compares the institutionalized respondents and the community-dwelling respondents. The table reveals considerable differences between the two domains. In our empirical population, more women than men live in institutions, whereas the community-dwelling population is more gender balanced. The institutionalized population is more than 12 years older on average.

Table 3.1: Description of Our Empirical Population and Comparison of the Two Populations

		Total population	**Private households**	**Institutions**	
		%	%	%	Difference
Gender	Male	45.3	45.5	34.4	11.1^{***}
	Female	54.7	54.5	65.6	-11.1^{***}
Age	50 to 59 years	25.4	25.9	7.3	18.7^{***}
	60 to 69 years	33.1	33.7	12.3	21.3^{***}
	70 to 79 years	25.1	25.2	21.6	3.6^{***}
	80 to 89 years	13.9	13.3	39.3	-26.0^{***}
	Older than 90 years	2.4	2.0	19.5	-17.5^{***}
	Mean value	*67.9*	*67.6*	*79.7*	*-12.1****
Limitations with ADL	No limitations	84.3	85.4	41.3	44.1^{***}
	1 to 3 limitations	11.1	10.8	25.4	-14.7^{***}
	More than 3 limitations	4.6	3.8	33.3	-29.5^{***}
	Mean value	*0.46*	*0.40*	*2.71*	*-2.3****
Self-rated health	Poor	11.9	11.5	27.4	-15.9^{***}
	Fair	27.8	27.6	35.3	-7.7^{***}
	Good	35.3	35.5	27.4	8.2^{***}
	Very good	17.6	17.9	6.7	11.2^{***}
	Excellent	7.3	7.4	3.2	4.2^{***}
	Mean value	*2.81*	*2.82*	*2.23*	*0.6****
Number of observations		98,666	96,225	2,441	

Note: *p <0.05, **p <0.01, ***p <0.001; t-test of differences in unweighted means, chi-squared test for unweighted categories

Concerning the two dependent variables, nearly 60% of all institutionalized respondents have at least one or more limitations in activities of daily living, compared to a significantly lower share of 14.6% in private households. As a consequence, the mean values of the ADL variable differ significantly between the two populations. Regarding the second dependent variable, more respondents

in institutions reported a poor or fair state of health compared to the community-dwelling population. Only 10% of all institutionalized respondents rated their health as very good or excellent. The aggregate differences in the mean value of self-rated health are less distinct compared to the ADL variable, although they are still statistically significant.

The first column of Table 3.1 provides the values of the variables of our total empirical population, which includes community-dwelling and institutionalized respondents. A comparison of this column with the second column depicting the aggregate values for the community-dwelling population gives a first impression of the coverage bias in the case of noncoverage of the institutionalized population. Only slight differences occur in terms of gender distribution and the number of children. The community-dwelling population definitely deviates from the true values with respect to other variables. The joint share of the two oldest age cohorts differs by one percentage point. Regarding the two dependent variables, only the variable measuring limitations in ADL seems to be significantly biased. Adding the institutionalized respondents to the community-dwelling respondents decreases the share of people without any limitations in ADL from 85.4% to 84.3%. The differences in reports of self-rated health are less diverging if the community-dwelling population is compared with the total population. The results of Table 3.1 can be interpreted as a census of our empirical population. The following analyzes show how survey samples perform, if they do not cover institutionalized residents at all or cover them insufficiently.

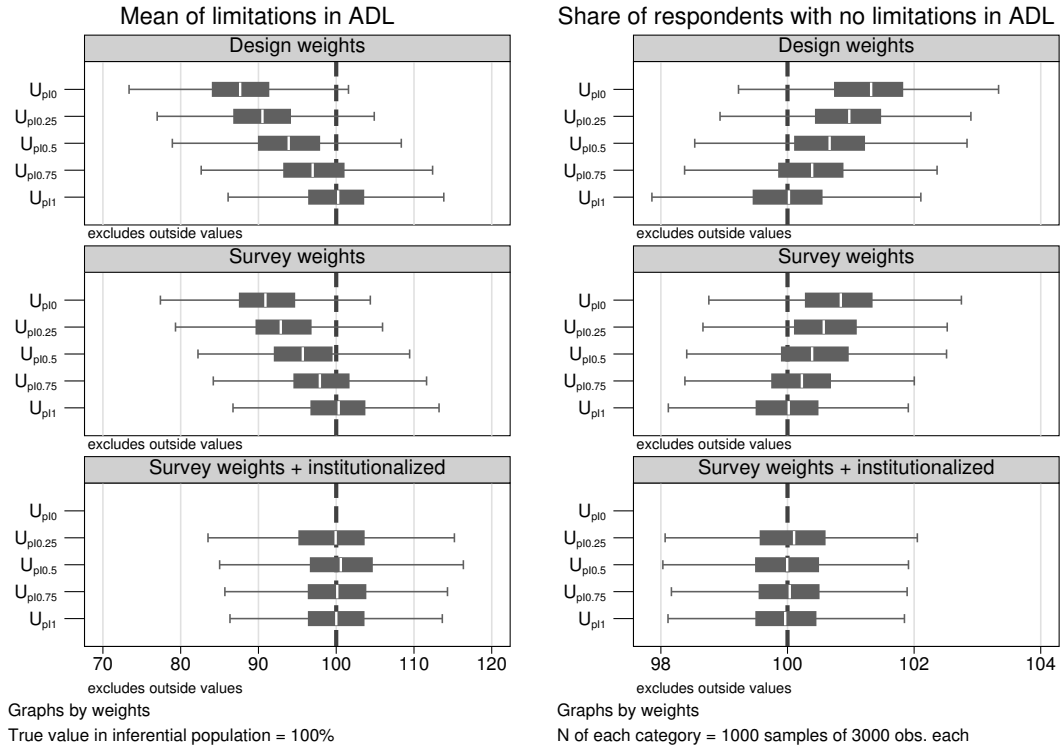
Noncoverage

Figure 3.1 and Table 3.6 (see Appendix C) provide the results of the simulation for the first dependent variable. Each box plot in Figure 3.1 contains the percentage deviation of estimates of 1,000 samples with the given coverage condition and the given weighting scheme compared to the true values. Table 3.6 provides the mean values from those 1,000 samples and the share of samples that were rejected on the basis of their confidence intervals. In the case of noncoverage, nearly two thirds of all samples failed to predict the true mean of the ADL variable for our statistical population. The estimation of the share of respondents without any limitations in ADL is less sensitive to bias. The simulation replicated the differences shown in Table 3.1, since the samples missed the true value (84.34%) by about one percentage point on average. The deviation caused a significant bias in 41.4% of the samples, which missed the true value with their confidence intervals. The graphical analysis in Figure 3.3 shows that these 414 samples estimated a share of respondents without limitations in ADL to be between 85.6% to 87.6%.

In line with our hypothesis, using estimator $\overline{y}_{I_0,\mathrm{re}_1}$ improved the precision under the condition of noncoverage ($\mathcal{U}_{p\,I_0}$), but could not eliminate the bias entirely. Both boxes move closer to the dashed lines that indicate the true values. As it can be seen in Table 3.6, using age and gender as auxiliary variables reduced the number of estimates for the mean of ADL whose CIs did not include the true mean of 42.3%. Regarding the estimated share of respondents without limitations in ADL, only 21.6% of all CIs did not include the true value. As pointed out in Section 3.6, the estimator $\overline{y}_{I_0,\mathrm{re}_2}$, which adds institutionalization to the survey weights, cannot be applied under

the condition of noncoverage, since the domain of institutionalized respondents is empty.

Figure 3.1: Relative Bias in Two Aggregated Point Estimates Calculated with Different Coverage Rates and Weighting Schemes

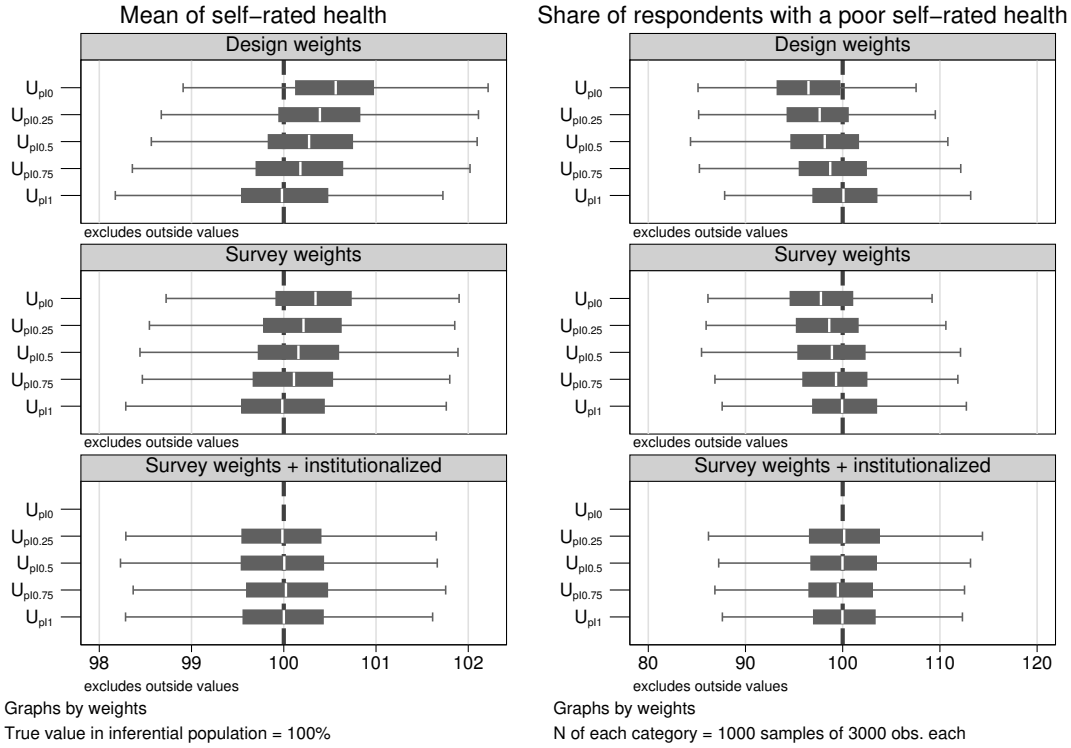


ADL is a relatively strong predictor of institutionalization and should be biased with a higher likelihood than other variables as a consequence. According to Table 3.1, self-rated health is less different between the community-dwelling population and the overall population. Indeed, our simulation confirms that the bias in estimates related to self-rated health is weaker than in those for ADL. The design-weighted estimator $\overline{y}_{I_0,\,d}$ overestimated the average self-rated health and underestimated the share of persons with poor self-rated health (see Figure 3.2). However, the estimations are closer to the true values than for ADL. 10.9% and 11.5% of all CI estimates using $\overline{y}_{I_0,\,d}$ missed the true share of poor self-rated health and the true mean of self-rated health, respectively (see Table 3.7). Although $\overline{y}_{I_0,\,\mathrm{re}_1}$ still displays bias for the statistics of ADL, it made a substantial difference in the measurement of self-rated health. Using $\overline{y}_{I_0,\,\mathrm{re}_1}$ decreased the number of CI estimates that did not include the true share of poor self-rated health to 6.4%.

The results underline the varying impact of bias on two different health-related variables and point estimates. With respect to the ADL variable, the estimation of the mean was more prone to bias than the estimation of the share of respondents without any limitations in ADL. With respect to self-rated health, the estimation of the share of poor self-rated health and the prediction of the

mean were equally sensitive to bias. This finding can be explained also by the different scales of the two variables. ADL is a count variable with a large 9-level scale and many respondents with a zero value. Every third institutionalized respondent reached a value larger than 3 on the ADL scale (see Table 3.1), and this group drove the true mean upwards. The self-rated health variable has a smaller 5-level scale, and smaller values were more prevalent among the institutionalized respondents. As a consequence, the true mean is less affected by the group of institutionalized respondents.

Figure 3.2: Relative Bias in Two Aggregated Point Estimates Calculated with Different Coverage Rates and Weighting Schemes



Undercoverage

Following the scenario of noncoverage using the sampling frame $\mathcal{U}_{p\,I_0}$, we increased the proportion of the institutionalized population that we included in the sampling frame in four steps. The frames $\mathcal{U}_{p\,I_{0.25}}$, $\mathcal{U}_{p\,I_{0.5}}$, and $\mathcal{U}_{p\,I_{0.75}}$ can be be described as situations of undercoverage. For the three frames 18 ($\alpha = 0.25$), 37 ($\alpha = 0.50$), and 55 ($\alpha = 0.75$) institutionalized persons were expected to be part of the samples. Figures 3.1 and 3.2 show the percentage deviation of estimates for each undercoverage scenario, the respective Tables in Appendix C provide the share of CI estimates that did not include the true value. Figures 3.3 and 3.4 present the distribution of all the sample

Figure 3.3: Estimated Share of Respondents **Without Any Limitations in ADL** in 1,000 Samples Per Histogram (*Light grey*: True share within 95% CI ; *Dark grey*: True share beyond 95% CI; *Dashed line*: True share in inferential population)

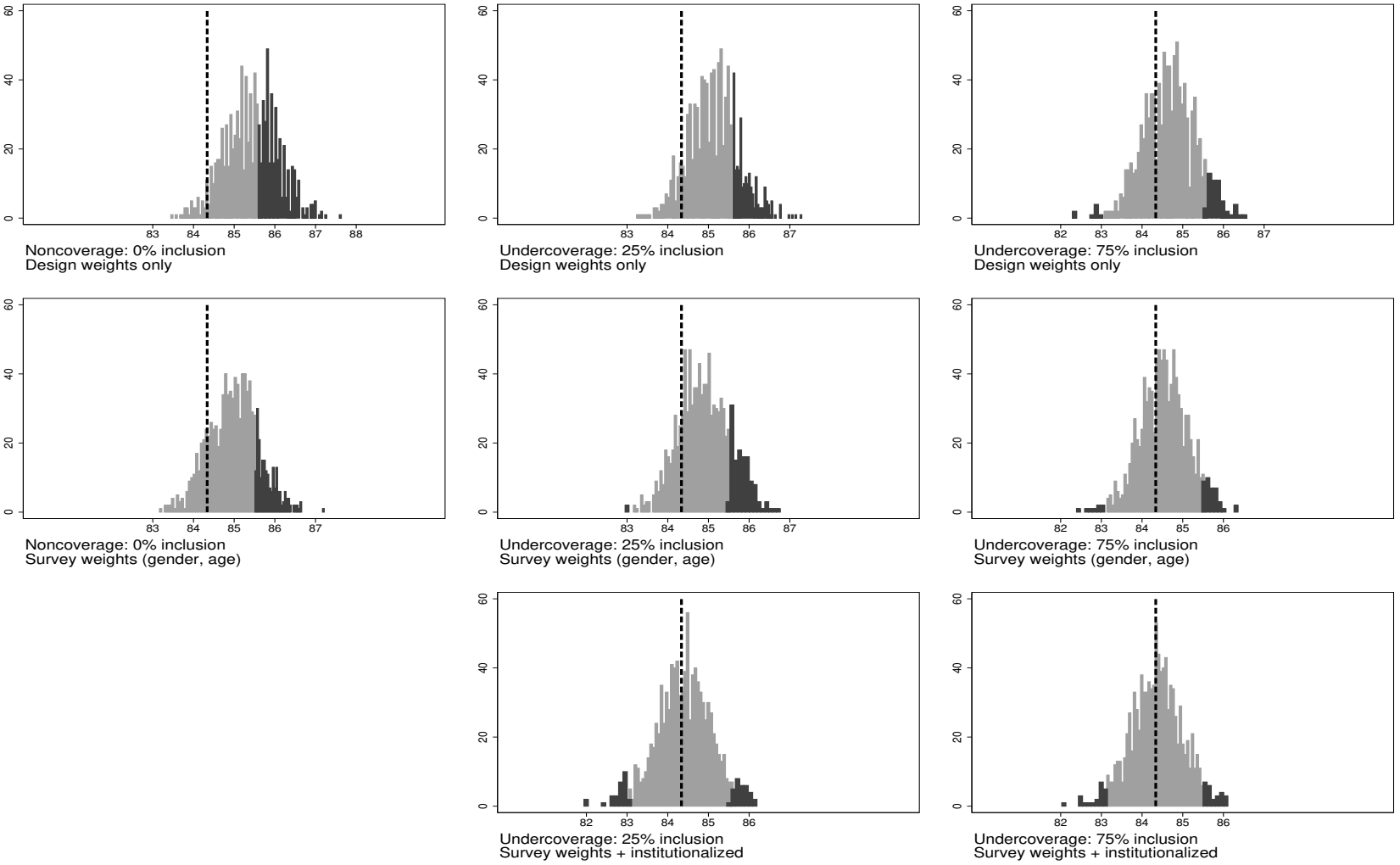
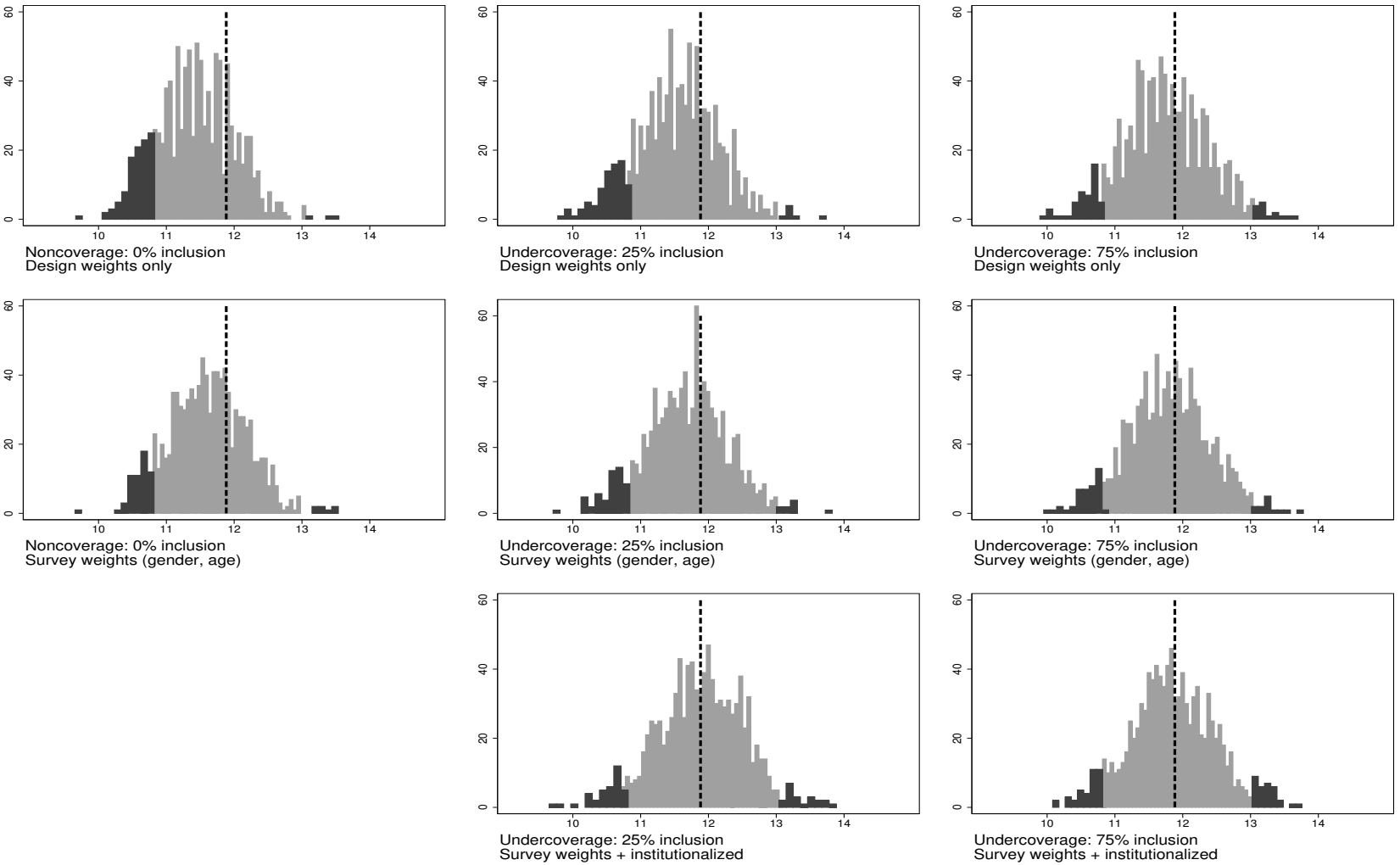


Figure 3.4: Estimated Share of Respondents with **Poor Self-Rated Health** in 1,000 Samples Per Histogram (*Light grey*: True share within 95% CI ; *Dark grey*: True share beyond 95% CI; *Dashed line*: True share in inferential population)



estimates for different undercoverage scenarios, and the estimators for the share of respondents without any limitations in ADL and the share of individuals with poor self-rated health, respectively (see Appendix C for the respective Figures for the remaining coverage conditions).

As we assumed, the bias in the ADL variable becomes significantly smaller with an increasing coverage of the institutionalized population (see Figure 3.1). Using $\overline{y}_{I_{0.25},\,d}$, an expected 18 institutionalized individuals in a sample of nearly 3,000 respondents made a difference and decreased the CI noncoverage rate to 43.6% of all samples. The precision of $\overline{y}_{I_{0.5},\,d}$ and $\overline{y}_{I_{0.75},\,d}$ got even better. Figure 3.3 shows how the distributions of estimates were shifted to the left, bringing their center closer to the true value. As with noncoverage, using auxiliary information for estimation decreased the likelihood of bias in all the undercoverage frames. $\overline{y}_{I_{0.25},\,\mathrm{re}_1}$ reduced the CI noncoverage rate of the share of respondents without limitations in ADL by about 50% compared to $\overline{y}_{I_{0.25},\,d}$ (see second row in Figure 3.3). The extended survey weights resulted in nearly unbiased samples from $\overline{y}_{I_{0.25},\,\mathrm{re}_2}$ onward. Only 6.0% of all samples missed the true value with their confidence interval, and 6.8% of all samples missed the true mean with their CI when survey weights accounted for the undercovered group. For the traditional survey weights ($\overline{y}_{I_\alpha,\,\mathrm{re}_2}$), the CI noncoverage rates approached the 5% threshold for $\alpha = 0.75$ (see Table 3.6).

Regarding self-rated health, the bias in the case of noncoverage was smaller than for the ADL variable. This observation also holds for the different scenarios of undercoverage. With an increasing α, the bias in $\overline{y}_{I_\alpha,\,\mathrm{d}}$ is reduced in both point estimates. The slight overestimation of the mean of self-rated health and the underestimation of the share of respondents with poor self-rated health were progressively corrected. This is in line with our expectations. In addition, most of the weighted estimators of self-rated health were estimated almost without any bias of an α of at least 0.5 (Table 3.7). Weighting with age, gender and institutionalization as auxiliary variables produced almost unbiased estimates in every scenario of undercoverage (see Figure 3.2 and last row of Figure 3.4).

The results for $\overline{y}_{I_\alpha,\,\mathrm{re}_1}$ promote the overall conclusion that conventional survey weights might indeed help to counterbalance the bias in health-related variables. In the case of undercoverage, the few institutionalized residents received larger weights because they were older and more often female than the average respondent. The weights significantly improved the precision of the estimates compared to simple design-weights. This improvement can only be achieved if institutionalized residents are included in the auxiliary variables used for weighting. For instance, the Health Survey for England excludes institutionalized residents from the population estimates used for calibration weighting (Craig et al. 2015, 25) and misses the chance to counterbalance the impact of noncoverage at least indirectly. In addition to traditional survey weights, weighting directly for the hard-to-survey group proves to be even more efficient in our simulation. The estimator $\overline{y}_{I_\alpha,\,\mathrm{re}_2}$ yielded unbiased results for all scenarios of undercoverage. However, one drawback of this method is the necessity to have at least some respondents from the hard-to-survey population in the sample.

3.8 Conclusion

The present study quantified the bias caused by the noncoverage or undercoverage of the population living in institutions for the elderly and tested whether survey weights can help to counterbalance the bias. Given the very small share of institutionalized respondents, we pooled the panel data of the *Survey of Health, Ageing and Retirement in Europe* (SHARE). We obtained a dataset of nearly 100,000 respondents, among them more than 2,400 institutionalized respondents. Taking this data as the basis for our simulation, we randomly sampled 5,000 samples of 3,000 observations with different coverage rates of institutionalized respondents. For each of the 5,000 samples, we calculated descriptive statistics and applied three different types of weights.

The simulation results prove that coverage bias affects variables to a different extent and even influences different kinds of point estimates in different ways. A variable measuring the limitations in activities of daily living (ADL) is heavily biased if respondents living in institutions are excluded from the samples. Our second variable measuring the self-rated health of respondents also was biased, but the impact of noncoverage is less pronounced. All other things being equal, the precision of all the point estimates improved with an increasing coverage rate. We tested three conditions of undercoverage and found a clear pattern of a decreasing coverage bias when more institutionalized residents were included.

