

Response Behavior and Quality of Survey Data: Comparing Elderly Respondents in Institutions and Private Households

Schanze, Jan-Lucas

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schanze, J.-L. (2023). Response Behavior and Quality of Survey Data: Comparing Elderly Respondents in Institutions and Private Households. *Sociological Methods & Research*, 52(3), 1519-1555. <https://doi.org/10.1177/0049124121995534>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

Response Behavior and Quality of Survey Data: Comparing Elderly Respondents in Institutions and Private Households

Sociological Methods & Research
2023, Vol. 52(3) 1519–1555
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00491241211995534
journals.sagepub.com/home/smr



Jan-Lucas Schanze¹ 

Abstract

An increasing age of respondents and cognitive impairment are usual suspects for increasing difficulties in survey interviews and a decreasing data quality. This is why survey researchers tend to label residents in retirement and nursing homes as hard-to-interview and exclude them from most social surveys. In this article, I examine to what extent this label is justified and whether quality of data collected among residents in institutions for the elderly really differs from data collected within private households. For this purpose, I analyze the response behavior and quality indicators in three waves of Survey of Health, Ageing and Retirement in Europe. To control for confounding variables, I use propensity score matching to identify respondents in private households who share similar characteristics with institutionalized residents. My results confirm