Weighting the samples for age and gender improved all the point estimates in the case of noncoverage or undercoverage. If survey researchers make sure that the population margins they use for weighting *include* institutionalized residents, they can improve the precision of their estimates even without including the elderly institutionalized population in their survey. The likelihood of unbiased results gets even better if institutionalized residents are included in the samples, even if this domain suffers from undercoverage. Undercoverage also allows researchers to weight for the hard-to-survey population as such, if good population estimates of the share and relevant characteristics are available. In our simulation, the extended survey weights eliminated bias for both health-related variables. This shows the high potential of calibrating the weights on the size of undercovered populations to reduce bias.

Our simulation-based analysis has a number of limitations. By analyzing the respondents of SHARE with a simulation-based approach, we wanted to arrive at some general conclusions about the risk of biased estimates in social surveys. One advantage of our simulation is that it does not rely on synthetic observations, but considers real respondents instead. Especially in the case of an understudied group like institutionalized populations, a simulation with synthetic observations would require a large number of assumptions and thus raise many question about the validity of simulation results.

Nevertheless, our conclusions only hold true for the empirical population we compiled. The potential to generalize our results beyond the simulation depends on the data quality and the validity of our methods. As we mentioned previously, some of the SHARE countries reported undercoverage of institutionalized respondents in the first wave and refreshment waves (De Luca et al. 2015; Klevmarken et al. 2005; Lynn et al. 2013). Panel attrition due to non-contact and refusals is a sec-

ond major factor that could lead to a biased longitudinal sample of institutionalized respondents. Moving to an institution constitutes an external shock that probably increases the likelihood of non-contact and/or refusal (see Lugtig 2014).[14] Due to undercoverage and panel attrition, our statistical population certainly misses institutionalized residents in the general population, although the share of these respondents in our simulation (2.5%) is close to the share of institutionalized residents in the overall population older than 50 years (1.9% according to the European census).
Regarding our method, the data we used was collected during a period of 11 years between 2004 and 2015. We considered this data as cross-sectional data, and thus, might have missed the changes of contextual factors that could influence our results as confounding variables. In this study, we only analyzed a limited number of two variables. Both variables are health-related variables, and therefore, they have a higher likelihood of being biased with respect to the noncoverage or undercoverage of the elderly institutionalized population. We suggest taking our results of the ADL variable as the maximum effect coverage bias can have on survey estimates. We assume that the results for our second variable shows a more common influence of coverage bias. The presentation of research findings in Section 3.3 puts forward more variables that could be biased and should be tested, such as marital status, family composition, objective health, and socio-economic status.
A lot of research remains to be done on the institutionalized population. We only analyzed mean values and the distribution of variables. It is easier to interpret bias in descriptive statistics, but since they are only the starting point of most scientific papers, it would be very informative to extend the analysis to cover inferential statistics as well. For instance, the bias of regression coefficients could be tested by comparing the predicted outcomes of an equation with the respective results obtained in samples with equal coverage. Our study focused on the question of whether it is *necessary* to include institutionalized residents. Future research should examine whether it is also *feasible* to include this domain.
Our results could have implications for survey research in Europe because most social surveys in Europe deliberately exclude the institutionalized population, since it is considered as hard-to-survey (see Tourangeau 2014). Within our study, we asked whether survey programs need to make additional efforts to extend their coverage to the institutionalized population. Our results show for two health-related variables that surveys of an aging population indeed risk to obtain biased survey estimates if the institutionalized population is excluded. Moreover, the results suggest to abstain from the strategy to exclude institutionalized residents, because of the concern to cover this population insufficiently. Undercoverage is less sensitive to bias than noncoverage, especially when a sample is weighted for age and gender or even for institutionalization itself.

References

Agüero-Torres, Hedda, Eva Von Strauss, Matti Viitanen, Bengt Winblad, and Laura Fratiglioni (2001). “Institutionalization in the elderly: The role of chronic diseases and dementia. Cross-

[14] Please note, empirical evidence about the effect of institutionalization on panel membership is lacking.

sectional and longitudinal data from a population-based study". *Journal of Clinical Epidemiology* 54(8): 795–801.

Andersen, Ronald (1995). "Revisiting the Behavioral Model and Access to Medical Care: Does it Matter?" *Journal of Health and Social Behavior* 36(1): 1–10.

Angelini, Viola and Anne Laferrère (2012). "Residential mobility of the European elderly". *CESifo Economic Studies* 58(3): 544–569.

Asakawa, Keiko, David Feeny, Ambikaipakan Senthilselvan, Jeffrey Johnson, and Darryl Rolfson (2009). "Do the determinants of health differ between people living in the community and in institutions?" *Social Science and Medicine* 69(3): 345–353.

Börsch-Supan, Axel (2017a). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1*. Data Release: 6.0.0. SHARE-ERIC. DOI: 10.6103/SHARE.w1.600.

– (2017b). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2*. Data Release: 6.0.0. SHARE-ERIC. DOI: 10.6103/SHARE.w2.600.

– (2017c). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4*. Data Release: 6.0.0. SHARE-ERIC. DOI: 10.6103/SHARE.w4.600.

– (2017d). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5*. Data Release: 6.0.0. SHARE-ERIC. DOI: 10.6103/SHARE.w5.600.

– (2017e). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6*. Data Release: 6.0.0. SHARE-ERIC. DOI: 10.6103/SHARE.w6.600.

Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber (2013). "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)". *International Journal of Epidemiology* 42(4): 992–1001.

Börsch-Supan, Axel and Mathis Schröder (2011). "Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe". In: *SHARELIFE Methodology*. Ed. by Mathis Schröder. Mannheim Research Institute for the Economics of Ageing (MEA): 5–10.

Bravell, Marie, Stig Berg, Bo Malmberg, and Gerdt Sundström (2009). "Sooner or later? A study of institutionalization in late life". *Aging Clinical & Experimental Research* 21(4-5): 329–337.

Castora-Binkley, Melissa, Hongdao Meng, and Kathryn Hyer (2014). "Predictors of long-term nursing home placement under competing risk: Evidence from the health and retirement study". *Journal of the American Geriatrics Society* 62(5): 913–918.

Craig, Rachel, Jennifer Mindell, and Vasant Hirani (2015). *Health Survey for England: Methods and documentation*. Vol. 2. Leeds (UK): Health and Social Care Information Centre Health.

De Luca, Giuseppe, Claudio Rossetto, and Frederic Malter (2015). "Sample design and weighting strategies in SHARE Wave 5". In: *SHARE Wave 5: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy: 75–84.

Désesquelles, Aline and Nicolas Brouard (2003). "The Family Networks of People aged 60 and over Living at Home or in an Institution". *Population* 58(2): 201–227.

Einio, Elina, Christine Guilbault, Pekka Martikainen, and Michel Poulain (2012). "Gender Differences in Care Home Use Among Older Finns and Belgians". *Population* 67(1): 75–101.

ESS Sampling Expert Panel (2016). "Sampling Guidelines: Principles and Implementation for the European Social Survey". London: ESS ERIC Headquarters.

Eurostat (2015). *People in the EU: who are we and how do we live?* Tech. rep. Luxembourg.

– (2016). *European Statistical System - Census Hub*. URL: https://ec.europa.eu/CensusHub2.

Feskens, Remco (2009). "Difficult Groups in Survey Research and the Development of Tailor-made Approach Strategies". PhD thesis. Universiteit Utrecht.

Gaugler, Joseph, Sue Duval, Keith Anderson, and Robert Kane (2007). "Predicting nursing home admission in the U.S: a meta-analysis". *BMC Geriatrics* 7(13): 1–14.

Geerts, Joanna and Karel van den Bosch (2012). "Transitions in formal and informal care utilisation amongst older Europeans: The impact of national contexts". *European Journal of Ageing* 9(1): 27–37.

Groves, Robert, Floyd Fowler, Mick Couper, James Lepkowski, Eleanor Singer, and Roger Tourangeau (2009). *Survey Methodology*. Second edi. New York: John Wiley & Sons, Inc.

Grundy, Emily and Mark Jitlal (2007). "Socio-demographic variations in moves to institutional care 1991-2001: A record linkage study from England and Wales". *Age and Ageing* 36(4): 424–430.

Hancock, Ruth, Anthony Arthur, Carol Jagger, and Ruth Matthews (2002). "The effect of older people's economic resources on care home entry under the United Kingdom's long-term care financing system". *Journals of Gerontology, Series B* 57(5): 285–293.

Hays, Judith, Carl Pieper, and Jama Purser (2003). "Competing risk of household expansion or institutionalization in late life." *Journals of Gerontology, Series B* 58(1): S11–S20.

Himes, Christine, Gert Wagner, Douglas Wolf, Hakan Aykan, and Deborah Dougherty (2000). "Nursing home entry in Germany and the United States". *Journal of Cross-Cultural Gerontology* 15(2): 99–118.

Kasper, Judith, Liliana Pezzin, and J. Bradford Rice (2010). "Stability and changes in living arrangements: Relationship to nursing home admission and timing of placement". *Journals of Gerontology, Series B* 65 B(6): 783–791.

Kelfve, Susanne, Mats Thorslund, and Carin Lennartsson (2013). "Sampling and non-response bias on health-outcomes in surveys of the oldest old". *European Journal of Ageing* 10(3): 237–245.

Klevmarken, Anders, Bengt Swensson, and Patrik Hesselius (2005). "The SHARE Sampling Procedures and Calibrated Design Weights". In: *The Survey of Health, Ageing and Retirement in Europe – Methodology*. Ed. by Axel Börsch-Supan and Hendrik Jürges. September. Mannheim. Chap. 5: 28–37.

Laferrère, Anne, Aaron Van Den Heede, and Karel Van Den Bosch (2012). "Entry into institutional care: predictors and alternatives". In: *Active ageing and solidarity between generations in Europe*. Ed. by Axel Börsch-Supan, Martina Brandt, Howard Litwin, and Guglielmo Weber: 253–264.

Lessler, Judith T. and William D. Kalsbeek (1992). *Nonsampling Error in Surveys*. New York: John Wiley & Sons, Inc.

Lugtig, Peter (2014). "Panel attrition: separating stayers, fast attriters, gradual attriters, and lurkers". *Sociological Methods & Research* 43(4): 699–723.

Luppa, Melanie, Tobias Luck, Herbert Matschinger, Hans-Helmut König, and Steffi Riedel-Heller (2010a). "Predictors of nursing home admission of individuals without a dementia diagnosis before admission - results from the Leipzig Longitudinal Study of the Aged (LEILA 75+)." *BMC health services research* 10: 186.

Luppa, Melanie, Tobias Luck, Siegfried Weyerer, Hans Helmut König, Elmar Brähler, and Steffi Riedel-Heller (2010b). "Prediction of institutionalization in the elderly. A systematic review". *Age and Ageing* 39(1): 31–38.

Lynn, Peter (2011). "Maintaining Cross-Sectional Representativeness in a Longitudinal General Population Survey".

Lynn, Peter, Giuseppe De Luca, Matthias Ganninger, and Sabine Häder (2013). "Sample Design in SHARE Wave Four". In: *SHARE Wave 4 Innovations & Methodology*. Ed. by Frederic Malter and Axel Börsch-Supan. München: MEA, Max Planck Institute for Social Law and Social Policy: 74–123.

Martikainen, Pekka, Heta Moustgaard, Michael Murphy, Elina Einiö, Seppo Koskinen, Tuija Martelin, and Anja Noro (2009). "Gender, Living Arrangements, and Social Circumstances as Determinants of Entry into and Exit from Long-Term Institutional Care at Older Ages: A 6-Year Follow-Up Study of Older Finns". *Gerontologist* 49(1): 34–45.

Maxwell, Colleen, Andrea Soo, David Hogan, Walter Wodchis, Erin Gilbart, Joseph Amuah, Misha Eliasziw, Brad Hagen, and Laurel Strain (2013). "Predictors of Nursing Home Placement from Assisted Living Settings in Canada". *Canadian Journal on Aging* 32(04): 333–348.

McCann, Mark, Emily Grundy, and Dermot O'Reilly (2012). "Why is housing tenure associated with a lower risk of admission to a nursing or residential home? Wealth, health and the incentive to keep 'my home'". *Journal of Epidemiology & Community Health* 66(2): 166–169.

Neuert, Cornelia, Otto Wanda, Angelika Stiegler, Clara Beitz, Robin Schmidt, Katharina Meitinger, and Natalja Menold (2016). "SHARE Gatekeeper-Project: Cognitive Pretest". Mannheim.

Nihtilä, Elina, Pekka Martikainen, Seppo Koskinen, Antti Reunanen, Anja Noro, and Unto Häkkinen (2008). "Chronic conditions and the risk of long-term institutionalization among older people". *European Journal of Public Health* 18(1): 77–84.

Nöel-Miller, Claire (2010). "Spousal Loss, Children, and the Risk of Nursing Home Admission". *Journals of Gerontology, Series B* 65(3): 370–380.

Pimouguet, Clément, Debora Rizzuto, Pär Schön, Behnaz Shakersain, Sara Angleman, Marten Lagergren, Laura Fratiglioni, and Weili Xu (2016). "Impact of living alone on institutionalization and mortality: A population-based longitudinal study". *European Journal of Public Health* 26(1): 182–187.

Pot, Anne Margriet, France Portrait, Geraldine Visser, Martine Puts, Marjolein Broese van Groenou, and Dorly J H Deeg (2009). "Utilization of acute and long-term care in the last year of life: comparison with survivors in a population-based study". *BMC Health Services Research* 9(1): 139.

Riedel-Heller, Steffi G, Astrid Schork, Herbert Matschinger, and Matthias C Angermeyer (2000). "Recruitment procedures and their impact on the prevalence of dementia: Results from the Leipzig Longitudinal Study of the Aged (LEILA75+)". *Neuroepidemiology* 19(3): 130–140.

Rodríguez-Sánchez, Beatriz, Viola Angelini, Talitha Feenstra, and Rob Alessie (2017). "Diabetes-Associated Factors as Predictors of Nursing Home Admission and Costs in the Elderly Across Europe". *Journal of the American Medical Directors Association* 18(1): 74–82.

Sangl, Judith et al. (2007). "The Development of a CAHPS® Instrument for Nursing Home Residents (NHCAHPS)". *Journal of Aging & Social Policy* 19(2): 63–82.

Särndal, Carl-Erik (2007). "The calibration approach in survey theory and practice". *Survey Methodology* 33(2): 99–119.

Särndal, Carl-Erik, Bengt Swensson, and Jan Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Schanze, Jan-Lucas (2017). *Report on sampling practices for the institutionalized population in social*. Tech. rep. Deliverable 2.16 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. URL: https://seriss.eu/resources/deliverables/.

Schnell, Rainer (1991). "Wer ist das Volk? Zur faktischen Grundgesamtheit bei „allgemeinen Bevöllkerungsfragen": Undercoverage, Schwererreichbare und Nichtbefragbare". *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 1(43): 106–137.

Suanet, Bianca, Marjolein Broese van Groenou, and Theo van Tilburg (2012). "Informal and formal home-care use among older adults in Europe: can cross-national differences be explained by societal context and composition?" *Ageing and Society* 32(03): 491–515.

Thomeer, Mieke Beth, Stipica Mudrazija, and Jacqueline Angel (2015). "How do race and hispanic ethnicity affect nursing home admission? Evidence from the health and retirement study". *Journals of Gerontology, Series B* 70(4): 628–638.

Toot, Sandeep, Tom Swinson, Mike Devine, David Challis, and Martin Orrell (2017). "Causes of nursing home placement for older people with dementia: a systematic review and meta-analysis". *International Psychogeriatrics* 29(02): 195–208.

Tourangeau, Roger (2014). "Defining hard-to-survey populations". In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy Johnson, Kirk Wolter, and Nancy Bates. Cambridge University Press: 3–20.

Von Bonsdorff, Mikaela, Taina Rantanen, Pia Laukkanen, Timo Suutama, and Eino Heikkinen (2006). "Mobility limitations and cognitive deficits as predictors of institutionalization among community-dwelling older people". *Gerontology* 52(6): 359–365.

Appendix A

Table 3.2: List of socio-demographic variables that have a positive statistical impact on the likelihood of living in institutions for the elderly

Variables	**Studies finding a significant positive impact in multivariate models**
Age	Agüero-Torres et al. 2001; Angelini and Laferrère 2012; Asakawa et al. 2009 Castora-Binkley et al. 2014; Désesquelles and Brouard 2003 Einio et al. 2012; Gaugler et al. 2007; Grundy and Jitlal 2007 Hays et al. 2003; Kasper et al. 2010; Laferrère et al. 2012 Luppa et al. 2010a; Martikainen et al. 2009; Maxwell et al. 2013 McCann et al. 2012; Pot et al. 2009; Riedel-Heller et al. 2000 Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015
Civil status (not married, widowed, divorced)	Angelini and Laferrère 2012; Asakawa et al. 2009; Bravell et al. 2009 Castora-Binkley et al. 2014; Einio et al. 2012; Gaugler et al. 2007 Grundy and Jitlal 2007; Luppa et al. 2010a; Nöel-Miller 2010 Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015
Education	*Lower*: Asakawa et al. 2009; Einio et al. 2012 *Higher*: Castora-Binkley et al. 2014; Thomeer et al. 2015
Ethnicity (white)	Asakawa et al. 2009; Castora-Binkley et al. 2014; Hays et al. 2003 Kasper et al. 2010; Luppa et al. 2010a; Nöel-Miller 2010 Thomeer et al. 2015
Gender	*Female*: Bravell et al. 2009; Einio et al. 2012; Grundy and Jitlal 2007 Kasper et al. 2010; McCann et al. 2012 *Male*: Désesquelles and Brouard 2003; Einio et al. 2012 Gaugler et al. 2007; Luppa et al. 2010a; Martikainen et al. 2009 Pot et al. 2009
Housing (not home owner)	Einio et al. 2012; Gaugler et al. 2007; Grundy and Jitlal 2007 Luppa et al. 2010a; Martikainen et al. 2009; McCann et al. 2012 Thomeer et al. 2015
Income	*Lower*: Angelini and Laferrère 2012; Gaugler et al. 2007 Laferrère et al. 2012; Martikainen et al. 2009; Thomeer et al. 2015 *Higher*: Rodríguez-Sánchez et al. 2017
Living alone (without partner or children)	Angelini and Laferrère 2012; Désesquelles and Brouard 2003 Gaugler et al. 2007; Grundy and Jitlal 2007; Kasper et al. 2010 Laferrère et al. 2012; Martikainen et al. 2009; McCann et al. 2012 Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015
Parenthood (Not having (grand-) children (living nearby))	Bravell et al. 2009; Hays et al. 2003; Kasper et al. 2010 Laferrère et al. 2012; Nöel-Miller 2010; Rodríguez-Sánchez et al. 2017
Small social networks, no informal caregiver	Hays et al. 2003; Laferrère et al. 2012; Luppa et al. 2010a Maxwell et al. 2013; Nöel-Miller 2010; Thomeer et al. 2015

Table 3.3: List of health-related variables that have a positive statistical impact on the likelihood of living in institutions for the elderly

Variables	**Studies finding a significant positive impact in multivariate models**
Cognitive impairment	Agüero-Torres et al. 2001; Castora-Binkley et al. 2014 Einio et al. 2012; Gaugler et al. 2007; Hays et al. 2003 Kasper et al. 2010; Luppa et al. 2010a; Maxwell et al. 2013 Nihtilä et al. 2008; Nöel-Miller 2010; Rodríguez-Sánchez et al. 2017 Thomeer et al. 2015; Toot et al. 2017
Dementia	Agüero-Torres et al. 2001; Einio et al. 2012; Laferrère et al. 2012 Luppa et al. 2010a; Nihtilä et al. 2008; Riedel-Heller et al. 2000 Toot et al. 2017
Functional impairments	Luppa et al. 2010a; Maxwell et al. 2013; Pot et al. 2009
Medical conditions	Agüero-Torres et al. 2001; Angelini and Laferrère 2012 Einio et al. 2012; Gaugler et al. 2007; Grundy and Jitlal 2007 Hays et al. 2003; Luppa et al. 2010a; Maxwell et al. 2013 Nihtilä et al. 2008; Rodríguez-Sánchez et al. 2017; Toot et al. 2017
Limitations in activities of daily living (ADL)	Bravell et al. 2009; Gaugler et al. 2007; Hays et al. 2003 Kasper et al. 2010; Laferrère et al. 2012; Nöel-Miller 2010 Rodríguez-Sánchez et al. 2017; Thomeer et al. 2015; Toot et al. 2017
Limitations in instrumental activities of daily living (IADL)	Castora-Binkley et al. 2014; Laferrère et al. 2012; Nöel-Miller 2010 Thomeer et al. 2015; Toot et al. 2017
Physical dependency and mobility difficulties	Angelini and Laferrère 2012; Désesquelles and Brouard 2003 Hays et al. 2003; Thomeer et al. 2015; Toot et al. 2017
Self-rated health	Castora-Binkley et al. 2014; Einio et al. 2012; Luppa et al. 2010a McCann et al. 2012; Nöel-Miller 2010

Appendix B

Table 3.4: Origin of respondents used to compile our empirical population. *In brackets*: Institutionalized respondents among those (For the sampling designs of most countries in the first wave, see Klevmarken et al. 2005).

	Wave 1	Wave 2	Wave 4	Wave 5	Wave 6	Total
Austria	315	366	902	1,226	3,310	6,119
	(6)	(13)	(43)	(31)	(58)	(151)
Germany	1,195	1,086	496	1,363	4,285	8,425
	(13)	(27)	(8)	(45)	(54)	(147)
Sweden	607	706	334	985	3,788	6,420
	(31)	(50)	(30)	(36)	(67)	(214)
Netherlands	837	881	517	3,977	-	6,212
	(82)	(75)	(58)	(198)	(-)	(413)
Spain	515	498	483	1,428	5,416	8,340
	(13)	(9)	(14)	(29)	(61)	(126)
Italy	598	653	760	1,050	5,126	8,187
	(16)	(5)	(17)	(22)	(36)	(96)
France	638	766	1,365	1,271	3,796	7,836
	(15)	(21)	(31)	(29)	(56)	(152)
Denmark	300	659	270	756	3,608	5,593
	(26)	(62)	(24)	(43)	(79)	(234)
Greece	292	1,251	-	-	4,777	6,320
	(5)	(55)	(-)	(-)	(2)	(62)
Switzerland	226	366	654	441	2,717	4,404
	(5)	(14)	(24)	(24)	(57)	(124)
Belgium	737	816	996	1,239	5,628	9,416
	(11)	(24)	(69)	(61)	(155)	(320)
Israel	501	430	-	709	1,862	3,502
	(10)	(35)	(-)	(21)	(35)	(101)
Czech Republic	-	1,283	1,079	1,180	4,732	8,274
	(-)	(28)	(21)	(44)	(76)	(169)
Luxembourg	-	-	-	500	1,539	2,039
	(-)	(-)	(-)	(22)	(30)	(52)
Estonia	-	-	917	1,135	5,527	7,579
	(-)	(-)	(15)	(13)	(52)	(80)
Total	6,761	9,761	8,773	17,260	56,111	98,666
	(233)	(418)	(354)	(618)	(818)	(2,441)

Table 3.5: Operationalization of all variables used in the empirical analysis

Variable	**Operationalization in our analysis**
Index on limitations in ADL	Count variable, reaching from 0 (no limitations) to 9 limitations with the following items: ADL: Dressing, including putting on shoes and socks; Walking across a room; Bathing or showering; Eating, such as cutting up your food; Getting in or out of bed; Using the toilet, including getting up or down; IADL: Making telephone calls; Taking medications; Managing money, such as paying bills and keeping track of expenses
Self-rated health	1 = Poor, 2 = Fair, 3 = Good, 4 = Very good, and 5 = Excellent
Age	1 = 50 to 59 years, 2 = 60 to 69 years, 3 = 70 to 79 years, 4 = 80 to 89 years, and 5 = older than 90 years
Gender	0 = male, 1 = female
Institutionalization	0 = Community-dwelling respondents, 1 = Institutionalized respondents

Appendix C

Table 3.6: Aggregated Estimations for Two Point Estimates Across 1,000 Simulations Per Cell (In brackets, percentage of samples that missed the true value with their 95 % confidence interval)

	Mean of limitations in ADL			**Share of respondents with no limitations in ADL**		
	$\overline{y}_{I_\alpha, d}$	$\overline{y}_{I_\alpha, \text{re}_1}$	$\overline{y}_{I_\alpha, \text{re}_2}$	$\overline{y}_{I_\alpha, d}$	$\overline{y}_{I_\alpha, \text{re}_1}$	$\overline{y}_{I_\alpha, \text{re}_2}$
$\mathcal{U}_{p\,I_0}$	0.4018	0.4170	-	85.43	85.02	-
	(63.4)	(42.3)	-	(41.4)	(21.6)	-
$\mathcal{U}_{p\,I_{0.25}}$	0.4145	0.4266	0.4563	85.15	84.84	84.39
	(43.6)	(27.2)	(6.8)	(24.5)	(12.8)	(6.0)
$\mathcal{U}_{p\,I_{0.50}}$	0.4302	0.4389	0.4613	84.88	84.66	84.34
	(22.1)	(13.7)	(5.1)	(16.0)	(9.5)	(5.1)
$\mathcal{U}_{p\,I_{0.75}}$	0.4447	0.4492	0.4589	84.64	84.53	84.36
	(7.8)	(6.7)	(5.2)	(6.0)	(4.7)	(4.6)
$\mathcal{U}_{p\,I_1}$	0.4584	0.4589	0.4576	84.34	84.33	84.32
	(5.6)	(5.3)	(5.6)	(5.1)	(5.5)	(4.5)
True values	0.4580			84.34		

Table 3.7: Aggregated Estimations for Two Point Estimates Across 1,000 Simulations Per Cell (In brackets, percentage of samples that missed the true value with their 95 % confidence interval)

	Mean of self-rated health			**Share of respondents with poor self-rated health**		
	$\overline{y}_{I_\alpha, d}$	$\overline{y}_{I_\alpha, \text{re}_1}$	$\overline{y}_{I_\alpha, \text{re}_2}$	$\overline{y}_{I_\alpha, d}$	$\overline{y}_{I_\alpha, \text{re}_1}$	$\overline{y}_{I_\alpha, \text{re}_2}$
$\mathcal{U}_{p\,I_0}$	2.8223	2.8158	-	11.47	11.63	-
	(11.5)	(6.3)	-	(10.9)	(6.4)	-
$\mathcal{U}_{p\,I_{0.25}}$	2.8179	2.8128	2.8065	11.59	11.71	11.88
	(9.3)	(6.1)	(4.3)	(9.0)	(6.3)	(5.8)
$\mathcal{U}_{p\,I_{0.50}}$	2.8145	2.8110	2.8061	11.66	11.75	11.89
	(6.9)	(5.4)	(6.6)	(7.0)	(6.2)	(5.6)
$\mathcal{U}_{p\,I_{0.75}}$	2.8119	2.8100	2.8079	11.76	11.80	11.85
	(7.0)	(7.1)	(5.4)	(6.3)	(5.6)	(7.0)
$\mathcal{U}_{p\,I_1}$	2.8070	2.8068	2.8068	11.89	11.90	11.89
	(5.6)	(5.5)	(4.5)	(5.2)	(5.2)	(6.3)
True values	2.8069			11.88		

Figure 3.5: Estimated Share of Respondents **Without Any Limitations in ADL** in 1,000 Samples Per Histogram (*Light grey*: True share within 95% CI ; *Dark grey*: True share beyond 95% CI; *Dashed line*: True share in inferential population)

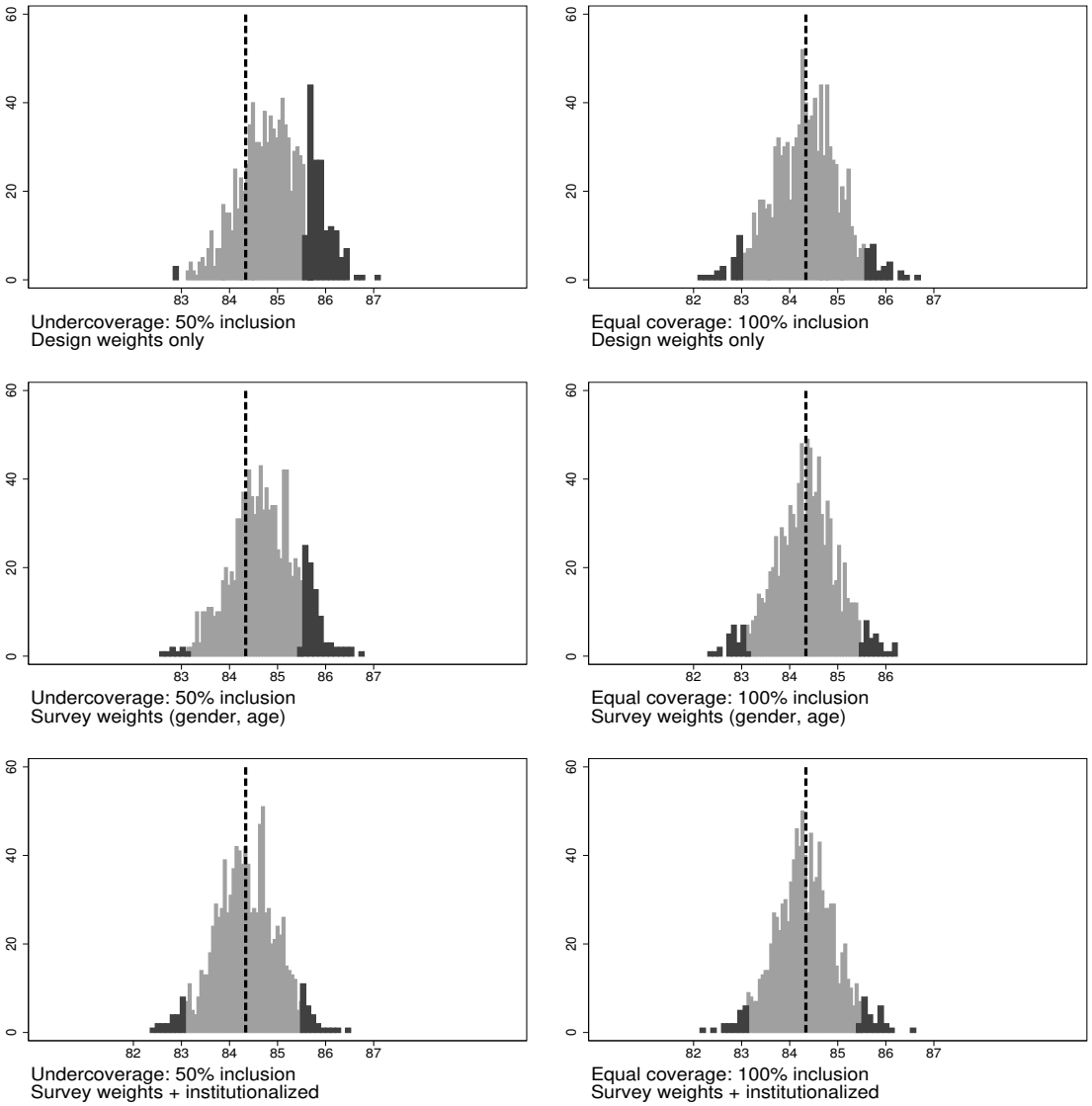
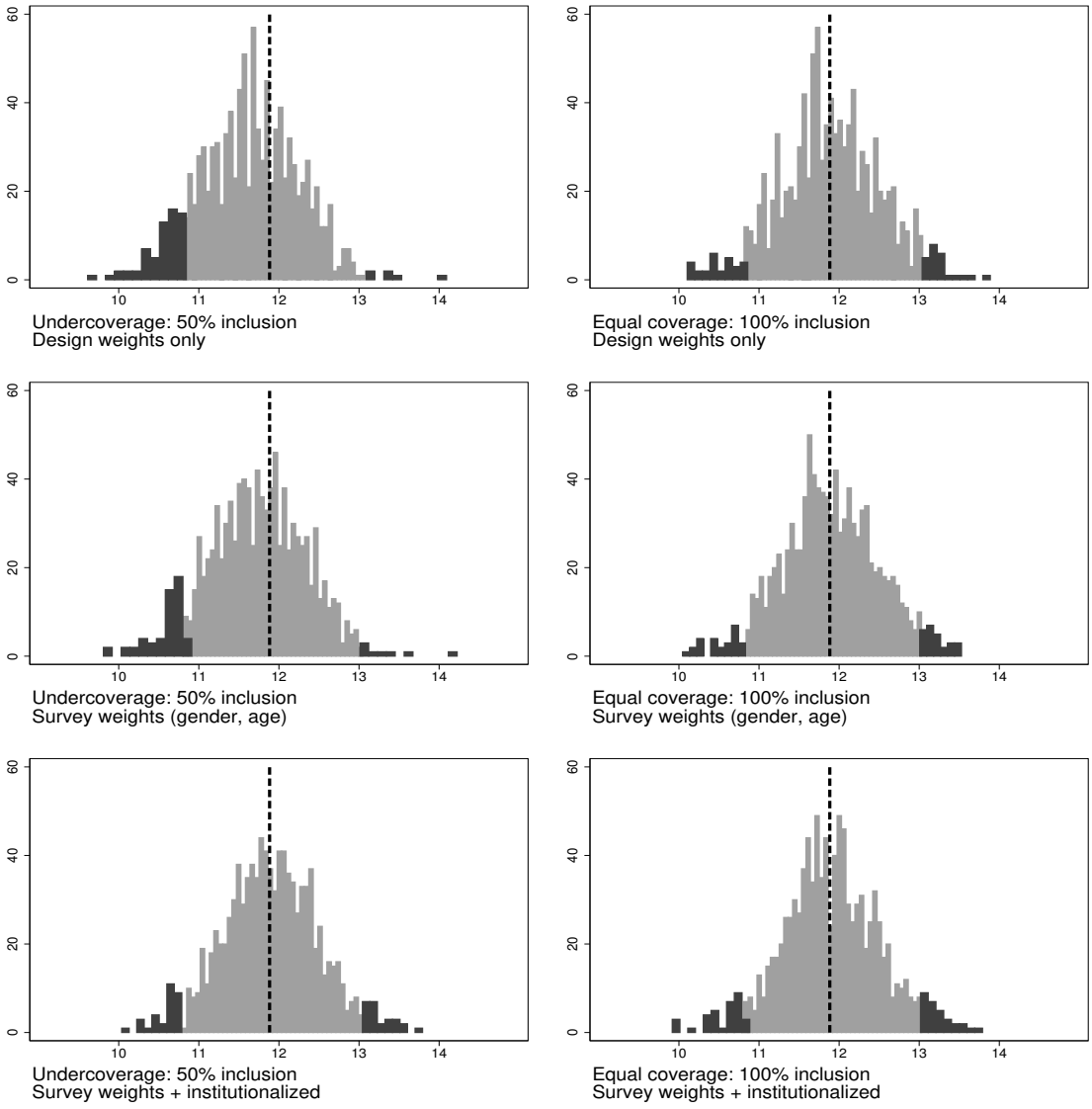


Figure 3.6: Estimated Share of Respondents with **Poor Self-Rated Health** in 1,000 Samples Per Histogram (*Light grey*: True share within 95% CI ; *Dark grey*: True share beyond 95% CI; *Dashed line*: True share in inferential population)



Chapter 4

Response behavior and quality of survey data: Comparing elderly respondents in institutions and private households

The following chapter is a pre-print of a paper accepted for publication by Sociological Methods & Research in December 2020.

4.1 Introduction

Residents living in retirement and nursing homes are often labeled as hard-to-survey and are excluded from many social surveys (Schanze and Levinson 2019). They are a very small population and hard-to-sample, they are hard-to-reach because the access to the institutions can be restricted by gatekeepers (ibid., (Tourangeau 2014)). Another important reason for the label and the exclusion are characteristics that have been confirmed in many previous research studies: Compared to private household residents, residents of retirement and nursing homes are older on average, suffer from worse health and live with functional and cognitive impairment more often (Agüero-Torres et al. 2001; Einio et al. 2012; Gaugler et al. 2007; Laferrère et al. 2012; Lang 2014; Luppa et al. 2008; Maxwell et al. 2013; Toot et al. 2017). Those characteristics obviously make them more difficult to interview (Tourangeau 2014). In an expert survey carried out among survey researchers with experiences in nursing and retirement homes, 14 out of 21 researchers (66.7%) agreed that residents of long-term care institutions are at least somewhat or even very hard to in-

terview (Schanze and Levinson 2019).[1]

Nevertheless, the EU-financed SERISS project identified nearly 100 survey programs that already collected data in retirement homes, nursing homes, health-related institutions or assisted living facilities (Schanze and Levinson 2019). Those survey programs sometimes have a broad research interest or a more specific focus on health or ageing, and they are conducted at a national level like the French *Enquêtes Capacités, Aides et REssources des seniors* (CARE), the U.S. *Health and Retirement Survey* (HRS), and the *UK Household Longitudinal Study* (UKHLS) or with a cross-national design like the *European Health Interview Survey* (EHIS) or the *Survey of Health, Ageing and Retirement in Europe* (SHARE). If institutionalized elderly are potentially harder to interview, can this be detected in the survey data?

When respondents answer survey questions, they need to cope with a complex task. This task and the various intellectual steps respondents need to perform were described multiple times in great detail (Alwin 2007; Alwin and Krosnick 1991; Knäuper et al. 2016; Schwarz 1999; Tourangeau et al. 2000): As the first step, respondents need to understand the meaning of the question and identify which information are required to answer it. Following the comprehension, respondents need to access relevant information from their memory and retrieve it. The retrieved information needs to be formatted to match the given answer scale of the survey question. Before voicing their reply to the interviewer, respondents need to judge whether the information are correct and whether it might be socially undesirable to answer the question honestly. As the last step, respondents need to communicate their reply.

The intensity and accuracy of running through this cognitive process is moderated by several factors (Alwin and Krosnick 1991; Krosnick 1991), namely (1) the difficulty of the survey item as such, whether is uses complicated, abstract, or ambiguous words (Alwin and Krosnick 1991; Colsher and Wallace 1989; Fuchs 2009; Lenzner 2012; Schwarz 2007), (2) the ability and characteristics of respondents, (3) their motivation to answer the item correctly (Fox et al. 2007; Knäuper et al. 2016), and (4) the impact of interviewer-respondent interactions (Beullens et al. 2018; Loosveldt 1997; Ongena and Dijkstra 2007), at least in survey modes that rely on interviewers.

The difficulty of specific survey questions and interviewer effects are not at the heart of this paper. I will examine the impact of respondents' characteristics on difficulties to interview them and on the quality of survey data by comparing survey responses of elderly respondents living in private households and institutions in SHARE.[2] My research question is to what extent the response behavior and the quality of data of respondents in institutions for the elderly differs from that of respondents living in private households? Do the characteristics of institutionalized residents make them hard-to-interview and does it also lead to a worse data quality? And are residents of private households comparably hard-to-interview when taking into account their socio-demographic characteristics and health condition?

In the following section, I will continue by summarizing previous research on the question of

[1] Schanze and Levinson (2019) only report aggregate results for the entire institutionalized population. The above mentioned numbers are based on own calculations.

[2] Both groups are confronted with the same survey items and their interviewers received the same training. Hence, task difficulty and the impact of interviewers should be rather similar across those two groups.

why respondents sometimes fail to answer survey questions appropriately, how age and cognitive impairment affect those processes, and whether institutionalization as such influences response behavior and data quality. The subsequent section describes the survey data I used and my research design, which mainly consists of matching institutionalized respondents to comparable respondents living in private households. I present results on a number of indicators for hard-to-interview respondents and data quality and will conclude with an answer to the overall question, whether institutionalized residents are really unique in terms of response behavior and whether it is feasible to conduct survey interviews with this "hard-to-interview" group in general social surveys.

4.2 Why do respondents fail to answer survey questions properly?

Throughout decades of research, scientist examined why respondents sometimes fail to run through the process of answering survey questions in a proper way. Instead of answering questions in an optimal way, respondents sometimes defect to a more careless and less motivated answering behavior coined as weak or strong *satisficing* by Krosnick (1991). As indicators of satisficing, Krosnick reviewed earlier studies on primacy and recency effects, acquiescence, which is the tendency to agree to questions irrespective of their content, endorsing the status quo, non-differentiation, Don't know replies, and mental coin flipping resulting in random responses. As mentioned above, question difficulty, respondents' characteristics and ability, their motivation, and interviewers can all lead to such a response behavior.

Looking at respondents' characteristics, many researchers assume that older respondents differ from younger respondents in terms of question comprehension, accessibility and retrieval of information, ability to translate behaviors or attitudes into the given scales of questions, communication abilities, and even motivation (Beullens et al. 2018; Knäuper et al. 2016). Age is one of the most common usual suspects for a lower data quality (Alwin 2007; Alwin and Krosnick 1991; Andrews and Herzog 1986; Beullens et al. 2018; Colsher and Wallace 1989; Fuchs 2009; Knäuper 1999; Knäuper et al. 2016; Kühn and Porst 1999; Lang 2014; Loosveldt 1997; Motel-Klingebiel et al. 2014; Schwarz 1999).

Several studies investigated the impact of age on a range of quality indicators. The degree of item nonresponse, especially "Don't know" responses, increases significantly with age (Colsher and Wallace 1989; Fuchs 2009; Knäuper 1999; Knäuper et al. 1997; Koch and Blohm 2009; Pickery and Loosveldt 1998), while early papers found a negative impact of age on reliability and valid variance (Alwin and Krosnick 1991; Andrews and Herzog 1986; Colsher and Wallace 1989; Rodgers et al. 1988). This finding was confirmed in a recent analysis by Kroh and colleagues (2016). In a meta-analysis of survey experiments, Knäuper (1999) found stronger response order effects (recency and primacy effects) for older respondents, which results in the increasing preference for extreme responses with advancing age (He et al. 2014; Kieruj and Moors 2013). Another experiment found no differences in the response behavior by older respondents depend-

ing on whether labels of a question were numbered or not, which was not the case for respondents younger than 70 years (Schwarz and Knäuper 1999). Both studies relate those findings to an aging-related decrease in working memory capacity and cognitive resources. Beullens et al. (2018) identified age as a driving factor of an increased need for clarification, interview length, and stronger interviewer effects. This may indicate a shift from from a standardized, but artificial interview process to a more conversational approach resembling everyday conversations (cf. ibid., Kühn and Porst 1999), which in fact can also help to improve respondents' comprehension of survey items and thereby improve the accuracy of responses (Schober and Conrad 1997; West et al. 2018).

As a positive response behavior, a negative impact of age on acquiescence was found in three studies using the Big-Five inventory and other items (Costello and Roodenburg 2015; He et al. 2014; Lechner and Rammstedt 2015), whereas another study could not find any significant impact of age on acquiescence in two different surveys when controlling for various other variables (Olson and Bilgen 2011). To explain acquiescence, respondents' psychological traits, the desire to be perceived in a positive light, and a lack of cognitive sophistication are cited as relevant factors next to interviewer effects (Olson and Bilgen 2011). However, acquiescence is sometimes difficult to find (cf. Kieruj and Moors 2013) and difficult to explain. As Roberts et al. (2019, 615) conclude, "*there is likely considerable variation in individual and cultural propensity to acquiescence for reasons other than lack of motivation or ability [...]*".

Despite the strong theoretical expectations and some confirmed hypotheses, the literature on the impact of age on the quality of survey responses shows mixed or weak effects. This can be explained by potential confounding of age effects with other important explanatory variables, most importantly education and cognitive impairment (Knäuper 1999). Regarding education, Alwin (2007) found that the difference between age groups are not significant when controlling for education. This can be related to cohort effects, since older respondents enjoyed a shorter education while being in their adolescence compared to the younger age cohorts. Secondly, even though age or education are not sufficient indicators for cognitive impairment (Kühn and Porst 1999; Loosveldt 1997), an advanced age is confounded with an increased cognitive decline (Salthouse 2009). When cognitive functioning declines, a problematic respondent behavior becomes more likely. For instance, the extent of item nonresponse increases (Fuchs 2009). On the other hand, a worse working memory also reduces the likelihood of question order effects, since respondents do not remember previous questions accurately (Fuchs 2009; Schwarz 1999).

Moreover, as previous research shows, data quality does not decrease linearly with age. Early research with U.S. panel data already showed an insignificant impact of age on test-retest reliability in 96 attitudinal measures with one exception: "*[...] declines in reliability accompanying aging [...] primarily occur in the oldest age group*" (Alwin and Krosnick 1991, 173). In a more recent paper, Beullens and colleagues (2018, 9) examined several indicators of respondents' behavior and arrived at a similar conclusion regarding the nonlinear impact of age, concluding that "[...*] up to the 54–62 age- group, little difference can be found among the respondents. It is only in the two oldest age-groups that the differences become very clear, with the oldest age-group (71 and above)*

being the most distinct." Those findings can probably be explained with an accelerated cognitive decline at older ages (Salthouse 2009), even though the reliability of survey items can also be negatively influenced by more rapid changes of the living situation and respondents' characteristics in old age, for instance, when moving to an institution due to the health condition.

When it comes to old age and cognitive impairment, institutionalized residents living in retirement and nursing homes are especially challenging *and* interesting. As many studies showed, age, health, and functional or cognitive impairment are significantly related to the likelihood of living in or moving to a nursing or retirement home (Agüero-Torres et al. 2001; Einio et al. 2012; Gaugler et al. 2007; Laferrère et al. 2012; Luppa et al. 2008; Maxwell et al. 2013; Toot et al. 2017).

Few scientific studies explicitly analyzed whether residents living in long-term care institutions have the capacity to be interviewed and how the resulting data quality looks like. In Austria, a small-scale survey in nursing homes examined residents' ability to participate in a survey on their quality of life (Lang 2014). Again, age as such was *not* a significant predictor of a lower response rate: Residents who could not take part in the survey needed a higher level of care, were more disoriented, and in worse general health (ibid.). This study emphasized the importance of cognitive impairment and health.

Regarding potential difficulties of survey questions for institutionalized residents, Sangl et al. (2007) concluded that nursing home residents mostly ignored reference periods given by questions but rather referred to single events, thereby simplifying the response task (ibid.). Moreover, residents preferred numbered scales from 0 to 10 instead of fully labeled scales, because they "*were not distracted by the meaning or emotional content of the words*" (Sangl et al. 2007, 71). Fox and colleagues (2007) tested various survey questions in a similarly challenging target group consisting of older adults in hospitals. The authors came to the conclusion that this population especially struggled with long questions with multiple parts. The comprehension of respondents was also worse for negatively framed items, because for those questions, disagreement indicates a positive statement. In addition, they found that questions with a low relevance to respondents caused frustration and tiredness among residents.

Most large-scale survey programs cannot take into account specific requirements of subgroups of respondents by adapting questions or survey procedures. Thus, considering the specific characteristics of residents in retirement and nursing homes, I expect more difficult interviews in institutions and a worse data quality compared to private households. This expectation is in line with the hard-to-interview label of institutionalized elderly used by many survey researchers. Moreover, it is my second hypothesis that those differences are mainly driven by the age and related factors such as health and cognitive condition of institutionalized residents. Hence, I expect much smaller differences between institutions and private households when controlling for socio-demographic and health-related variables. Given the absence of previous literature on those research questions, it remains to be seen whether differences will completely disappear.

4.3 Data and Analysis

Earlier studies of data quality and satisficing in the field of survey methodology have analyzed a wide set of indicators, as Roberts and colleagues (2019) summarized in a recently published systematic review. Those indicators comprise primacy and recency effects, acquiescence, non-differentiation, extreme responses, replies to trap question, or correlations of several variables (ibid.). This paper examines the share of proxy interviews and ratings of respondents' performance by interviewers as indicators of hard-to-interview respondents, the selection of middle and positive extreme categories to survey items as response patterns, as well as item nonresponse and acquiescence as indicators of data quality. I will also analyze how the indicators of hard-to-interview respondents influence the data quality indicators.

4.3.1 Data preparation and variables used

I use waves 5, 6 and 7 of the *Survey of Health, Ageing and Retirement in Europe* (SHARE), the three last waves of this survey collected in a number of European countries and Israel in 2013, 2015, and 2017 (Börsch-Supan 2019a,b,c). SHARE is a cross-national panel survey, which covers institutionalized respondents living in retirement and nursing homes and respondents living in private households older than 50 years old (Börsch-Supan et al. 2013).[3] In a similar approach to that of Schanze and Zins (2019), I pooled the SHARE waves 5, 6, and 7 and kept only the most recent observation for each respondent. I used this method to increase the number of institutionalized respondents in my analysis. Omission of information about waves and change across waves is justified because I am mainly interested in the response behavior as such and not in substantive analyses of institutionalization or changes over time within respondents. Following the pooling of waves, I dropped countries with 15 or less institutionalized respondents in the three waves, resulting in a dataset with more than 77,500 unique observations in 18 different countries. 2,002 of those observations were identified as institutionalized residents using three different variables in SHARE (2.5% of the pooled waves).[4]

As dependent variable I will start with the share of proxy interviews as a first indicator of a hard-to-interview respondent. Proxy interviews are a tool to safeguard the data quality when being confronted with difficult respondents. According to the SHARE Release Guide, "*only if physical and/or cognitive limitations make it too difficult for a respondent to complete the interview her-/himself, it is possible that the respondent is assisted by a so-called proxy respondent to complete*

[3] I dropped all cases younger than 50 years old when the interview took place. Those respondents are interviewed as partners of anchor respondents.

[4] In the technical variables module, a variable captures whether an interview was conducted in a nursing home (mn024). In addition, I use interviewer observations on the type of building and coded respondents as institutionalized, if they lived either in "*a housing complex with services for elderly*" or in a "*special housing for elderly (24 hours attention)*". Finally, I label those respondents as institutionalized, who indicated that they have been permanently living in a nursing home during the last 12 months.

the interview" (MEA 2019, 13). Even in cases of hearing impairment or Alzheimer's disease some parts of the SHARE interview are non-proxy modules and need to be completed by the respondent, if possible (ibid.).

As the second and third indicator of hard-to-interview respondents I analyze how interviewers rated respondents in terms of requiring additional clarifications[5] and their willingness to answer questions.[6] Asking for clarifications while interpreting a question is not a direct indicator of bad data quality, because respondents asking for additional explanations might intend to answer survey questions more accurately and with a higher quality compared to respondents who refrain from asking questions (Ongena and Dijkstra 2007). In any case this kind of response behavior will make the survey interview longer and more difficult for interviewers. Moreover, previous research indicated a link between an increased need for clarification with the tendency to provide inadequate answers (Loosveldt 1997) and with a more limited understanding of questions and larger interviewer effects (Beullens et al. 2018), especially among older respondents. The rating of respondents' willingness serves as a proxy for their motivation, which was defined as an important explanatory factor for satisficing alongside task difficulty and respondents' ability (Krosnick 1991). When analyzing ratings by interviewers, it should be kept in mind that interviewers sometimes use heuristic cues while replying to evaluation questions (Kirchner et al. 2017). In those cases, a higher age, and especially institutionalization, could lead to a more negative rating of respondents, even though their behavior might be comparable to the behavior of younger respondents or respondents living in private households.

Following the first three measures of hard-to-interview respondents, I will move on to indicators of data quality. First, I will investigate two different batteries of items in SHARE, namely the CASP inventory and a short version of the Big-Five inventory. The CASP inventory was developed before the first wave of SHARE was fielded in 2004 to measure the health-related quality of life (Knesebeck et al. 2005). In SHARE, the CASP inventory consists of 12 items measuring 4 different dimensions, namely control, autonomy, self-realization, and pleasure. Each of the items gives a short statement to respondent (e.g., "I feel that life is full of opportunities"), respondents are asked to rate how often this statement applies to them on a 4-point Likert-type scale. A short version of the Big-Five inventory was introduced in the seventh wave of SHARE.[7] This Big-Five inventory contains 11 items measuring five different dimensions of personality (abbreviated as OCEAN). Respondents receive brief statements asking them to what extent they agree or disagree a certain trait applies to them on a 5-level Likert-type scale. Both item batteries are in a non-proxy part of the questionnaire and should only be administered to the respondent.

As the first step, I compare the item nonresponse within the two batteries across the two housing situations. I will mainly focus on the average number of missing values across all CASP and Big-Five items and the share of respondents having at least one missing item in the two batteries. I will

[5]This variable is measured on a 6-level scale. I merged the levels "almost never" and "never" resulting in a 5-level scale.

[6]This variable is measured on a 4-level scale, reaching from "Bad" to "Very good"

[7]This is why the sample size reduces from 79,655 cases (2,002 of those in institutions) to 52,446 cases (919 of those in institutions).

also consider different types of item nonresponse, namely refusals, Don't know responses, and missing values caused by the incapacity of respondents to reply. If respondents are not capable of answering CASP or Big-Five questions themselves, SHARE asks their interviewers to either assign a system missing value or indicate another type of item nonresponse, usually "Don't know". According to the SHARE user support team, the routing of respondents who relied on proxies did not work properly in wave 7, which is why those respondents received more Don't know replies and nearly no system missing values. This routing error should not affect my aggregate analysis of item nonresponse and the conclusions.

As second step, I compare response patterns in Big-Five items. Due to the uneven number of scale points in Big-Five items, a midpoint is intended to measure a neutral attitude of respondents ("Neither agree nor disagree"). While previous analysis showed clear advantages of a 5-point Agree-Disagree scale compared to longer scales (Revilla et al. 2014), other authors found a large number of respondents hiding their non-attitude by selecting the midpoint (Raaijmakers et al. 2000; Sturgis et al. 2014). When probing respondents who selected the midpoint, it turned out that a majority of those respondents did actually not know whether they should agree or disagree to the given statement (Sturgis et al. 2014). This could help to explain the correlation of an increasing selection of midpoints when the clarity of items decreases (Velez and Ashworth 2007). The selection of middle response options is labeled as weak satisficing (Roberts et al. 2019) and was shown to have a negative impact on the reliability of attitude measures (Alwin and Krosnick 1991). This is why I compare the share of middle response selection between institutionalized respondents and respondents living in private households. Additionally, I briefly examine the selection of endpoints for Big-Five items as another response pattern, which proved to be more prevalent among older respondents (He et al. 2014; Kieruj and Moors 2013). The proportion of midpoints, strong disagreement, and strong agreement across 11 items was calculated within respondents. I excluded missing values from this calculation and only used their substantive replies as denominator. As a limitation, I cannot directly associate response patterns with data quality without a proper experimental designs or probing respondents to provide more information about the underlying reason for their choices. Still, this comparison also serves to prepare the subsequent analysis of acquiescence. Respondents making overly use of midpoints will obtain lower scores of acquiescence.

Batteries of items that intend to measure a latent concept with multiple dimensions are well suited to analyze data quality. Any deviations of empirical findings from theory might indicate measurement errors and a worse data quality. In first exploratory analyses of the CASP inventory the 12 items did not load as expected by theory (Hyde et al. 2003). This finding is not unique and was already observed in an earlier study (Borrat-Besson et al. 2015). In the light of those empirical difficulties I decided to restrict my analysis of the CASP inventory to the analysis of item nonresponse and focus on the better developed Big-Five inventory. The short version of the Big-Five inventory was already used in earlier studies to analyze data quality (Lechner and Rammstedt 2015; Rammstedt et al. 2010; Rammstedt and Kemper 2011; Soto et al. 2008) because the items are balanced within each latent dimension (e.g., "being lazy" and "doing a thorough job"). In theory, indicating agreement with one item of a dimension would require respondents to answer with disagreement

to the second item of the respective dimension, resulting in a highly negative correlation. When measuring the Big-Five items on the 5-point scale reaching from 1 to 5, the mean of items within a dimension is expected to be 3.0. A deviation from this pattern within dimensions and across the 5 dimensions, in other words any values higher than 3.0, indicate a tendency to agree to mutually exclusive statements to a larger extent than expected by theory.[8] This kind of acquiescence was labeled as weak satisficing and is a sign of a worse data quality.

4.3.2 Propensity score matching

In the course of my analysis I will first compare various indicators of response behavior and data quality for institutions and private households to check whether respondents in institutions are harder to interview as expected by most survey researcher. Likely differences between the two populations are also driven by confounding variables, since institutionalized elderly differ from respondents in private households on average, in particular in terms of age, health, and cognitive impairment. The statistical method of propensity score matching (PSM) is commonly used to overcome the selection bias in non-experimental studies that are interested in isolating the impact of a certain treatment or quasi-treatment by creating a control group post hoc (Caliendo and Kopeinig 2005; Dehejia and Wahba 2002). PSM helps to control for differences in confounding variables by identifying cases living in private households that bear a greater resemblance to institutionalized residents in a number of pre-defined characteristics. This method helps to examine whether institutionalization leads to a more problematic response behavior or whether this is rather caused by other characteristics that can be found in institutions as well as in private households.
After comparing all respondents, I match on a small number of basic socio-demographic variables, namely respondents' age, their gender, whether they are widowed or not, and their level of education (labeled as *Basic PSM* in this paper). Age is grouped into 4 categories, reaching from 50 to 59 years, 60 to 69 years, 70 to 79 years and 80 years and older. Gender is coded as 0 if respondents were male and 1 if respondents were female, while widowhood is also dichotomous, identifying widowed respondents with a value of 1. The level of education is measured with the ISCED scheme that aims to generate a comparable measurement of educational achievements across countries. SHARE uses the 7-level ISCED-97 scale, which ranges from 0 (pre-primary education) to 6 (second stage of tertiary education).
As the third step, I add three health-related variables to the above-mentioned set of socio-demographic variables in the propensity score matching (*Full PSM*). In previous studies, a combination of socio-demographic and health-related variables proved to be decisive in increasing the likelihood of institutionalization in older age cohorts (Schanze and Zins 2019). In this analysis, the

[8] When calculating the acquiescence scores *within dimensions*, a missing value in one or both of the paired items led to an exclusion of the respective respondent for this given score. For the *overall acquiescence* score, I only included respondents if they answered at least five or more out of the 10 Big-Five items used for this score. This led to the exclusion of a small number of respondents (N = 91) with a higher degree of item nonresponse. Nevertheless, in a robustness check it did not lead to substantive changes in the results presented in Table 4.8.

diagnosis of dementia serves as a dichotomous measure of cognitive illness and potential cognitive decline. A 7 point count variable on limitations in activities of daily living (ADL) measures the ability of respondents to cope with 6 different tasks of daily living.[9] And finally, the self-rated health of respondents is a third health-related indicator measuring the subjective perception of respondents regarding their state of health on a 5 point scale reaching from poor, fair, good, and very good to excellent.

When selecting variables for the propensity score matching, a number of conditions need to be fulfilled. As a first requirement variables used for matching should influence the treatment and the outcome simultaneously (cf. Caliendo and Kopeinig 2005). All variables used in the matching procedure are expected to influence the housing situation as quasi-treatment *and* the response behavior and data quality as outcomes. This is especially the case for age, education, diagnosis of dementia, and the self-rated health. I only included explanatory variables with a statistically significant impact on the likelihood of institutionalization and tried to limit the number of matching variables as much as possible to avoid an increase of variance in propensity scores resulting in a lower number of matched cases (cf. (Caliendo and Kopeinig 2005)).

Following the conditional independence assumption, the variables used for matching should not be influenced by the treatment or the anticipation of the treatment (Caliendo and Kopeinig 2005). Age, gender, widowhood and education should definitely be exogenous and independent from the housing situation. The same holds for dementia as a health condition, which should not be caused by institutionalization. The perception of what people are able to do (limitations in ADL) and their self-rated health could be more influenced by their context. However, it is difficult to postulate whether respondents will rate their health as better or worse due to their institutional environment. Some respondents might rate their abilities or health worse, simply because they live in a retirement or nursing home, whereas other institutionalized respondents might compare themselves with their cohabitants who are worse off and therefore arrive at a better self-rated health. However, I do not expect a systematic distortion into one direction and assume that the potential deviation is negligible for the following analyses.

As matching algorithm I used radius matching (Dehejia and Wahba 2002). The conditional propensity of being institutionalized is calculated in a multivariate logistic regression with the above-mentioned explanatory variables within each country. Radius matching excludes control cases that are too different from institutionalized cases by applying a caliper, which is often defined as 25 percent of the standard deviation of the propensity scores explaining institutionalization (Lunt 2014). I also applied the caliper of this size. To avoid matches across countries and different health-care systems, I used the countries as strata and only matched institutionalized respondents within their respective countries.[10] institutionalized residents, because the group of institutionalized residents is too small to run a meaningful logistic regression as part of the PSM. The stratified matching is meaningful to avoid an undetected impact of confounding contextual variables that

[9] Dressing, including putting on shoes and socks; Walking across a room; Bathing or showering; Eating, such as cutting up your food; Getting in or out of bed; Using the toilet, including getting up or down

[10] Which is why I excluded countries with 15 or less

exert a differing influence between countries. To increase the number of control cases for each institutionalized case within countries, I did not stratify for waves but matched across the three waves. An institutionalized respondents in Austria in wave 5 could only be matched to another Austrian respondent, but the control case could also be interviewed in wave 7 or vice versa. This matching strategy yielded a sufficiently large number of control cases in all of the 18 countries (Lunt 2014). Only in the Netherlands a little less than 20 respondents in private households were available as a potential match for each institutionalized respondents. In all the other countries, the number of control cases was well above this threshold, reaching 115 control cases in Finland as the maximum. As a result of the radius matching, all control cases within the bounds of the caliper receive a weight determined by their distance to the members of the *institutionalized* group measured by the propensity score. The weight for institutionalized cases equals 1, while the weight for respondents in private households ranges between 0.00047 and 1.5 in the basic PSM (SD = 0.056) and between 0.001 and 7.5 in the full PSM (SD = 0.11).

4.3.3 Exploration of the Propensity Score Matching

Table 4.1 shows the number of unique cases observed in institutions in the three waves and the respondents I lost in the process of matching. The large number of control cases in the pooled SHARE samples allowed for excellent matches with a very small number of matches beyond the border of the pre-defined caliper. Depending on the type of matching, 33 institutionalized respondents in the socio-demographic PSM (Basic PSM) and 113 institutionalized respondents in the second, more advanced iteration (Full PSM) are excluded because they had missing values in one of the variables used for matching. The caliper prevents matches for 3 institutionalized respondents for the basic socio-demographic characteristics. While matching with socio-demographic and health-related variables, 44 institutionalized respondents are not matched due to the absence of comparable control cases living in private households. Checking those 44 institutionalized respondents, the data shows that unmatched institutionalized residents do not differ from matched cases in terms of age, gender, widowhood, education and self-perceived health. However, unmatched institutionalized residents have a higher likelihood of dementia and face more limitations in ADL. Adding it up, 1.8 percent of the institutionalized respondents could not be matched in the Basic PSM due to missing values and the absence of control cases, while 7.8 percent could not be matched in the Full PSM due to those reasons. To control for the impact of the systematic drop-outs, I will always show and discuss the results for the unmatched sample. Hence, I will refer to unmatched institutionalized cases in the presentation of results to assess whether the drop-outs are not at random.

Propensity score matching should help to counterbalance the impact of confounding variables while investigating the statistical impact of the housing situation on the response behavior. In a logistic regression analysis following both rounds of matching the large majority of explanatory variables did no longer have a statistically significant influence on the likelihood of institutionaliza-

tion. For the socio-demographic matching only age (Odds Ratio = 1.10, p = .003) and widowhood (OR = .88, p = .023), for the socio-demographic and health-related matching only age (OR = 1.11, p = .004) had a very small, nearly absent but statistically significant explanatory power.

Table 4.1: Overview of institutionalized cases used in PSM and reasons for losing cases in the PSM

		Total inst. pop.	Missing values	Beyond caliper	**Matched cases**
Unique cases	**Basic PSM**: Socio-dem. var.	2,002	33	3	**1,966**
in waves 5 to 7	**Full PSM**: Socio-dem. + health var.		113	44	**1,845**

Since I am using panel data, a previous participation of respondents and resulting panel experience could have a presumably positive impact on the response behavior. Panel respondents are more experienced with the various steps they need to perform while answering survey questions (Kroh et al. 2016). While Kroh and colleagues (2016) found more reliability with growing panel experience, another study did not detect any consistent impact of panel experience on multiple quality indicators in two U.S. panels (Sun et al. 2019). Checking the panel experience of respondents in my unmatched dataset, institutionalized respondents have a slightly longer period of panel participation than respondents in private households, partly related to the higher age of institutionalized respondents in the data. When matching institutionalized respondents with respondents living in private households the panel experience does not differ any longer. On average, both groups participated in 3.7 previous waves. Hence, previous panel experience should have no impact on the results I will present in the following section. The same holds for the origin of respondents. If the survey interview is conducted in a language other than the native language of respondents, additional difficulties arise and might have a negative impact on the quality of survey data. In the pooled dataset, roughly 12 percent of the respondents were not born in the country of interview in private households and institutions. When applying the two PSM conditions, this share does not change much. Hence, differences between the two populations cannot be explained with underlying differences in the number of foreign-born residents with potential language difficulties.

4.4 Results

4.4.1 Proxy interviews and interviewer observations

Interviewers decide whether a proxy respondent is required who can answer a part of the questionnaire on behalf of the respondent. Moreover, following the survey interview, interviewers in SHARE rate their respondents in terms of need for clarification and their willingness to answer

the questions. As mentioned earlier, those measures can be classified as subjective measures of data quality and might also be driven by a more negative perception of institutionalized residents by the survey interviewers.

Table 4.2 shows the share of proxy respondents for the entire pooled sample and for the respondents analyzed under the two different matching conditions. In the first row of Table 4.2, the difference between respondents living in private households and institutions is obvious. Proxy respondents are required with a much higher likelihood in institutions than in private households. Every third interview relied on proxy informants in contrast to only 6.5 percent in all private households. However, when controlling for age, gender, widowhood, and education, and even more so when matching respondents additionally with health-related variables, the share of proxy interviews increases dramatically in private households as well. Under the second matching condition, proxy interviews occur only 4.1 percentage points more often in institutions than in private households. As a consequence of systematic item nonresponse and unmatched cases, the share of proxy interviews slightly decreases in institutions. This observation can be made for several of the following results, indicating that institutionalized respondents who are lost in the course of PSM are more difficult to interview.

Table 4.2: Proxy interviews in three waves of SHARE, comparing private households and institutions

		Private households	**Institutions**	**Diff.**
All cases	**Proxy used** (%)	6.5	33.1	26.6***
	Total N	77,653	2,002	
Basic PSM	**Proxy used** (%)	10.8	32.7	22.0***
	N (weighted)	1,966	1,966	
Full PSM	**Proxy used** (%)	27.4	31.5	4.1*
	N (weighted)	1,845	1,845	

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate logistic regressions

Table 4.3 shows the rating of interviewers regarding respondents' need for additional clarifications while answering the survey questions. The table displays a similar pattern like Table 4.2 on proxy interviews. Large differences occur in the first row comparing the two groups of respondents without matching them. Those differences get smaller when matching respondents on socio-demographic variables and continue to decline when matching on health-related variables in addition to the socio-demographic variables. Health seems to be the stronger driver of more difficult interviews compared to age, gender, education, or widowhood. Even though the gap between institutionalized and residents in private households closes, a statistically significant difference

between the two group still remains. In the Full PSM condition, 9.1 percent of institutionalized respondents ask for clarification always or very often, whereas this is only the case for 7.5 percent of comparable respondents living in private households. A check with a dummy variable capturing the origin of respondents[11] showed a positive impact on the need for clarification for unweighted data. However, in both PSM conditions the origin of respondents did not have a statistically significant influence on the dependent variable next to institutionalization.

Table 4.3: Respondents asking for clarification, comparing private households and institutions across three waves of SHARE

		Private households	**Institutions**
All cases	**(Almost) never** (%)	77.0	55.5
	Now and then (%)	16.6	25.3
	Often (%)	3.4	9.5
	Very often (%)	1.5	5.0
	Always (%)	1.5	4.7
	Coefficient	0.44***	
	Total N	76,880	1,839
Basic PSM	**(Almost) never** (%)	66.2	55.8
	Now and then (%)	23.8	25.3
	Often (%)	5.7	9.3
	Very often (%)	2.6	4.9
	Always (%)	1.8	4.7
	Coefficient	0.27***	
	N (weighted)	1,935	1,809
Full PSM	**(Almost) never** (%)	60.7	55.7
	Now and then (%)	23.5	25.8
	Often (%)	8.4	9.4
	Very often (%)	3.9	4.6
	Always (%)	3.6	4.5
	Coefficient	0.10**	
	N (weighted)	1,769	1,728

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate OLS regressions with institutionalization as explanatory variable.

[11] 0 = Born in country of interview, 1 = Not born in country of interview

Table 4.4: Respondents' willingness to answer questions, comparing private households and institutions across three waves of SHARE

		Private households	**Institutions**
All cases	**Bad** (%)	2.0	8.5
	Fair (%)	8.5	17.2
	Good (%)	22.0	26.1
	Very good (%)	67.5	48.2
	Coefficient	-0.41***	
	Total N	76,926	1,841
Basic PSM	**Bad** (%)	2.9	8.6
	Fair (%)	11.0	17.0
	Good (%)	23.9	26.1
	Very good (%)	62.2	48.4
	Coefficient	-0.31***	
	N (weighted)	1,937	1,812
Full PSM	**Bad** (%)	6.7	8.0
	Fair (%)	15.9	16.4
	Good (%)	24.5	26.2
	Very good (%)	52.9	49.4
	Coefficient	-0.06*	
	N (weighted)	1,776	1,731

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate OLS regressions with institutionalization as explanatory variable.

As the last indicator of hard-to-interview respondents, I consider interviewer ratings regarding respondents' willingness to answer (see Table 4.4). In contrast to the previous indicator, this variable might be more difficult to measure but also advances a behavior that can be clearly linked to a worse data quality. As Krosnick (1991) pointed out, respondents' motivation is a crucial aspect in explaining satisficing and its negative impact on survey data. Apparently around 8.5 percent of all institutionalized residents and 2 percent of respondents in all private households are very unwilling to answer. At the other end of the scale, every second respondent in institutions and two out of three respondents in private households are rated as very cooperative. The lower willingness in institutions cannot be fully related to institutionalization as such. As the two steps of the propensity score matching show, socio-demographic variables and health are the true explanatory factors for respondents' willingness to answer, whereas institutionalization only has a marginal impact on the dependent variable.

4.4.2 Item nonresponse

As Koch and Blohm (2009) emphasized in their analysis of item nonresponse, the specific type of item nonresponse matters and should make a difference in the analysis of missing values. Tables 4.10 and 4.11 in the Appendix show the average proportion of the different types of item nonresponse for the CASP and Big-Five items for the two different types of housing and the different PSM conditions.

For the CASP items, a large fraction of the missing values are system missing values, especially in institutions (roughly 10%). The majority of those cases can be explained with the dropout of respondents who relied on proxies in other parts of the survey interview in wave 6. Due to the malfunctioning routing mentioned above, Big-Five items and CASP items administered in wave 7 did not receive many "Other missing" codes. For this wave the large majority of proxy cases were recorded with a Don't know answer in both populations (see Table 4.11 in the Appendix).

Since system missing values indicate that respondents were not capable to answer survey questions themselves, I decided to keep this code in the aggregate analysis of item nonresponse. Tables 4.5 and 4.6 show the share of respondents with at least one missing value for 12 CASP items and for 11 Big-Five items respectively. Moreover, the tables also show the average number of missing values for both item batteries.

Table 4.5: Item nonresponse in 12 items of the CASP inventory in three waves of SHARE

		Private households	Institutions	Diff.
All cases	**At least 1 mv** (%)	7.3	35.7	28.4***
	Mean no. of mv (%)	0.5	3.3	2.8***
	Total N	77,653	2,002	
Basic PSM	**At least 1 mv** (%)	12.4	35.6	23.2***
	Mean no. of mv (%)	0.9	3.3	2.4***
	N (weighted)	1,966	1,966	
Full PSM	**At least 1 mv** (%)	24.5	32.5	8.0***
	Mean no. of mv (%)	2.3	3.0	0.7**
	N (weighted)	1,845	1,845	

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate logistic regressions and negative binomial regressions

In institutions, roughly every third respondent did not reply to a least one item in the CASP item

battery. On average, 3.3 items out of the 12 items were suffering from item nonresponse within this group of respondents. The equivalent numbers are much lower among the 77,653 respondents living in private households. A large majority of about 92.7 percent of the respondents did not have any missing values in the CASP items, on average only 0.5 items had a missing value. The gap in item nonresponse gets closer when respondents are matched on socio-demographic variables and health-related variables. The share of respondents in private households with a least one missing value nearly doubles when matching for socio-demographic variables, and it doubles again when adding health-related variables to the PSM. In the full PSM condition, 24.5 percent of all respondents living in private households had at least one missing values, which is still 8 percentage points less than their institutionalized counterparts. Again, the systematic drop-out of certain institutionalized respondents in the matching leads to an improvement of this data quality indicator, namely to a decrease of missing values in the institutionalized group.

Table 4.6: Item nonresponse in 11 items of the Big-Five inventory in wave 7 of SHARE

		Private households	**Institutions**	**Diff.**
All cases	**At least 1 mv** (%)	4.1	31.8	27.7***
	Mean no. of mv (%)	0.3	2.9	2.5***
	Total N	51,527	919	
Basic PSM	**At least 1 mv** (%)	7.8	31.5	23.8***
	Mean no. of mv (%)	0.6	2.8	2.2***
	N (weighted)	1,184	907	
Full PSM	**At least 1 mv** (%)	17.5	27.6	10.1***
	Mean no. of mv (%)	1.6	2.4	0.8***
	N (weighted)	972	829	

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate logistic regressions and negative binomial regressions

For the Big-Five items the overall level of item nonresponse is a bit lower compared to the CASP items. Apart from that, the pattern looks similar to Table 4.5. The size of the gap between both groups closes in the process of matching respondents, but its size is still large enough to be statistically significant. When matching on socio-demographic and health-related variables, item nonresponse for at least one item is more likely by about 57.7 percent in institutions. Institutionalized residents answer nearly one item less on average than comparable respondents living in private households. In both item batteries a significant difference in the degree of item nonresponse persists even though the differences shrink markedly.

Hard-to-interview respondents who require information by proxies in other parts of the survey

achieve a higher level of item nonresponse in both item batteries. The differences to all other respondents amount to 4 more missing items in private households (p = 0.00) and 2.5 more missing items in institutions (p = 0.00). Respondents with a better rating of their willingness abstain less often from replying in both item batteries (coeff. = -0.89, p = 0.00 for the CASP items), while those respondents asking more often for clarification also achieve a higher average number of missing values in CASP and Big-Five items (coeff. = 0.43, p = 0.00 for the CASP items). As for the proxy indicator, differences for interviewer ratings are less distinct in institutions than in private households, even not statistically significant for the interviewer rating of asking for clarification. In private households, the impact of the three indicators gets also constantly smaller when matching respondents with their institutionalized counterparts. In the group of respondents who are more challenging to interview due to their age or health, even those respondents who are not using the support by proxy or are very willing to be interviewed still produce more missing values than respondents in the full sample.

4.4.3 Response patterns in Big-Five items

Before examining acquiescence in the Big-Five inventory, I briefly investigate the response patterns in this item battery in addition to item nonresponse. Table 4.7 and Table 4.12 (in the Appendix) show the share of middle response options and strong agreements in 11 Big-Five items in relation to all substantive responses. Thus, items with missing values are excluded from the denominator.

Table 4.7: Share of Big-Five items answered with middle response options of all substantive replies (excluding missing values) in wave 7 of SHARE

		Private households		**Institutions**		**Diff.**
		(%)	SD	(%)	SD	
All	**Middle resp. option**	15.4	0.16	17.4	0.18	2.0**
resp.	Total N	50,038		703		
Basic	**Middle resp. option**	15.2	0.16	17.4	0.18	2.2**
PSM	N (weighted)	1,122		695		
Full	**Middle resp. option**	16.2	0.18	17.4	0.18	$1.2^{n.s.}$
PSM	N (weighted)	841		666		

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate OLS regressions

Institutionalized respondents pick the middle response option more often than respondents living in private households (see Table 4.7). This difference amounts to nearly 2 percentage points and

it is statistically significant (p = .001). The gap gets even bigger when matching respondents for socio-demographic variables, before shrinking when adding health-related variables to the set of socio-demographic variables in the PSM. In the full PSM condition the difference between the two types of housing is not statistically significant any longer (p = .188).
Looking at strong agreement and strong disagreement, institutionalized respondents tend to give more negative extreme responses ("Strongly disagree") in Big-Five items (results not shown). The difference in the respective proportion amounts to 1.2 percentage points in the full sample (p = .013) and gets smaller and statistically insignificant when matching respondents. On the other *end of the response scale*, institutionalized respondents use the "Strongly Agree" response option slightly less often than respondents living in private households: 18.4 percent of all items answered by institutionalized respondents are answered with strong agreement, while this is the case for 19 percent of the items answered by respondents in private households (see Table 4.12 in the Appendix). However, this difference is not statistically significant (p = .375), the same observation holds for the two matching conditions.
Respondents relying on proxy interviews in other parts of the survey interview tend to reply more often with middle response categories (coeff. = 0.07, p = 0.00) and less often with strong agreement (coeff. = -0.03, p = 0.00). The same pattern can be observed for those respondents who ask more often for clarification, while respondents with a better rating of their willingness respond significantly less often with the middle response option in private households (coeff. = -0.06, p = 0.00) and institutions (coeff. = -0.05, p = 0.00). Notwithstanding a lack of further information about the underlying reasons for this response pattern, those results could lead to a cautious support of previous analyses on the background of middle response options (Alwin and Krosnick 1991; Sturgis et al. 2014), linking it to respondents who are more difficult to interview in institutions and private households. In the Big-Five items, it is used more often by institutionalized respondents and by older respondents with a worse health condition.

4.4.4 Acquiescence and factor structure in the Big-Five inventory

Similar to earlier studies, I use the Big-Five inventory to check for acquiescence as another behavior leading to a worse data quality (Rammstedt et al. 2010; Rammstedt and Kemper 2011; Soto et al. 2008). Table 4.8 shows a simple measure of acquiescence with values higher than 3.0 indicating acquiescence. A comparable level of acquiescence can be found in the Big-Five items in SHARE as in previous analyses of this item battery (Rammstedt et al. 2010; Rammstedt and Kemper 2011; Soto et al. 2008). Moreover, somewhat surprisingly it is more prevalent in private households than in retirement and nursing homes. The extent of acquiescence differs across the different Big-Five dimensions, with the highest degree of acquiescence for two items measuring extraversion and two items measuring agreeableness. Those differences in acquiescence between domains could be caused by different living circumstances, namely if certain concepts do not apply to respondents' lives, or by comprehension issues (Soto et al. 2008). However, both explanations

would rather lead to the hypothesis that institutionalized respondents with their more special living circumstances and larger comprehension issues (see Table 4.3) show a more acquiescent response behavior. This is apparently not the case.

In contrast to all previous indicators of data quality analyzed in this paper, acquiescence does *not* change when running through the two stages of propensity score matching. Significant differences between institutionalized respondents and respondents living in private households do not get smaller when matching for socio-demographic variables and even when adding health-related variables.

Thus, earlier conclusions regarding mixed results of acquiescence (Roberts et al. 2019) can be cautiously confirmed in the present analysis. Apparently acquiescence is independent from respondents' ability as measured by age, gender, widowhood, education, dementia, self-rated health, and limitations in activities of daily living. Even when controlling for those variables, institutionalized residents show a more desirable respondent behavior than their counterparts in private households. Interestingly, the hard-to-interview indicators exert the opposite influence on acquiescence compared to earlier findings for item nonresponse and response patterns. Respondents with proxy interviews in other parts of the interview achieve a lower level of acquiescence (coeff. = -0.05, p = 0.00), while respondents with a better rating of their willingness have a slightly stronger tendency for acquiescence (coeff. = 0.04, p = 0.00). The less pronounced motivation in institutions (see Table 4.4) does not lead to a more acquiescent response behavior. Respondents asking more often for additional clarification have a slightly smaller level of acquiescence (coeff. = -0.02, p = 0.00). Due to the small effect size, the latter indicator is not statistically significant in institutions.

Which factors could explain the persisting gap in acquiescence between the two types of housing? First, the higher item nonresponse among institutionalized respondents causes the exclusion of the most critical respondents (cf. Table 4.6). Second, the higher share of midpoints among the group of institutionalized respondents leads to a mean value closer to 3.0. While both explanations are certainly true, they do not explain why the gap between institutions and private households remains as it is throughout the PSM, while the gaps in item nonresponse and selection of midpoints get much smaller. Thus, the lower acquiescence in institutions could be also explained with the housing situation as such or - more likely - with additional confounding variables I did not control for or that were not measured by SHARE. Referring to earlier research, this could be a psychological trait to agree to statements to be perceived more positively by interviewers, a lower cognitive interest in answering the questions properly (satisficing), or interviewer effects (Olson and Bilgen 2011). In other words, institutionalized respondents may be a little less worried about the opinions that interviewers might form about them while conducting the interview. The lower acquiescence among hard-to-interview respondents could point in the same direction.

Table 4.8: Acquiescence in various latent dimensions of the Big-Five inventory and across all Big-Five items in wave 7 of SHARE

		Private households		**Institutions**		**Diff.**
		Mean	SD	Mean	SD	
All resp.	Extraversion	3.36	0.70	3.28	0.71	-.07**
	Agreeableness	3.20	0.63	3.17	0.61	-.03$^{n.s.}$
	Conscientiousness	3.13	0.63	3.09	0.65	-.04$^{n.s.}$
	Neuroticism	3.07	0.67	3.03	0.69	-.03$^{n.s.}$
	Openness	3.06	0.82	3.00	0.88	-.07*
	Mean N (SD)	49,819 (61.8)		670 (8.9)		
	Big-Five, all items	3.16	0.35	3.11	0.37	-.05***
	Total N	49,961		688		
Basic PSM	Extraversion	3.38	0.70	3.28	0.70	-0.10***
	Agreeableness	3.19	0.61	3.17	0.61	-.01*
	Conscientiousness	3.08	0.63	3.09	0.65	.01$^{n.s.}$
	Neuroticism	3.08	0.68	3.03	0.69	-.05***
	Openness	3.01	0.83	3.00	0.88	-.02$^{n.s.}$
	Mean weighted N (SD)	1,110 (2.9)		662 (8.4)		
	Big-Five, all items	3.15	0.35	3.11	0.38	-.04***
	N (weighted)	1,116		680		
Full PSM	Extraversion	3.37	0.70	3.27	0.71	-.10***
	Agreeableness	3.21	0.62	3.17	0.61	-.04***
	Conscientiousness	3.10	0.65	3.09	0.65	-.01*
	Neuroticism	3.09	0.70	3.04	0.69	-.05***
	Openness	3.05	0.83	3.00	0.88	-.05***
	Mean weighted N (SD)	821 (3.3)		638 (7.8)		
	Big-Five, all items	3.16	0.35	3.11	0.38	-.05***
	N (weighted)	829		656		

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate OLS regressions

The analysis of the factor structure of the Big-Five items sheds a light at the impact of acquiescence on the survey data. Table 4.9 shows the results of a principal component factor analysis with a varimax rotation for 10 items of the Big-Five inventory. Each block refers to one of the three configurations of the propensity score matching.

Table 4.9: Principal component factor analysis (varimax rotated) of Big-Five items (not corrected for acquiescence)

	I see myself as sb. who...	**Private households**					**Institutions**				
All cases	...is reserved	**.89**	.02	.07	-.08	.08	**-.86**	.03	.12	.04	.10
	...is outgoing sociable	**-.63**	.04	.45	.12	.05	**.69**	.27	.20	-.18	.17
	...tends to find fault w. others	-.10	**.70**	-.07	-.25	.08	.19	**-.50**	-.13	.29	.42
	...is considerate and kind	-.11	-.18	**.67**	.14	.06	.13	**.71**	.18	-.14	.19
	...does a thorough job	.05	-.13	**.73**	.05	-.02	.06	**.69**	-.20	.01	.21
	...tends to be lazy	.18	**.63**	-.35	.13	.05	-.15	**-.50**	.46	-.09	.40
	...is relaxed, handles stress well	-.04	.05	.16	**.82**	.10	.10	.11	.15	**-.80**	.12
	...gets nervous easily	.14	.15	.03	**-.79**	.11	-.06	.00	.26	**.80**	.13
	...has an active imagination	-.03	.47	.48	.06	-.43	-.03	.17	-.12	.00	**.83**
	...has few artistic interest	.05	.04	.01	.01	**.93**	.01	.03	**.80**	.10	-.18
	% of explained variance	*12.8*	*11.9*	*15.9*	*14.3*	*10.9*	*13.2*	*16.0*	*11.0*	*14.3*	*12.0*
	Total N			49,439					630		
Basic PSM	...is reserved	**.86**	-.06	.11	-.08	-.01	**-.87**	.03	.09	.03	.11
	...is outgoing sociable	**-.70**	-.07	.32	.09	.09	**.68**	.29	.18	-.19	.22
	...tends to find fault w. others	-.03	**.89**	-.05	-.10	-.02	.18	-.48	.44	.30	-.13
	...is considerate and kind	-.24	-.40	**.51**	.09	.10	.12	**.72**	.17	-.14	.17
	...does a thorough job	.06	.04	**.81**	.12	.05	.06	**.70**	.18	.02	-.20
	...tends to be lazy	.11	.20	**-.60**	.09	.13	-.16	-.48	.41	-.09	.46
	...is relaxed, handles stress well	-.05	-.02	.10	**.84**	-.02	.10	.12	.12	**-.80**	.14
	...gets nervous easily	.12	.13	-.01	**-.79**	-.02	-.06	.01	.13	**.80**	.26
	...has an active imagination	-.09	.34	.29	.09	**.62**	-.02	.19	**.82**	-.01	-.12
	...has few artistic interest	.02	.12	.07	.03	**-.87**	.01	.02	-.18	.11	**.80**
	% of explained variance	*13.3*	*11.6*	*15.0*	*13.9*	*11.7*	*13.0*	*16.1*	*12.1*	*14.4*	*11.1*
	N (weighted)			1,091					624		
Full PSM	...is reserved	**.90**	.02	.07	.10	.06	**-.87**	.04	.09	.02	.12
	...is outgoing sociable	**-.64**	-.06	.50	-.07	.06	**.68**	.31	.17	-.19	.19
	...tends to find fault w. others	-.09	**.62**	-.08	.29	-.18	.18	**-.51**	.49	.24	-.07
	...is considerate and kind	-.05	-.21	**.69**	-.15	.17	.14	**.74**	.15	-.12	.12
	...does a thorough job	.07	-.37	**.61**	-.02	-.20	.06	**.66**	.20	.01	-.27
	...tends to be lazy	.15	**.75**	-.16	-.07	.19	-.17	-.40	.36	-.09	**.59**
	...is relaxed, handles stress well	-.04	.03	.21	**-.79**	.04	.09	.17	.10	**-.80**	.15
	...gets nervous easily	.14	.09	.07	**.82**	.09	-.06	.04	.13	**.80**	.25
	...has an active imagination	-.05	.36	**.54**	-.02	-.48	-.04	.20	**.82**	.00	-.07
	...has few artistic interest	.03	.09	.02	.04	**.87**	.03	.11	-.25	.15	**.74**
	% of explained variance	*12.7*	*12.8*	*14.9*	*14.3*	*11.4*	*13.1*	*15.8*	*12.3*	*14.1*	*11.3*
	N (weighted)			802					602		

Relying on the Kaiser criterion requesting an eigenvalue higher than 1 to be considered as a distinct factor, the factor analysis would only suggest the extraction of 4 factors for respondents living in private households in *all* the three conditions (cf. Rammstedt and Kemper 2011). The fifth factor only reaches eigenvalues between 0.95 and 0.9. In contrast, the extraction of 5 factors is suggested for institutionalized respondents. For the purpose of comparability and in the light of the well-established theory of the concept, I nevertheless extract 5 factors for private households.

Table 4.9 shows the factor loadings for the Big-Five items that are not corrected for acquiescence. In both populations only the extraversion and neuroticism items load as expected and form their discrete latent dimensions. In private households, the openness items also load nearly as expected, whereas those two items load on different latent dimensions in institutions. The two dimensions agreeableness and conscientiousness are muddled with cross-loadings and do not form distinct dimensions in both populations. Ambiguous factor loadings were also found in earlier studies for those Big-Five items (Lechner and Rammstedt 2015; Rammstedt and Kemper 2011).
The changes in the factor structure following the propensity score matching can be considered as minor. For institutionalized respondents, the factor structure and the strength of factor loadings does not change substantially. For private households, changes are a bit stronger for some of the items, however, this can probably be explained with the reweighting of cases in the course of propensity score matching. The factor structure as such does not undergo drastic changes. The total share of explained variance remains very stable (65.8% with all cases compared to 66.1% in the Full PSM). Like for the analysis of acquiescence presented in Table 4.8, the variables used for matching do not have a strong impact on the Big-Five factor structure and related measurement error.
Does the acquiescence lead to the blurred factor structure? Table 4.9 does not control for acquiescence as suggested by Rammstedt and Kemper (2011). This is done in a second step by ipsatizing the data. The overall mean of a respondent is subtracted from each of the Big-Five items to account for the response tendency. The result of the subtraction is divided by respondents' standard deviation. Indeed, the factor analysis for the ipsatized Big-Five items clearly results in the factor structure as predicted by the theory (results not shown). This suggests a strong impact of acquiescence on the Big-Five items.
Summarizing the results for the two populations, acquiescent response behavior is significantly less prevalent in institutions than in private households, even when comparing institutions to private households while applying PSM. The factor structure in the Big-Five inventory is negatively affected by acquiescence in both populations. It is rather difficult to conclude whether the factor structure is closer to the theory in institutions or private households. Apparently, the number of factors is more in line with the theory in institutions than in private households, where only 4 instead of 5 factors should have been extracted. However, the factor loadings of items does not show any distinct advantage of the theory in institutions as compared to private households.

4.5 Conclusion

The present paper investigated whether institutionalized residents are more difficult to interview than respondents living in private households. This claim is often made by survey researchers, usually followed by an exclusion of this group from the target population in many social surveys. Difficulties to interview institutionalized residents can be expected due to their higher age, a larger

number of cognitive and functional impairment, and a worse health condition. However, those characteristics are not unique to institutionalized residents but can also be observed in private households. This paper aimed to answer the research questions whether institutionalized respondents perform worse in survey interviews and whether institutionalization as such leads to a worse quality of survey data or whether it can be rather explained with confounding variables.

As there is not the *one* indicator for bad or good response behavior, I compiled evidence on a number of different indicators to get a good overview of difficulties in survey interviews and the data quality. I made use of interviewer observations of respondents' performance, compared the rate of proxy interviews in private households and institutions, analyzed item nonresponse in two item batteries, and also had a look at the response patterns and acquiescence in the Big-Five inventory. To isolate the impact of institutionalization, I used propensity score matching to identify respondents in private households who bear a greater resemblance with institutionalized residents than the entire SHARE sample of respondents aged 50 years or older. In a first step, I matched respondents on age, gender, education, and widowhood. In a second step, I added dementia, the self-rated health of respondents, and the number of limitations in activities of daily living to the set of variables used for matching. The large number of cases in 18 European countries and Israel allowed for excellent matches with very few discarded observations.

This analysis comes with a number of limitations. A little less than 8 percent of institutionalized respondents could not be matched in the Full PSM due to missing values in the variables used as predictors and some additional observations without control cases within the borders of the caliper. In the less elaborate Basic PSM only using socio-demographic variables less than 2 percent of institutionalized dropped out. The drop out of those cases is not random but systematic, as the results show. When comparing response behavior and quality indicators for all institutionalized respondents with the matched institutionalized respondents, an improvement in most indicators can be observed. Hence, the impact of systematic drop-outs can be quantified, but we should be cautious whenever interpreting the degree of differences between the two types of housing and related statistical significance.

As a second limitation, SHARE most probably conducts interviews with a selective subset of elderly respondents. In the first wave of SHARE, the group of institutionalized respondents might be underrepresented due to undercoverage or noncoverage in sampling frames (Bergmann et al. 2019). The survey loses potentially hard-to-interview respondents due to nonresponse in the first wave and panel attrition in the following waves, especially in the oldest age cohort (ibid.). It is difficult to quantify the resulting bias, since external benchmark data would be required. Even though the implementation of proxy interviews is likely to reduce the extent of nonresponse bias, I am certainly missing nonrespondents, potentially those with even larger difficulties in the interview. This applies to institutions *and* private households.

Coming to a conclusion, institutionalized respondents are indeed harder-to-interview than respondents in private households on average. In line with my first hypothesis, they require proxy interviews more often than respondents in private households, they are depicted as less motivated by their interviewers, need additional clarification more often than respondents living in private

households, generate a larger number of missing values in two item batteries, and picked the middle response option more often than respondents living in private households. Those results are mostly in line with the perceptions of survey researchers experienced with research in institutions for the elderly (Schanze and Levinson 2019). The indicators of hard-to-interview response behavior also result in different response patterns and lower data quality, increasing the overall number of missing values and the share of midpoints in Big-Five items.

Nevertheless, three outcomes of this paper should be mentioned to arrive at a more differentiated conclusion: First of all, as suggested by my second hypothesis, the analysis shows that institutionalized respondents are less particular when comparing them to respondents in private households who are more alike in a number of ways. Socio-demographic characteristics can only explain a small fraction of data quality issues. When comparing institutions and private households while controlling for socio-demographic *and* health-related variables, differences in indicators of data quality between the two populations get very small and less often statistically significant. A certain group of residents in private households is nearly as difficult to interview as institutionalized residents.

Secondly, acquiescence seems to be less of a problem in institutions than in private households, even though ability and motivation tend to be lower here than in private households. Apparently institutionalized respondents feel less inclined to present themselves in a more positive light. The variables used in the propensity score matching could not help to explain differences in the degree of acquiescence. Acquiescence led to a blurred factor structure in the Big-Five inventory with marginal differences between the two types of housing. Only in private households the Kaiser criterion would demand the extraction of four instead of five dimensions, while results in institutions are in line with the theory. Those findings, as well as the unexpectedly reverse impact of hard-to-interview indicators, would very much call for further research on this quality indicator.

As a third observation, a generic hard-to-interview label for all institutionalized residents would be seriously misleading. Two out of three institutionalized residents did not require a proxy interview, and every second institutionalized respondent was rated as very willing to answer by the interviewers. The data quality in institutions is comparable to private households with similar basic characteristics. In addition to a lower acquiescence in nursing homes, between 65 to 70 percent of institutionalized residents did not have any missing values in Big-Five or CASP items. A small-scale study in nursing homes in Vienna, Austria, also came to the conclusion that two thirds of all institutionalized residents could be interviewed (Lang 2014), especially when adapting survey instruments to account for aging-related requirements and cognitive impairment (Fox et al. 2007; Lenzner 2012; Sangl et al. 2007; Schanze and Levinson 2019). Once a sample has been drawn and access and cooperation has been secured, it is indeed possible to conduct survey interviews with residents in institutions for the elderly.

References

Agüero-Torres, Hedda, Eva Von Strauss, Matti Viitanen, Bengt Winblad, and Laura Fratiglioni (2001). "Institutionalization in the elderly: The role of chronic diseases and dementia. Cross-sectional and longitudinal data from a population-based study". *Journal of Clinical Epidemiology* 54(8): 795–801.

Alwin, Duane (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Alwin, Duane and Jon Krosnick (1991). "The reliability of survey attitude measurement: The Influence of Question and Respondent Attributes". *Sociological Methods & Research* 20(1): 139–181.

Andrews, Frank and Regula Herzog (1986). "The quality of survey data as related to age of respondent". *Journal of the American Statistical Association* 81(394): 403–410.

Bergmann, Michael, Annette Scherpenzeel, and Axel Börsch-Supan (2019). *SHARE Wave 7 Methodology: Panel innovations and life histories*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Beullens, Koen, Geert Loosveldt, and Caroline Vandenplas (2018). "Interviewer Effects Among Older Respondents in the European Social Survey". *International Journal of Public Opinion Research*: 1–17.

Borrat-Besson, Carmen, Valérie-Anne Ryser, and Judite Gonçalves (2015). "An evaluation of the CASP-12 scale Ageing and Retirement in Europe used in the Survey of Health, (SHARE) to measure Quality of Life among people aged 50+". Lausanne.

Börsch-Supan, Axel (2019a). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5*. SHARE-ERIC. DOI: 10.6103/SHARE.w5.700.

– (2019b). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6*. SHARE-ERIC. DOI: 10.6103/SHARE.w6.700.

– (2019c). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7*. SHARE-ERIC. DOI: 10.6103/SHARE.w7.700.

Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber (2013). "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)". *International Journal of Epidemiology* 42(4): 992–1001.

Caliendo, Marco and Sabine Kopeinig (2005). *Some Practical Guidance for the Implementation of Propensity Score Matching*. Tech. rep. DIW Berlin.

Colsher, Patricia and Robert Wallace (1989). "Data quality and age: Health and psychobehavioral correlates of item nonresponse and inconsistent responses". *Journals of Gerontology* 44(2): 45–52.

Costello, Shane and John Roodenburg (2015). "Acquiescence Response Bias - Yeasaying and Higher Education". *Australian Educational and Developmental Psychologist* 32(2): 105–119.

Dehejia, Rajeev H. and Sadek Wahba (2002). "Propensity Score-Matching Methods for Nonexperimental Causal Studies". *Review of Economics and Statistics* 84(1): 151–161.

Einio, Elina, Christine Guilbault, Pekka Martikainen, and Michel Poulain (2012). "Gender Differences in Care Home Use Among Older Finns and Belgians". *Population* 67(1): 75–101.

Fox, Mary, Souraya Diani, and David Streiner (2007). "Using Standardized Survey Items With Older Adults Hospitalized for Chronic Illness". *Research in nursing & health* 30: 468–481.

Fuchs, Marek (2009). "Item-Nonresponse in einer Befragung von Alten und Hochbetagten - Der Einfluss von Lebensalter und kognitiven Fähigkeiten". In: *Umfrageforschung - Herausforderungen und Grenzen*. Ed. by Martin Weichbold, Johann Bacher, and Christof Wolf: 333–349.

Gaugler, Joseph, Sue Duval, Keith Anderson, and Robert Kane (2007). "Predicting nursing home admission in the U.S: a meta-analysis". *BMC Geriatrics* 7(13): 1–14.

He, Jia, Dave Bartram, Ilke Inceoglu, and Fons J.R. van de Vijver (2014). "Response Styles and Personality Traits: A Multilevel Analysis". *Journal of Cross-Cultural Psychology* 45(7): 1028–1045.

Hyde, M, R D Wiggins, P Higgs, and D B Blane (2003). "A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19)." *Aging Mental Health* 7(3): 186–194. ISSN: 1360-7863. DOI: 10.1080/1360786031000101157. URL: http://www.ncbi.nlm.nih.gov/pubmed/12775399.

Kieruj, Natalia D. and Guy Moors (2013). "Response style behavior: Question format dependent or personal style?" *Quality and Quantity* 47(1): 193–211.

Kirchner, Antje, Kristen Olson, and Jolene Smyth (2017). "Do interviewer postsurvey evaluations of respondents' engagement measure who respondents are or what they do? A behavior coding study". *Public Opinion Quarterly* 81(4): 817–846.

Knäuper, Bärbel (1999). "The Impact of Age and Education on Response Order Effects in Attitude Measurement". *Public Opinion Quarterly* 63: 347–370.

Knäuper, Bärbel, Robert Belli, Daniel Hill, and Regula Herzog (1997). "Question difficulty and respondent cognitive ability: The impact on data quality". *Journal of Official Statistics* 13(2): 181–199.

Knäuper, Bärbel, Kimberly Carrière, Melodie Chamandy, Zhen Xu, Norbert Schwarz, and Natalie Rosen (2016). "How aging affects self-reports". *European Journal of Ageing* 13(2): 185–193.

Knesebeck, Olaf von dem, Martin Hyde, Paul Higgs, Alexandra Kupfer, and Johannes Siegrist (2005). "Quality of life and well-being". In: *Health, ageing and retirement in Europe: First results from the survey of health, ageing and retirement in Europe*. Ed. by Axel Börsch-Supan, Agar Brugiavini, Hendrik Jürges, Johan Mackenbach, Johannes Siegrist, and Guglielmo Weber. Mannheim Research Institute for the Economics of Aging: 199–203.

Koch, Achim and Michael Blohm (2009). "Item Nonresponse in the European Social Survey". *Research & Methods* 18(1): 45–65.

Kroh, Martin, Florin Winter, and Jürgen Schupp (2016). "Using person-fit measures to assess the impact of panel conditioning on reliability". *Public Opinion Quarterly* 80(4): 914–942.

Krosnick, Jon (1991). “Response strategies for coping with the cognitive demands of attitude measures in surveys”. *Applied Cognitive Psychology* 5(3): 213–236.

Kühn, Konstanze and Rolf Porst (1999). *Befragung alter und sehr alter Menschen: Besonderheiten, Schwierigkeiten und methodische Konsequenzen. Ein Literaturbericht*. Tech. rep.: 1–40.

Laferrère, Anne, Aaron Van Den Heede, and Karel Van Den Bosch (2012). “Entry into institutional care: predictors and alternatives”. In: *Active ageing and solidarity between generations in Europe*. Ed. by Axel Börsch-Supan, Martina Brandt, Howard Litwin, and Guglielmo Weber: 253–264.

Lang, Gert (2014). “Zur Befragung und Befragbarkeit von kognitiv eingeschränkten und demenziell veränderten Menschen in Alten- und Pflegeheimen”. In: *Das erzwungene Paradies des Alters?* Ed. by Anton Amann and Franz Kolland. 2nd ed.: 207–229.

Lechner, Clemens and Beatrice Rammstedt (2015). “Cognitive ability, acquiescence, and the structure of personality in a sample of older adults”. *Psychological Assessment* 27(4): 1301–1311.

Lenzner, Timo (2012). “Effects of Survey Question Comprehensibility on Response Quality”. *Field Methods* 24(4): 409–428.

Loosveldt, Geert (1997). “Interaction Characteristics of the Difficult-to-Interview Respondent”. *International Journal of Public Opinion Research* 9(4): 386–394.

Lunt, Mark (2014). “Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching”. *American Journal of Epidemiology* 179(2): 226–235.

Luppa, Melanie, Tobias Luck, Elmar Brähler, Hans Helmut König, and Steffi G. Riedel-Heller (2008). “Prediction of institutionalisation in dementia: A systematic review”. *Dementia and Geriatric Cognitive Disorders* 26(1): 65–78.

Maxwell, Colleen, Andrea Soo, David Hogan, Walter Wodchis, Erin Gilbart, Joseph Amuah, Misha Eliasziw, Brad Hagen, and Laurel Strain (2013). “Predictors of Nursing Home Placement from Assisted Living Settings in Canada”. *Canadian Journal on Aging* 32(04): 333–348.

MEA (2019). “Release Guide 7.0.0”. Munich (DE). URL: www.share-project.org/fileadmin/pdf%7B%5C_%7Ddocumentation/SHARE%7B%5C_%7Drelease%7B%5C_%7Dguide%7B%5C_%7D7-0-0.pdf.

Motel-Klingebiel, Andreas, Daniela Klaus, and Julia Simonson (2014). “Befragung von älteren und alten Menschen”. In: *Handbuch Methoden der empirischen Sozialforschung*. Ed. by Nina Baur and Jörg Blasius. Springer Fachmedien Wiesbaden: 781–786.

Olson, Kristen and Ipek Bilgen (2011). “The role of interviewer experience on acquiescence”. *Public Opinion Quarterly* 75(1): 99–114.

Ongena, Yfke and Wil Dijkstra (2007). “A Model of Cognitive Processes and Conversational Principles in Survey Interview Interaction”. *Applied Cognitive Psychology* 21: 145–163.

Pickery, Jan and Geert Loosveldt (1998). “The impact of respondent and interviewer characteristics on the number of "no opinion" answers: A multilevel model for count data”. *Quality and Quantity* 32(1): 31–45.

Raaijmakers, Quinten A.W., Anne Van Hoof, Harm 'T Hart, Tom F.M.A. Verbogt, and Wilma A.M. Vollebergh (2000). "Adolescents' midpoint responses on Likert-type scale items: Neutral or missing values?" *International Journal of Public Opinion Research* 12(2): 207–215.

Rammstedt, Beatrice, Lewis Goldberg, and Ingwer Borg (2010). "The measurement equivalence of Big-Five factor markers for persons with different levels of education". *Journal of Research in Personality* 44(1): 53–61.

Rammstedt, Beatrice and Christoph Kemper (2011). "Measurement equivalence of the Big Five: Shedding further light on potential causes of the educational bias". *Journal of Research in Personality* 45(1): 121–125.

Revilla, Melanie, Willem Saris, and Jon Krosnick (2014). "Choosing the Number of Categories in Agree-Disagree Scales". *Sociological Methods and Research* 43(1): 73–97.

Roberts, Caroline, Emily Gilbert, Nick Allum, and Léïla Eisner (2019). "Satisficing in Surveys: A Systematic Review of the Literature". *Public Opinion Quarterly* 83(3): 598–626.

Rodgers, Willard, Regula Herzog, and Frank Andrews (1988). "Interviewing Older Adults: Validity of Self-Reports of Satisfaction". *Psychology and aging* 3(3): 264–272.

Salthouse, Timothy (2009). "When does age-related cognitive decline begin?" *Neurobiology of Aging* 30(4): 507–514.

Sangl, Judith et al. (2007). "The Development of a CAHPS® Instrument for Nursing Home Residents (NHCAHPS)". *Journal of Aging & Social Policy* 19(2): 63–82.

Schanze, Jan-Lucas and Ma'ayan Levinson (2019). *Report on what persons live in institutions and the most relevant institution types they live in, the availability of data to select them and their ability to be reached in practice*. Tech. rep. Deliverable 2.17 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. URL: https://seriss.eu/resources/deliverables/.

Schanze, Jan-Lucas and Stefan Zins (2019). "Undercoverage of the elderly institutionalized population: The risk of biased estimates and the potentials of weighting". *Survey Methods: Insights from the Field*. URL: https://surveyinsights.org/?p=10422.

Schober, Michael and Frederick Conrad (1997). "Does conversational interviewing reduce survey measurement error?" *Public Opinion Quarterly* 61(4): 576–602.

Schwarz, Norbert (1999). "Self-reports of Behaviors and Opinions: Cognitive and Communicative Processes". In: *Cognition, Aging, and Self-Reports*. Ed. by Norbert Schwarz, Denise Park, Bärbel Knäuper, and Seymour Sudman. Ann Arbor (USA): Taylor & Francis Group: 17–43.

– (2007). "Cognitive Aspects of Survey Methodology". *Applied Cognitive Psychology* 21: 277–287.

Schwarz, Norbert and Bärbel Knäuper (1999). "Cognition, Aging, and Self-Reports". In: *Cognitive Aging: A Primer*. Ed. by Denise Park and Norbert Schwarz. Psychology Press: 233–253.

Soto, Christopher J., Oliver P. John, Samuel D. Gosling, and Jeff Potter (2008). "The Developmental Psychometrics of Big Five Self-Reports: Acquiescence, Factor Structure, Coherence, and Differentiation From Ages 10 to 20". *Journal of Personality and Social Psychology* 94(4): 718–737.

Sturgis, Patrick, Caroline Roberts, and Patten Smith (2014). “Middle Alternatives Revisited: How the neither/nor Response Acts as a Way of Saying "I Don’t Know"?” *Sociological Methods and Research* 43(1): 15–38.

Sun, Hanyu, Roger Tourangeau, and Stanley Presser (2019). “Panel Effects: Do the Reports of Panel Respondents Get Better or Worse over Time?” *Journal of Survey Statistics and Methodology*(7): 572–588.

Toot, Sandeep, Tom Swinson, Mike Devine, David Challis, and Martin Orrell (2017). “Causes of nursing home placement for older people with dementia: a systematic review and meta-analysis”. *International Psychogeriatrics* 29(02): 195–208.

Tourangeau, Roger (2014). “Defining hard-to-survey populations”. In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau, Brad Edwards, Timothy Johnson, Kirk Wolter, and Nancy Bates. Cambridge University Press: 3–20.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski (2000). *The Psychology of Survey Response*. Cambridge University Press.

Velez, Pauline and Steven D. Ashworth (2007). “The impact of item readability on the endorsement of the midpoint response in surveys”. *Survey Research Methods* 1(2): 69–74.

West, Brady T., Frederick G. Conrad, Frauke Kreuter, and Felicitas Mittereder (2018). “Can conversational interviewing improve survey response quality without increasing interviewer effects?” *Journal of the Royal Statistical Society. Series A: Statistics in Society* 181(1): 181–203.

Appendix

Table 4.10: Average proportion of missing values across 12 CASP items in private households and institutions

		Private households		**Institutions**	
		Mean (%)	*SD*	*Mean (%)*	*SD*
All cases	**No missing**	95.5	0.3	72.1	1.1
	Refusal	0.2	0.0	0.8	0.1
	Don't know	2.6	0.3	17.2	1.0
	Other missing	1.7	0.0	9.9	0.1
	Total N	*77,653*		*2,002*	
Basic PSM	**No missing**	92.5	0.7	72.2	1.0
	Refusal	0.4	0.1	0.8	0.1
	Don't know	4.5	0.6	17.0	1.0
	Other missing	2.6	0.0	9.9	0.1
	N (weighted)	*1,966*		*1,966*	
Full PSM	**No missing**	81.0	0.8	75.4	1.1
	Refusal	0.9	0.1	0.7	0.1
	Don't know	11.3	0.7	15.4	1.0
	Other missing	6.8	0.0	8.6	0.1
	N (weighted)	*1,845*		*1,845*	

Table 4.11: Average proportion of missing values across 11 Big-Five items in private households and institutions

		Private households		**Institutions**	
		Mean (%)	*SD*	*Mean (%)*	*SD*
All cases	**No missing**	96.8	0.1	74.1	0.9
	Refusal	0.2	0.0	0.7	0.1
	Don't know	2.8	0.1	24.4	0.9
	Other missing	0.2	0.0	0.9	0.0
	Total N	*51,527*		*919*	
Basic PSM	**No missing**	94.1	0.2	74.2	0.9
	Refusal	0.5	0.0	0.7	0.1
	Don't know	5.2	0.2	24.2	0.9
	Other missing	0.2	0.0	0.9	0.0
	N (weighted)	*1,184*		*907*	
Full PSM	**No missing**	85.1	0.4	78.2	0.9
	Refusal	0.9	0.1	0.5	0.0
	Don't know	14.0	0.3	21.1	0.9
	Other missing	0.0	0.0	0.2	0.0
	N (weighted)	*972*		*829*	

Table 4.12: Share of Big-Five items answered with strong agreement of all substantive replies (excluding missing values) in wave 7 of SHARE

		Private households		**Institutions**		**Diff.**
		(%)	SD	(%)	SD	
All resp.	**Strongly agree**	19.0	0.17	18.4	0.17	-0.6$^{n.s.}$
	Total N	50,038		703		
Basic PSM	**Strongly agree**	19.3	0.17	18.3	0.17	-1.0$^{n.s.}$
	N (weighted)	1,122		695		
Full PSM	**Strongly agree**	19.4	0.17	18.4	0.17	1.0$^{n.s.}$
	N (weighted)	841		666		

Note: *p <0.05, **p <0.01, ***p <0.001, $^{n.s.}$ not significant; Tests of statistical significance conducted with bivariate OLS regressions

Chapter 5

A reply to “Do surveys accurately report voters over 80 years old?” – Testing for bias in probability-based surveys of private households

This chapter was published as follows: Schanze, Jan-Lucas. 2019. "A reply to “Do surveys accurately report voters over 80 years old?” – Testing for bias in probability-based surveys of private households" In: *French Politics*, Vol. 17 (1), DOI: https://doi.org/10.1057/s41253-019-00080-y

5.1 Introduction

In an article published in French Politics in early 2018, the authors Jean-Yves Dormagen and Laura Michel asked the research question whether older people are misrepresented in survey data that have been collected with a quota method in France (Dormagen and Michel 2018). The authors came up with this research question because they related difficulties in recruiting a part of the elderly population for survey interviews to a bias in samples. Residents with age-related characteristics, such as a worse cognitive performance, a decreasing health status and a lower activity level, might abstain from participation in surveys at higher rates. In addition to the self-selection of respondents, the quota sampling method allows interviewers to select those respondents who are “most-accessible and easy-to-interview” (Dormagen and Michel 2018, 5). According to Dormagen and Michel’s hypotheses, the self-selection and selection bias should reach a maximum impact in the oldest age cohorts. They examine the impact of bias on political behaviour because they assume that it could be especially biased, since younger elderly are highly active, whereas the group of oldest elderly withdraw from the social and political life.

To test their theory, Dormagen and Michel compare the demographic characteristics of survey respondents and self-reported voting behaviour in the *Panel électoral des Français* (PEF) in 2002 and 2007 and the TriElec survey in 2012 with benchmark data coming from the census-based *Échantillon Démographique Permanent* (EDP) and the large-scale administrative *Enquêtes Participation*. The tested election studies used a quota sampling method and crossed age with gender, education, and former occupation to select their respondents.
As results of their analysis, the authors find a demographic bias and a bias in electoral participation. The oldest cohort of respondents older than 85 years is clearly underrepresented in all three surveys, and the surveys strongly overestimate voting in this age cohort. The analysis reveals that the three surveys included too many young elderly (aged between 65 to 69 years), upper class retirees and respondents with a higher education. The bias of self-reported turnout is a direct consequence of the quota sampling and the underlying selection bias. Obviously, interviewers tend to prefer elderly who are easier to reach while attaining quotas that are defined too broad, for instance respondents should be "older than 65 years".
Randomly drawn probability-based surveys do not rely on the interviewers to select a gross sample of respondents. In line with Dormagen and Michel's argument, probability-based surveys should be less biased as a consequence, even though a self-selection of respondents also takes place in those probability-based samples, namely via nonresponse of selected persons. In this article, I compare the results generated with probability-based survey data of three cross-sections of the *European Social Survey* (ESS) in France to the quota samples and administrative benchmark data. Following the replication and the discussion of the results, I focus on the institutionalised population, a group which could be classified as hard-to-survey (Schanze 2017). Like many other surveys, the ESS excludes institutionalised residents from its target population. However, in France, 13.6% of the population older than 80 years live in an institution. After briefly elaborating on the characteristics of this group, I run a simple simulation to illustrate the potential impact the exclusion of institutionalised residents could have on the bias in different age cohorts.

5.2 Theory

As a first research question, this article examines if the bias in the sample composition and the bias in the electoral turnout described in Dormagen and Michel's article arise because of the non-probability sampling method used in the three surveys or whether probability-based surveys have similar issues in the group of elderly residents.
In quota surveys, researchers provide their interviewers with certain strata they need to achieve based on specific variables like age, gender, education or (former) occupation of the "head of the household" (cf. Dormagen and Michel 2018). As a consequence of this approach, the researchers lose control of the selection of respondents within the strata and cannot estimate whether their final sample is biased (Chambers and Skinner 2003). Even if the distribution of strata is close to the benchmark values, there might be an underlying bias in further secondary variables like voting

behaviour (cf. Lohr 2008). In theory, the selection of respondents should be less dependent upon interviewers in probability-based surveys (Levy and Lemeshow 2013). Interviewers do not need to find respondents with certain characteristics but are provided with a batch of households or persons they need to contact. In contrast to quota surveys, interviewers are under no circumstances allowed to substitute selected respondents with other persons (Lohr 2008), even if they share similar characteristics with the originally selected respondents. Random probability-based sampling helps to minimize the selection bias.

However, the Total Survey Error Framework lists several other sources of error that might cause bias in probability-based surveys (Biemer 2010; Groves et al. 2009; Groves and Lyberg 2010). Similar to quota surveys, a self-selection bias can arise in probability-based surveys because respondents might decide to abstain from participation in the survey interview. Nonresponse increases the risk of bias in survey estimates when it occurs systematically (Lynn 2008; Peytchev et al. 2018). Since we can assume that respondents have a differing response propensity based on their characteristics, such as age or health, it is very likely that probability-based surveys are biased as well. Nonetheless, I expect less biased demographic and turnout estimates in probability-based surveys compared to quota surveys due to a weaker selection bias.

Comparing the results obtained in the ESS with the three quota-based samples for the population older than 50 years, it is my hypothesis that the bias in self-reported turnout in the French presidential elections is less distinct in the ESS data. These data do not rely on the quota attainment of interviewers and therefore should offer a better representation of the French elderly population. In a nutshell, the oldest old have a higher, or rather more accurate probability to be selected as respondents in the ESS compared to the quota-based surveys.

In my second hypothesis, I assume that the bias in turnout in the ESS data is still stronger in the oldest age cohorts compared to all the other age cohorts. This is not only a consequence of the greater difficulty to reach and interview older respondents, resulting in a sample of elderly, who are healthier, better educated and less isolated, but could be also a result of the ESS definition of the target population. The ESS does not cover the institutionalised population. In France, 16% of the women older than 80 years and more than 8.5% of the respective men are institutionalised and live in retirement and nursing homes. As a second research question, I ask whether the bias in turnout in the oldest age cohort could be caused by the institutionalised residents, who are missing in the ESS data. I assume that this part of the population has a lower probability to vote than the elderly living in private households. Therefore, it is my hypothesis that their exclusion in the ESS increases the distortion of self-reported voting behaviour compared to the overall number of old aged voters in the *Enquête participation électorale*.

5.3 Data and methods

For testing the influence a different sampling design can have on the data, I analyse cross-sectional data of three rounds of the European Social Survey from 2002, 2008 and 2012 (ESS 2002, 2008,

2012). The ESS defines its target population in France as "[p]ersons aged 15 and over resident within private households, regardless of their nationality, citizenship, language or legal status" (Jowell and CCT 2003). ESS respondents were selected with multi-stage probability-based sampling approaches in France. In 2002 and 2008, a number of communities were sampled as primary sampling units; in those communities a random route procedure was used to draw a sample of households. Within these randomly selected households, the individual respondent was selected with the Last-Birthday-Method (ESS 2008; Jowell and CCT 2003). For the sixth ESS round in 2012, the INSEE sampling frame of the rotating census (Échantillon-maître) was used for a simple random sampling of households within randomly selected communities (ESS 2013). Within the selected households, a Kish grid was used to randomly assign the respondent (ibid.).
In all three ESS rounds analysed for this article, the French respondents were asked whether they voted in the first round of the last presidential election in France.[1] The items on the actual voting behaviour in the presidential elections in 2002, 2007 and 2012 allow a direct comparison of ESS data with the quota samples in the two rounds of the Panel électoral des Français and the TriElec study. For the analysis of voting behaviour, all respondents were excluded who did not have the right to vote due to their age or nationality at the point of data collection or said they were not registered at the voting list.[2] All results presented in this article are weighted with the post-stratification weights provided by the ESS (ESS 2014).
All interviews in the European Social Survey were conducted in the face-to-face mode like in the first wave of the *Panel électoral des Français*. In 2007 and 2012, the PEF and the TriElec survey were administered via the phone in a CATI mode (Dormagen and Michel 2018). It is important to keep in mind the differences in the interview mode, since they limit the comparability of designs between the quota surveys and the ESS. The survey mode has a separate effect on the demographic bias and the bias in turnout, and I cannot draw any conclusions about the isolated impact of the sampling design on the bias.

5.4 Results of the replication with probability-based survey data

In the first part of the empirical analysis, I investigate the degree of demographic bias in both data sources. Figure 5.1, Figure 5.2 and Figure 5.7[3] show the weighted results within the European Social Survey and the Panel électoral des Français in 2002 and 2012 in comparison to the administrative benchmark data. The numbers were calculated by dividing the weighted share of a given age cohort in the ESS by the share in the benchmark data, just as Dormagen and Michel did for the quota surveys (Dormagen and Michel 2018, 10). The dashed line indicates an equivalence of a given share in the survey data with the INSEE benchmark data.

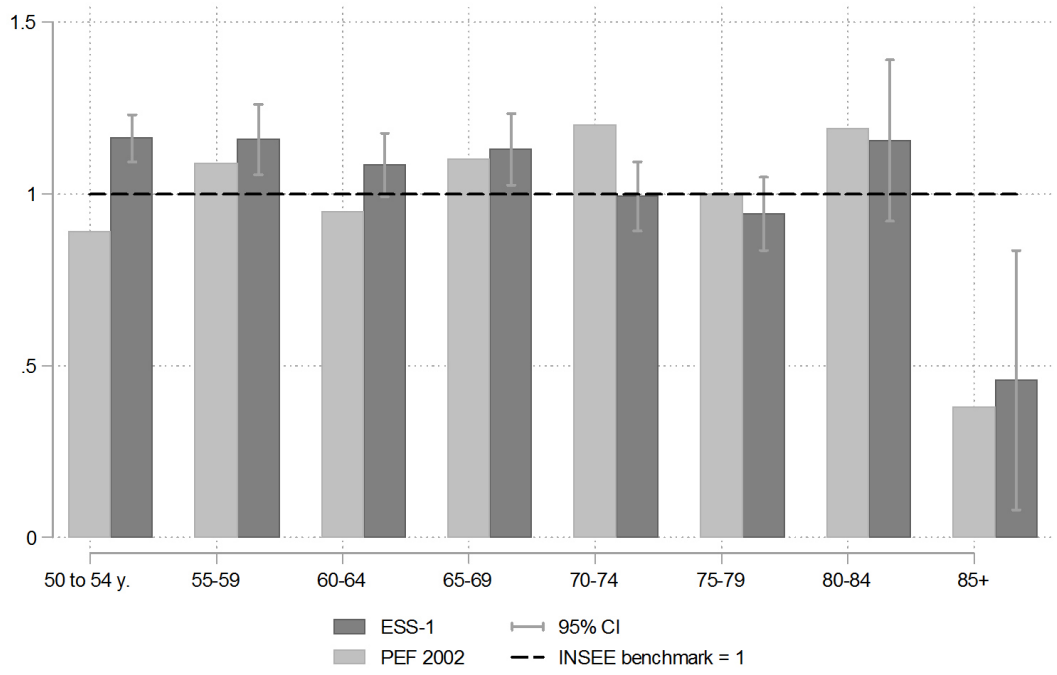
[1] In 2002 and 2012, this question was asked in the same year after the election took place. In 2008, the survey took place one year after the election. This might increase memory effects for the self-reported turnout (DeBell et al. 2018; Voogt 2004).

[2] The latter only affects a small number of 77 respondents in the first round of the ESS.

[3] see Appendix

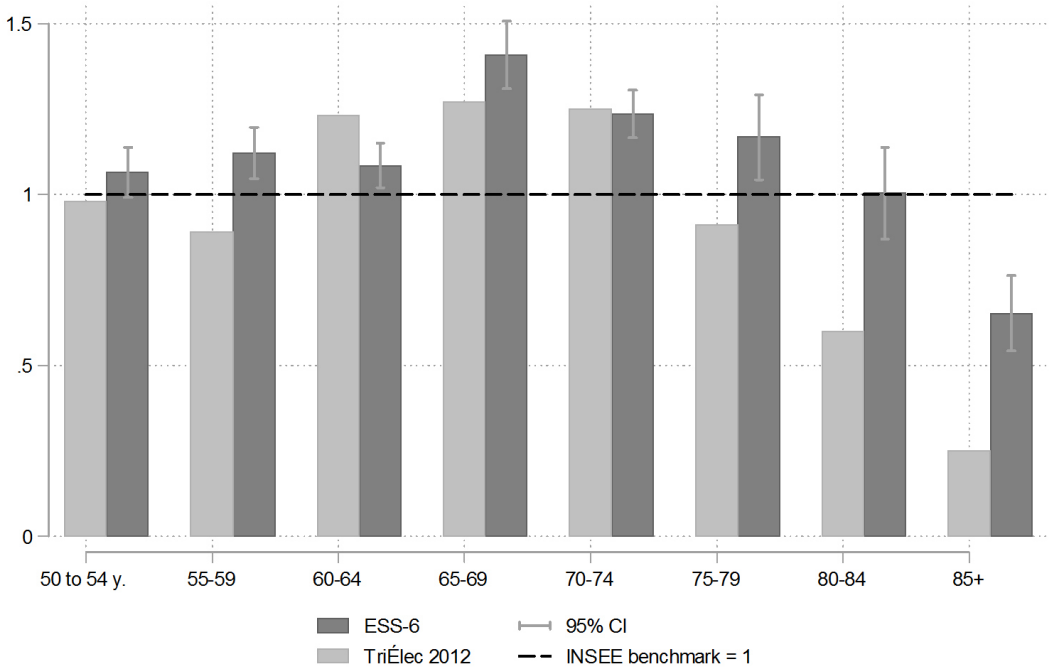
As shown in all three figures, the age cohort older than 85 years is the most critical age group in terms of representation in the surveys. Respondents from this group are underrepresented in all surveys irrespective of the sampling design applied. In probability-based survey data (ESS), the demographic representation of older respondents and the oldest age cohort is closer to the benchmark value than in quota-based surveys (PEF). In the first wave of the ESS in 2002, the difference between the two types of surveys is less distinct than in the two subsequent years. The difference amounts to approximately 11 percentage points in the oldest age cohort in 2002, whereas it grows to 16 percentage points in 2008 and 40 percentage points in 2012. The worse demographic representation in 2007 and 2012 in the PEF and the TriElec survey could also be partly attributed to the CATI mode, which might be a less preferred interview mode for old-aged respondents.

Figure 5.1: Demographic representation of age groups in the population eligible to vote in the PEF and ESS (2002, weighted)



In the 2008 PEF and in the 2012 TriElec survey, the representativeness clearly declines with age, something which cannot be observed in the ESS rounds 1 and 6. In these two ESS rounds, only the demographic representation of the oldest age cohort is critical and even the 95% confidence interval is rather far away from the benchmark value. In the 2008 ESS, however, the declining pattern of representation of the age cohorts older than 80 years bears a strong resemblance to the PEF (see Figure 5.7). Moreover, in the ESS rounds 1 and 4 the confidence intervals for the two oldest age cohorts are much larger due to the smaller number of respondents in this age group compared to the sixth round.

Figure 5.2: Demographic representation of age groups in the population eligible to vote in the TriElec Study and the ESS (2012, weighted)



It is obvious that post-stratification weights cannot fully correct for the bias in the oldest population, even when analysing age as dependent variable. In the European Social Survey, this can be explained with the construction of the post-stratification weights. The ESS calculates the weight with a very broad age group of respondents older than 55 years (ESS 2014). This weight cannot fully account for the heterogeneity of response rates in the older age cohorts as depicted in Figures 5.1, 5.2 and 5.7. Compared to INSEE benchmark data, respondents aged between 55 to 75 years are overrepresented in all unweighted ESS samples, whereas the oldest age cohort is underrepresented in the unweighted samples.[4] This underrepresentation of older respondents increases when post-stratification weights are applied.

As the second socio-demographic variable, education is very meaningful to assess bias. Dormagen and Michel concluded that respondents with a high school degree or an even higher level of education were overrepresented in the 2012 quota sample, and this overrepresentation increased with age (2018). Table 5.1 shows the results of an equivalent analysis with ESS-6 data. In the cases of lower and higher education all ESS estimates in the various age cohorts are closer to the benchmark value retrieved from the INSEE data.[5] The TriElec survey was only better in some of the estimates on the number of people with a high school diploma. The overall lesson from

[4]The unweighted ESS results are not shown in this article

[5]A value of 1 would indicate equivalence of shares in the survey data with the INSEE benchmark data.

the comparison with respect to level of education supports the assumption that probability-based survey data have a better representation than quota samples.

Table 5.1: Rates of overrepresentation compared to INSEE benchmark data by age groups and levels of education - comparison of TriElec (Dormagen and Michel 2018, 14) with ESS data (weighted)

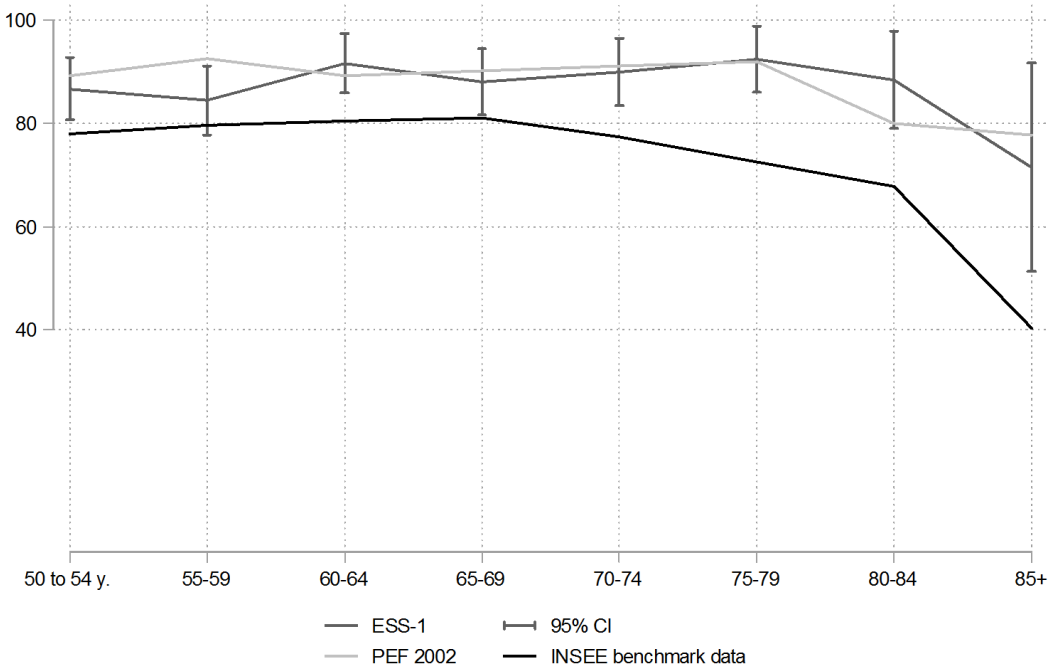
Age cohorts	Rate without high school diploma		Rate with high school diploma		Rate above high school diploma	
	TriElec	**ESS-6**	**TriElec**	**ESS-6**	**TriElec**	**ESS-6**
50 to 54 years	0.89	1.09	1.04	0.72	1.24	0.96
55 to 59 years	0.77	1.07	1.22	0.79	1.53	0.92
60 to 64 years	0.81	0.93	1.10	1.25	1.54	1.06
65 to 69 years	0.79	1.16	1.17	0.75	1.64	0.58
70 to 74 years	0.74	1.05	1.31	0.72	2.47	0.92
75 to 79 years	0.70	1.01	1.95	0.77	2.94	1.18
80 to 84 years	0.76	0.99	1.91	0.55	3.10	1.69
Older than 85 years	0.73	1.09	0.99	0.44	4.41	0.59
Total (>= 50 years)	0.76	1.04	1.27	0.83	1.81	0.95

In the second part of the analysis, I compare the self-reported turnout in the six survey samples (see Figure 5.3, Figure 5.4 and Figure 5.8). A very large majority of the French population aged between 50 to 70 years voted in the presidential elections according to the benchmark data. From 70 years onwards, the turnout constantly declines and in the oldest age cohorts only 40-60% of the population actually votes (INSEE 2012).

As expected, all surveys overestimated turnout in most age cohorts. In 2002, the ESS overestimated the turnout by about 5-11 percentage points in the age cohorts between 50 to 70 years. Figure 5.3 shows how the turnout gap gets constantly larger in the older age cohorts, although both surveys seem to reflect the dropping turnout rates among older persons. For both surveys the turnout gap reaches its maximum in the French population aged 85 years or older with a width of 31.2 percentage points in the ESS and 37.5 percentage points in the PEF.

In 2008 and 2012, the ESS estimates of turnout were closer to the benchmark data than the PEF and the TriElec survey (Figure 5.4 and Figure 5.8). This especially holds true for the younger age cohorts. Most of the ESS estimates are very close to the true value for the population younger than 69 years, and the ESS appears to suffer less from bias in these age cohorts than the two quota surveys. It should be mentioned that the high turnout in these two presidential elections also reduces the issue of overestimation. Similar to 2002, the turnout gap opens for the older age cohorts. In 2008, the PEF and the ESS yielded similar results for the oldest age cohorts, whereas in 2012 the TriElec survey strongly overestimated turnout in the population older than 80 years, estimating a *full* participation in these age cohorts. The European Social Survey was significantly closer to the true value in this round (see Figure 5.4).

Figure 5.3: Estimated turnout in the PEF and in the ESS in different age cohorts compared to true turnout (2002, weighted)

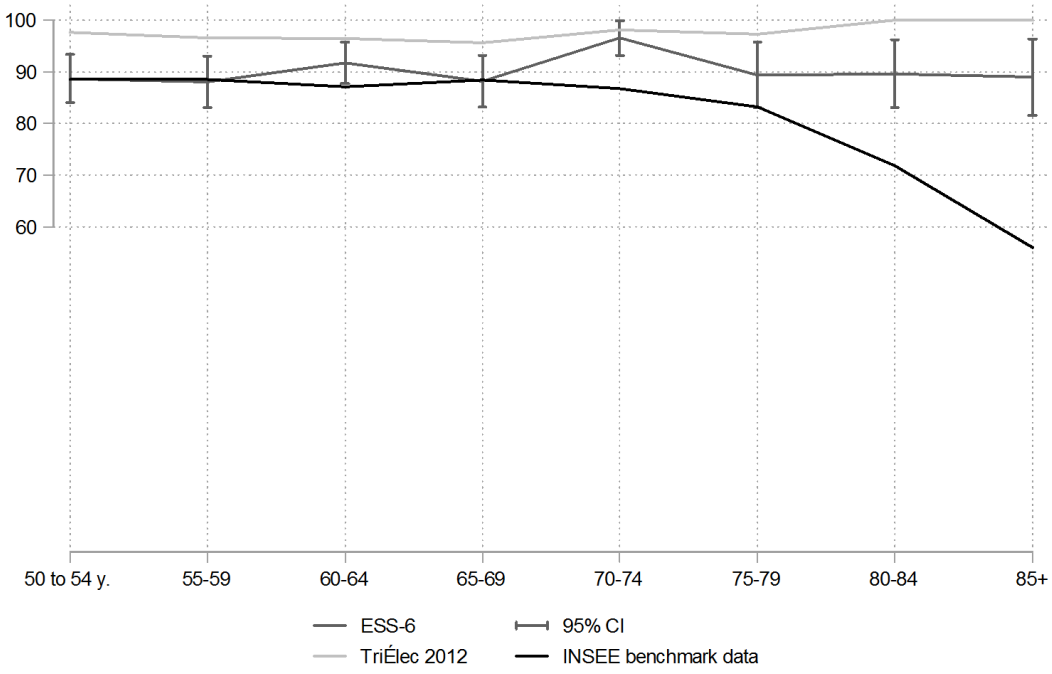


Summing it up, both surveys suffer from demographic bias and bias in turnout. As reported by Dormagen and Michel, the biases for age and turnout reach their maximum in the oldest age cohorts, irrespective of the underlying sampling principles. Older age segments of the population are worse represented than younger age cohorts. Both examined surveys overestimated the turnout, not only because of a bias in the sample due to the selection and self-selection of respondents (Sciarini and Goldberg 2016), but also due to social desirability and the perception of respondents that voting constitutes a norm (Bernstein et al. 2001; Voogt 2004), which could be positively influenced by the telephone mode used in the second PEF and the TriElec survey (cf. DeBell et al. 2018).

Comparing the two different types of sampling approaches, probability-based surveys perform better in terms of demographic representation and the estimation of turnout. Especially in 2012, when a sampling frame was used by the French ESS team instead of the random route procedure, bias in age only occurs in the oldest age cohort and is less extreme than in the TriElec survey. The probability-based survey data have a better representation of all levels of education. In contrast to the results in the TriElec survey, the pattern of an underestimation of lower educated respondents and a clear overestimation of highly educated respondents cannot be observed in the ESS data.

The turnout gap still occurred for respondents older than 75 years but affected the ESS data in a less extreme way. The better performance of probability-based data is in line with previous literature and should be – in addition to the face-to-face interview mode – due to the fact that

Figure 5.4: Estimated turnout in the PEF and in the ESS in different age cohorts compared to true turnout (2012, weighted)



the ESS does not rely on the selection of respondents by their interviewers. However, the self-selection bias caused by nonresponse and other sources of error named in the Total Survey Error Framework still poses a threat to the representativeness among older persons.

5.5 Missing institutionalised residents as a reason for bias?

In their study, Dormagen and Michel assume that the oldest age cohorts are more difficult to access for interviewers because "they are more often sick, more likely to have been hospitalized or to live in special institutions" (Dormagen and Michel 2018, 14). In the following analysis, I want to elaborate on the part of the population that lives in institutions, such as retirement homes or nursing homes.

Bias is the consequence of two factors: the size of an excluded population and its statistical distinctiveness compared to those respondents covered by the survey (Groves et al. 2009; Lessler and Kalsbeek 1992). In France, more than 2% of the total population lived in collective households according to the French census in 2009 and 2011 (Eurostat 2016; Pirou et al. 2013). The proportion of institutionalised residents increases with age. While only 1.2% of the age cohort between 60 to 69 years lives in institutions, 2.6% of the French population aged between 70 to 79 years is institutionalised. In the oldest age cohorts, older than 80 years, more than 13% of the French population was institutionalised in 2011 (Eurostat 2016). The mean age of the population living

in retirement homes was 84 years in 2009 (Pirou et al. 2013).

Regarding the statistical distinctiveness, a number of gerontological studies examined characteristics of the elderly institutionalised population and factors that lead to an institutionalisation in old age cohorts worldwide. Analyses of survey data as well as meta-analyses and systematic reviews have identified demographic, socio-economic and health-related variables, which increase the risk of institutionalisation among older persons. Cognitive impairments and dementia, medical conditions, limitations in activities of daily living and physical dependency increase the demand for institutional care (Désesquelles and Brouard 2003; Laferrère et al. 2012; Luppa et al. 2010; Rodríguez-Sánchez et al. 2017; Toot et al. 2017). Additionally, an older age, not being married, living alone and not having children, a smaller social network, a lower income and not being a homeowner increase the probability of institutionalisation in older age among other factors (Angelini and Laferrère 2012; Désesquelles and Brouard 2003; Laferrère et al. 2012; Luppa et al. 2010; McCann et al. 2012; Rodríguez-Sánchez et al. 2017).[6] It is easy to imagine that these factors have an impact on further variables like the political behaviour and the participation in elections (cf. Peytchev et al. 2018).

In fact, it could be the case that residents living in institutions for the elderly are covered by the two Panel électoral des Français and the TriElec survey, since their definitions of the target populations include the French population eligible to vote, which means they do *not* explicitly exclude institutionalised residents (Cautrès et al. 2003; l'Intérieure 2007). Nonetheless, I assume they are most probably underrepresented in the samples, because they are more difficult-to-reach – e.g., interviewers need to bypass gatekeepers - and less easy-to-interview very often.

In contrast, the European Social Survey explicitly excludes institutionalised residents in their target population like many other social surveys (Jowell and CCT 2003; Schanze 2017). Whenever turnout is estimated on the basis of ESS data the results only apply to the target population living in *private households*. However, in the administrative benchmark data used by Dormagen and Michel as well as by this article institutionalised residents are *included*. The exclusion of 13.5% of the population definitely contributes to the underestimation of the oldest age cohort in the ESS samples. However, the simulation focuses on the voting behaviour. Institutionalised residents do have the right to vote in France as long as they are registered at the voting lists.[7] The mismatch between the target population in the ESS and the population eligible to vote might be one of the factors leading to bias in survey samples.

Due to the usual exclusion of institutionalised residents in many surveys, information about their attitudes and behaviour is usually missing. To our knowledge, this is also the case for their political behaviour. Thus, to examine the potential impact the inclusion of institutionalised residents could have on the bias in turnout, I used a simple simulation. In a first step, I added synthetic or hy-

[6] In France, the Enquête Handicaps-Incapacités-Dépendance (HID) and the Enquête auprès des établissements d'hébergement pour personnes âgées (EHPA) allow analyses of the institutionalised population (Désesquelles and Brouard 2003; Wolff 2013).

[7] Please note, in the simulation I assume that all French institutionalised residents are registered to vote. This assumption does probably not hold for all institutionalised residents, especially if they suffer from dementia. According to INSEE data the oldest age cohorts show the highest rate of registration at the voting lists with 95% for the population older than 50 years old (INSEE 2012).

pothetic institutionalised respondents to the ESS samples. The number of synthetic cases follows the share of institutionalised residents in three different age cohorts provided by the 2011 census in France. I assume a similar extent of self-selection of respondents in institutions and in private households, which is why I let the number of synthetic cases follow the number of community-dwelling respondents in the ESS. Table 5.2 provides an overview of how many synthetic cases were added in the three ESS rounds within the three age cohorts.

Table 5.2: Adding synthetic institutionalised respondents to the ESS samples (based on census data)

Age cohorts	**Share of institutionalised respondents (2011 census)**	Additional synthetic institutionalised respondents (in brackets: original number of ESS respondents)		
		ESS-2 (2002)	**ESS-4 (2008)**	**ESS-6 (2012)**
60 to 69 years	1.23%	2.4 (192)	3.5 (279)	4.1 (328)
70 to 79 years	2.60%	4.0 (151)	5.8 (216)	5.8 (219)
> 80 years	13.56%	9.6 (61)	16.5 (105)	23.3 (149)

As the second step of the simulation, I need to make assumptions about the potential turnout of the additional synthetic respondents. On the basis of the demographic and health-related characteristics of institutionalised residents described above, I assume that institutionalised residents vote with a lower probability than their peer group living in private households.[8] I calculated the bias for three different scenarios. In the first scenario, I calculated the bias with a 5-percentage point lower turnout among the synthetic institutionalised respondents compared to their peers living in private households. For instance, I assumed that 80.5% of the 9.6 synthetic respondents older than 80 years would vote in this scenario compared to 85.5% of the community-dwelling respondents in the first round of the ESS. The second scenario assumes a 25-percentage point lower turnout among synthetic respondents, and the third scenario assumes a turnout of 50-percentage points lower (35.5% in the example mentioned above).
The simulation aims to illustrate the potential impact an inclusion of institutionalised residents could have in ESS samples. It could be seen rather as a maximum impact, since the model assumes a complete participation of institutionalised residents in the survey without any nonresponse of this group.

5.6 Results

Tables 5.3, 5.4 and 5.5[9] indicate the changes in the turnout in the different scenarios and also provide the deviation in brackets. These deviations are depicted in Figures 5.5, 5.9 and 5.10[10] for the three different age groups and the three different scenarios that only differ with respect to the

[8] A similar turnout would make no difference in the results compared to the original ESS samples with community-dwelling respondents only.
[9] The tables for ESS-1 and ESS-4 can be found in the Appendix.
[10] see Appendix

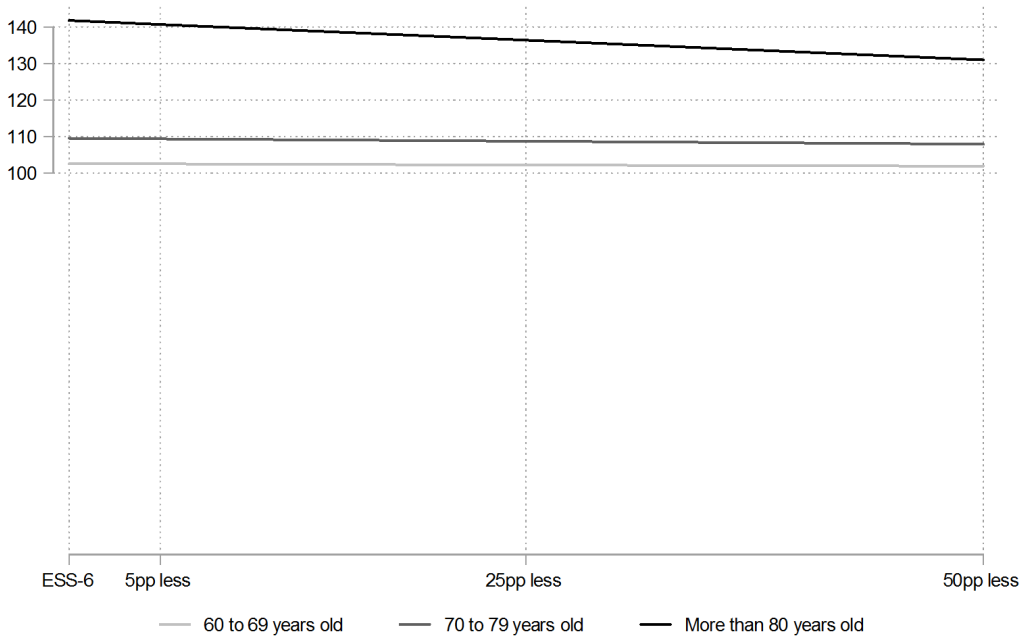
assumed turnout among the synthetic institutionalised respondents.

Table 5.3: Changes in turnout estimates and deviation (in brackets) when synthetic institutionalised respondents with given turnout are added to the ESS samples (2012)

Age cohorts	INSEE 2012	ESS-6		5%-points less		25%-points less		50%-points less	
60 to 69 years	87.7%	90.0%	(1.03)	89.9%	(1.03)	89.7%	(1.02)	89.4%	(1.02)
70 to 79 years	85.0%	93.2%	(1.10)	93.1%	(1.09)	92.5%	(1.09)	91.9%	(1.08)
> 80 years	63.0%	89.3%	(1.42)	88.7%	(1.41)	86.0%	(1.36)	82.6%	(1.31)

The numbers reveal the expected impact of an inclusion of institutionalised residents. The reductions in bias for the two younger age cohorts are rather small, but including institutionalised residents could make a significant difference for the oldest age cohort. For the sixth wave of the European Social Survey, the original turnout of 89.3% for the respondents older than 80 years would shrink to 82.6% if 23 additional institutionalised respondents were part of the survey sample and 39.8% of those additional respondents voted. This would account for a reduction in the deviation of nearly 11 percentage points in the ESS-6 sample. The results for the previous two rounds of the ESS show a similar influence of synthetic institutionalised residents.[11]

Figure 5.5: Reduction of bias in turnout when synthetic institutionalised residents with given turnout are added to the ESS-6 sample



[11] see Appendix

As a main conclusion from this simple simulation, an inclusion of institutionalised residents will certainly not counterbalance other sources of bias in the oldest age cohorts, like a stronger non-response. It could indeed help to reduce the bias when survey samples of community-dwelling residents are compared with large-scale administrative data. In France and most other European countries, the relative size of institutionalized residents in the oldest age cohorts might demand their inclusion, especially for those variables for which they are likely to deviate from respondents living in private households. Apart from variables related to the institutionalisations as such (see previous section), voting behaviour could be affected.

5.7 Conclusion

This article repeated an empirical analysis of demographic bias and bias in turnout in three French election surveys (Dormagen and Michel 2018). My analysis corroborates previous results of Dormagen and Michel, who concluded that the oldest age cohorts are a critical group in survey research. Survey samples include too few old-aged respondents, and the respondents they include are less representative for their age cohort than respondents in other age cohorts. As a result, the bias in electoral turnout reaches its maximum in this age cohort compared to younger age cohorts. The present article compared the estimates of probability samples in the European Social Survey with the quota samples in three French election surveys. Following the theory, the probability samples should not suffer from selection bias in contrast to the nonprobability samples. Indeed, the analyses revealed that probability-based surveys achieved results closer to the "true values" for age, education and turnout as measured in large-scale benchmark data than quota samples, probably also because they were conducted in a face-to-face interview mode. However, even the probability-based surveys cannot fully eliminate the demographic bias and also suffer from distortions in the voting behaviour. This could be the consequence of a stronger unit nonresponse among older respondents or due to other factors leading to a larger turnout gap. The exclusion of the institutionalised population could be one of the reasons for biased survey results.

Since institutionalised residents are a substantial group within the elderly French population, their exclusion might lead to an overestimation in reported turnout. This is especially the case when we assume a less active voting behaviour of institutionalised residents on the basis of their demographic and health-related characteristics. Institutionalised residents are older, frailer and less healthy.

To test the potential impact of institutionalised residents, I added synthetic respondents to ESS samples and simulated different degrees of statistical distinctiveness of voting behaviour because this type of information is missing in France. The simulation shows that an inclusion of institutionalised residents could make a significant difference in survey estimates, especially in the oldest age cohorts older than 80 years old. In younger age cohorts, the share of the population living in institutions is smaller and they only have a very small, probably statistically insignificant impact.

This article confirmed previous methodological literature underlining the superiority of probability-

based sampling methods in order to achieve less biased samples or respondents. Moreover, it highlighted the oldest age cohorts as a critical group for survey research in general. For the sake of drawing more valid conclusions for this group of respondents, it might be necessary to extent the definition of the target population to the institutionalised population. Otherwise, it would be necessary to emphasize that survey estimates only apply to the population living in private households and might consequently deviate from results observed in the general population.

References

Angelini, Viola and Anne Laferrère (2012). "Residential mobility of the European elderly". *CESifo Economic Studies* 58(3): 544–569.

Bernstein, Robert, Anita Chadha, and Roberg Montjoy (2001). "Overreporting voting: why it happens and why it matters." *Public opinion quarterly* 65(September 1998): 22–44. DOI: 10.1086/320036.

Biemer, Paul P. (2010). "Total survey error: Design, implementation, and evaluation". *Public Opinion Quarterly* 74(5): 817–848.

Cautrès, Bruno, Nonna Mayer, and Annie Claude Salomon (2003). "The French Electoral Panel 2002". *French Politics* 1(2): 243–245.

Chambers, R. and C. Skinner (2003). *Analysis of Survey Data*. John Wiley & Sons, Ltd.

DeBell, Matthew, Jon A. Krosnick, Katie Gera, David S. Yeager, and Michael P. McDonald (2018). "The Turnout Gap in Surveys". *Sociological Methods & Research*: 1–30.

Désesquelles, Aline and Nicolas Brouard (2003). "The Family Networks of People aged 60 and over Living at Home or in an Institution". *Population* 58(2): 201–227.

Dormagen, Jean-Yves and Laura Michel (2018). "Do surveys accurately report voters over 80 years old?" *French Politics* 16(2): 195–212.

– (2019). "Probability-based versus quota-based surveys? Selection and self-selection biases persist among the oldest old". *French Politics* 17(1): 45–49.

ESS (2002). *European Social Survey Round 1: Data file edition 6.5*. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

– (2008). *European Social Survey Round 4: Data file edition 4.4*. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

– (2012). *European Social Survey Round 6: Data file edition 2.3*. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

– (2013). "ESS6 - 2012 Documentation Report".

– (2014). "Documentation of ESS Post-Stratification Weights". URL: https://europeansocialsurvey.org/docs/methodology/ESS1-5%7B%5C_%7Dpost%7B%5C_%7Dstratification%7B%5C_%7Dweights%7B%5C_%7Ddocumentation.pdf.

Eurostat (2016). *European Statistical System - Census Hub*. URL: https://ec.europa.eu/CensusHub2.

Groves, Robert, Floyd Fowler, Mick Couper, James Lepkowski, Eleanor Singer, and Roger Tourangeau (2009). *Survey Methodology*. Second edi. New York: John Wiley & Sons, Inc.

Groves, Robert and Lars Lyberg (2010). "Total survey error: Past, present, and future". *Public Opinion Quarterly* 74(5): 849–879.

INSEE (2012). *L'inscription et la participation électorales en 2012: Qui est inscrit et qui vote*. URL: https://www.insee.fr/fr/statistiques/1281060.

Jowell, Roger and CCT (2003). *European Social Survey 2002/2003: Technical Report*. Tech. rep. London: Centre for Comparative Social Surveys, City University: 341–373.

Laferrère, Anne, Aaron Van Den Heede, and Karel Van Den Bosch (2012). "Entry into institutional care: predictors and alternatives". In: *Active ageing and solidarity between generations in Europe*. Ed. by Axel Börsch-Supan, Martina Brandt, Howard Litwin, and Guglielmo Weber: 253–264.

Lessler, Judith T. and William D. Kalsbeek (1992). *Nonsampling Error in Surveys*. New York: John Wiley & Sons, Inc.

Levy, Paul and Stanley Lemeshow (2013). *Sampling of Populations: Methods and Applications*. 4th. John Wiley & Sons, Inc.

l'Intérieure, Ministère de (2007). "Enquête post électorale présidentielle 2007: Résultats d'ensemble".

Lohr, Sharon (2008). "Coverage and Sampling". In: *International Handbook of Survey Methodology*. Ed. by Edith de Leeuw, Joop Hox, and Don Dillman. European Association of Methodology: 97–112.

– (2010). *Sampling: Design and Analysis*. Boston (MA), USA: Brooks/Cole.

Luppa, Melanie, Tobias Luck, Herbert Matschinger, Hans-Helmut König, and Steffi Riedel-Heller (2010). "Predictors of nursing home admission of individuals without a dementia diagnosis before admission - results from the Leipzig Longitudinal Study of the Aged (LEILA 75+)." *BMC health services research* 10: 186.

Lynn, Peter (2008). "The Problem of Nonresponse". In: *International Handbook of Survey Methodology*. Ed. by Edith de Leeuw, Joop Hox, and Don Dillman. European Association of Methodology: 35–55.

McCann, Mark, Emily Grundy, and Dermot O'Reilly (2012). "Why is housing tenure associated with a lower risk of admission to a nursing or residential home? Wealth, health and the incentive to keep 'my home'". *Journal of Epidemiology & Community Health* 66(2): 166–169.

Peytchev, Andy, Stanley Presser, and Mengmeng Zhang (2018). "Improving Traditional Nonresponse Bias Adjustments: Combining Statistical Properties with Social Theory". *Journal of Survey Statistics and Methodology*: 1–25.

Pirou, David, Nadine Poullain, and Sandrine Rochelle (2013). "La vie en communauté". *INSEE Première*(1434): 6–9.

Rodríguez-Sánchez, Beatriz, Viola Angelini, Talitha Feenstra, and Rob Alessie (2017). "Diabetes-Associated Factors as Predictors of Nursing Home Admission and Costs in the Elderly Across Europe". *Journal of the American Medical Directors Association* 18(1): 74–82.

Schanze, Jan-Lucas (2017). *Report on sampling practices for the institutionalized population in social*. Tech. rep. Deliverable 2.16 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. URL: https://seriss.eu/resources/deliverables/.

Sciarini, Pascal and Andreas C. Goldberg (2016). "Turnout bias in postelection surveys: Political involvement, survey participation, and vote overreporting". *Journal of Survey Statistics and Methodology* 4(1): 110–137. ISSN: 23250992. DOI: 10.1093/jssam/smv039.

Toot, Sandeep, Tom Swinson, Mike Devine, David Challis, and Martin Orrell (2017). "Causes of nursing home placement for older people with dementia: a systematic review and meta-analysis". *International Psychogeriatrics* 29(02): 195–208.

Voogt, Robert (2004). "I'm not interested - Nonresponse bias, response bias and stimulus effects in election research". PhD thesis. Universiteit van Amsterdam.

Wolff, François-Charles (2013). "Well-Being of Elderly People Living in Nursing Homes: The Benefits of Making Friends". *Kyklos* 66(1): 153–171.

Appendix

Figure 5.6: Share of residents living in collective households within respective age and gender cohorts in France (2011; Eurostat 2016)

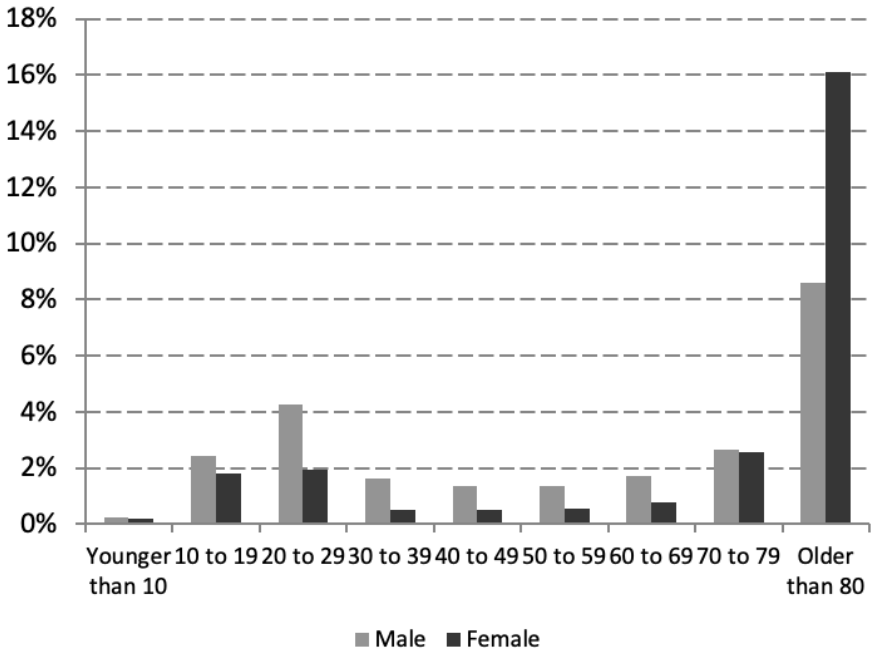


Figure 5.7: Demographic representation of age groups in the population eligible to vote in the PEF and ESS (2008)

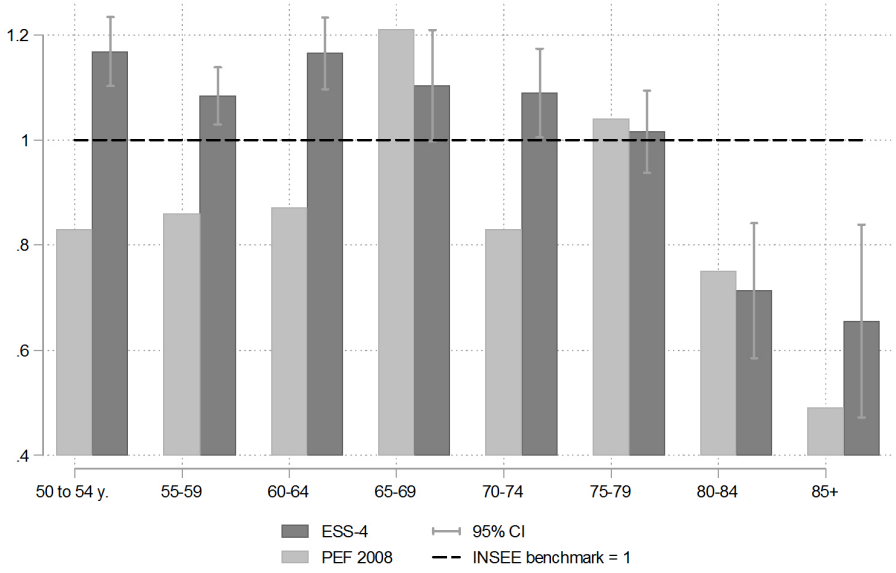


Figure 5.8: Estimated turnout in the PEF and in the ESS in different age cohorts compared to true turnout (2008, weighted)

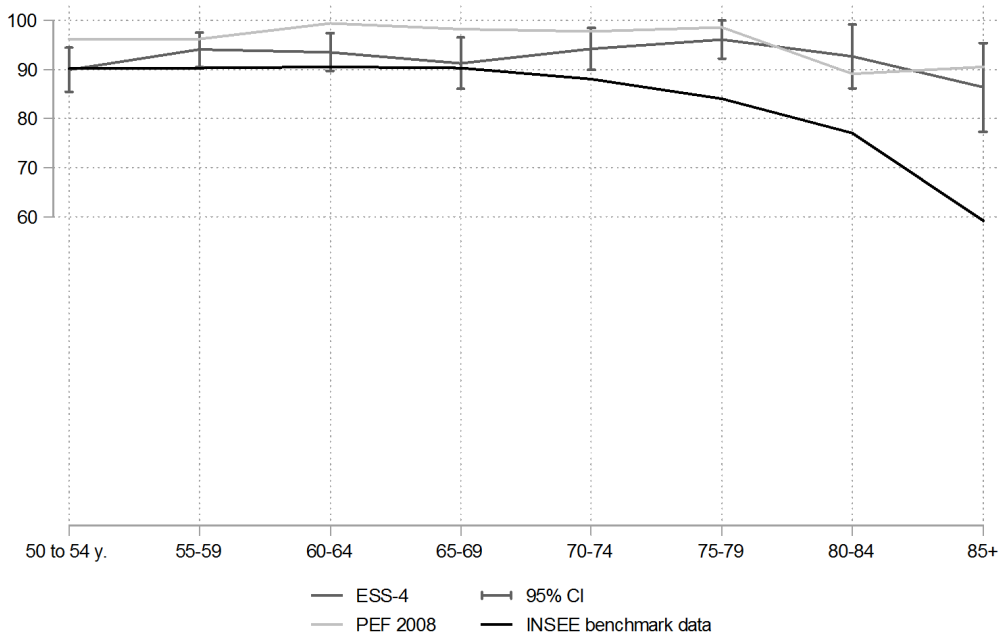


Figure 5.9: Reduction of bias in turnout when synthetic institutionalised residents with given turnout are added to the ESS-1 sample

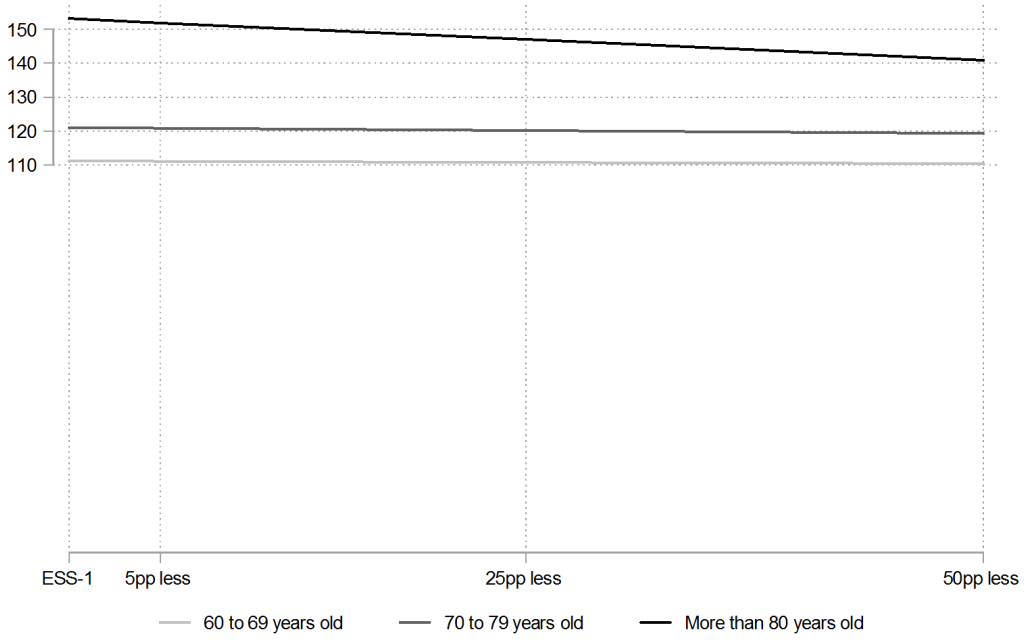


Figure 5.10: Reduction of bias in turnout when synthetic institutionalised residents with given turnout are added to the ESS-4 sample

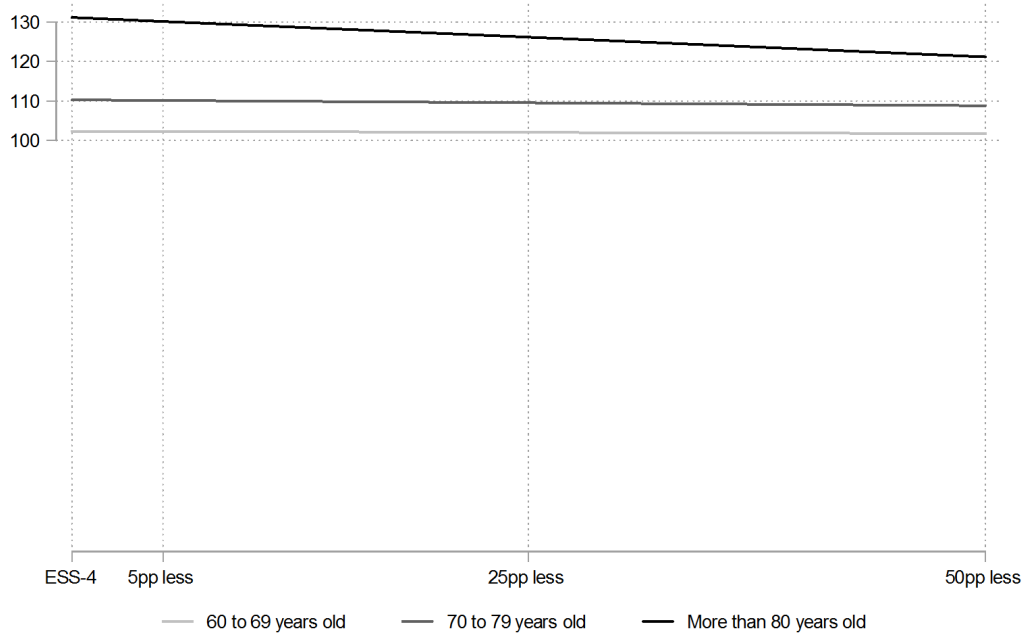


Table 5.4: Changes in turnout estimates and deviation (in brackets) when synthetic institutionalised respondents with given turnout are added to the ESS samples (2002)

Age cohorts	**INSEE 2002**	**ESS-1**		**5%-points less**		**25%-points less**		**50%-points less**	
60 to 69 years	80.8%	89.8%	(1.11)	89.8%	(1.11)	89.5%	(1.11)	89.2%	(1.20)
70 to 79 years	75.2%	91.0%	(1.21)	90.9%	(1.21)	90.4%	(1.20)	89.7%	(1.19)
> 80 years	54.9%	84.0%	(1.53)	83.3%	(1.52)	80.6%	(1.47)	77.2%	(1.41)

Table 5.5: Changes in turnout estimates and deviation (in brackets) when synthetic institutionalised respondents with given turnout are added to the ESS samples (2008)

Age cohorts	**INSEE 2008**	**ESS-4**		**5%-points less**		**25%-points less**		**50%-points less**	
60 to 69 years	90.4%	92.6%	(1.03)	92.6%	(1.02)	92.3%	(1.02)	92.0%	(1.02)
70 to 79 years	86.2%	95.1%	(1.10)	95.0%	(1.10)	94.4%	(1.10)	93.8%	(1.09)
> 80 years	68.5%	89.8%	(1.31)	89.1%	(1.30)	86.4%	(1.26)	83.0%	(1.21)

5.8 Addendum

Chapter 5 was a reply to a paper written by Dormagen and Michel (2018). This reply and extension led to another reply by the authors of the original paper (Dormagen and Michel 2019). In their second research note, Dormagen and Michel acknowledge the better representation of the oldest age group in probability-based samples when comparing them to quota samples, especially for the 2008 and 2012 European Social Survey (ESS). However, they interpret the added value of probability-based samples as rather small, also justified by a similar bias in estimates of electoral turnout in the oldest age group in ESS samples. Hence, Dormagen and Michel come to the conclusion that "*ESS probability-based surveys are just as ill-suited as PEF-type quota-based surveys to study the impact of aging on electoral participation*" (Dormagen and Michel 2019, 48). The representation bias in quota surveys could be solved by adding a quota for the population older than 85 years (ibid.).

Should social surveys in France and elsewhere use quota sampling instead of relying on probability-based sampling designs? Probably not, for a number of reasons. First, in their reply, the authors restrict their analysis of results to the surveys covering the 2002 and 2008 elections, because the quota survey TriÉlec was conducted with telephone interviews in 2012 (Dormagen and Michel 2019). I agree that differences between this survey and the 2012 ESS are likely to be caused by different sampling approaches *and* by different survey modes (cf. Section 5.3, DeBell et al. 2018). However, the 2012 TriÉlec survey shows even stronger issues with anticipating the turnout of the oldest age groups, while the ESS performs significantly better with its probability-based sample and the face-to-face mode. It is also important to mention that ESS estimates of the turnout of younger age groups between 50 to 70 years are closer to reality than the quota surveys in 2008 and 2012. Interviewers in quota surveys are able to select the most accessible respondents (cf. Lohr 2010, 97) in any age group, while hard-to-reach respondents have an lower probability to be selected under this design.

As another methodological limitation of their reply, Dormagen and Michel (2019, 47) compare unweighted estimates from the ESS and the quota surveys. While selection probabilities are unknown in quota surveys, one requirement and advantage of probability-based samples consists in knowing the exact inclusion probabilities of single sample units (Lohr 2010). This enables researchers to calculate design weights to control for unequal inclusion probabilities in the case of stratified or clustered samples. In all three ESS rounds in France, a sample design with three different stages was used, starting with a selection of regions as primary sampling units (PSU), followed by a selection of households within those PSU, and finally the selection of individual respondents within those households. In this case, only results weighted with design weights are proper probability-based estimates.

Summarized, it is certainly true that probability-based ESS samples also have a greater trouble to cover the oldest age groups adequately in contrast to younger age groups. This is due to a stronger nonresponse in this segment of the population, apparently also leading to a stronger bias in estimates of turnout. Compared to quota samples, a better demographic representation in the oldest

age group only has a small marginal effect on the bias in turnout estimates. Probability-based samples simply get "more of the same", more active, more healthy respondents who are also more likely to participate in presidential elections. A quota for the population older than 85 years as suggested by Dormagen and Michel would definitely help to improve the demographic representation in a quota sample, but interviewers would most probably fail in the attempt to improve secondary characteristics of their sample, like the participation in elections.

In addition to discussing the implications of Chapter 5, Dormagen and Michel were also able to verify the assumptions made in the simulation of turnout in institutions. The two authors have access to administrative data by the French National Statistical Office, which also contains information about the registration of voters and their turnout. As mentioned in Chapter 5, French citizens need to be registered as voters to be eligible to vote. 13.5% of the French population older than 80 years live in institutions and are not included in the probability-based samples by the ESS. To estimate the potential bias in electoral participation due to the exclusion of institutions, I make assumptions regarding the size of the institutionalized population eligible to vote and their turnout. I assume that all institutionalized residents are also registered as voters. This assumption is overly simplistic according to the administrative data. In fact, only 4.9% of the French voters older than 80 years are institutionalized (Dormagen and Michel 2019, 48). This is much lower than the share I use in my simulation for this age group.

Regarding the expected turnout in institutions, I assume a lower turnout than in private households justified with the characteristics of institutionalized residents. In Chapter 5, I use several degrees of differences in my calculation, starting with a turnout lower by about 5 percentage points in institutions than in private households up to a turnout lower by about 50 percentage points. Dormagen and Michel clearly confirmed the assumption of a lower turnout in institutions: "*25.8%* [voted] *in the first round of the 2012 presidential election — 37.2 points below the average of their age-group*" (Dormagen and Michel 2019, 48). Dormagen and Michel also assume a distortion of self-reported electoral participation in institutions leading to a higher reported turnout similar to the overreporting in private households (DeBell et al. 2018). In accordance with Chapter 5, they use the same factor for overreporting as in private households in the ESS, multiplying the 25.8% of voters in institutions by 1.38. This would result in a self-reported turnout of about 35.5% in a survey in institutions, which is lower by about 54 percentage points compared to the ESS estimate in private households. This number exactly replicates the most extreme condition in my simulation. Calculating with this turnout rate, Dormagen and Michel estimate a drop in turnout from 89.3% to 86.7% if institutionalized residents older than 80 years would be included in the French ESS. This reduction is smaller than the reduction I anticipated in Chapter 5 (see Table 5.3) due to the lower overall share of institutionalized voters.

The inclusion of institutionalized residents would have a significant impact on survey estimates of turnout in France, especially in the oldest age group and not so much in younger age groups. An evaluation of the reduction of bias by about 2.6 percentage points might not be extremely favorable, given that the overreporting of voting amounts to 26.3 percentage points in this age group (cf Section 1.6.1). In the conclusion of their reply, Dormagen and Michel also emphasize

the importance of selection mechanisms when securing responses of elderly respondents. The nonresponse bias is apparently larger in this group, because the growing group of people with a declining health, a lower cognitive performance, a smaller social capital, and less involvement in society are likely to abstain from taking part in surveys *and* in presidential elections (Dormagen and Michel 2019).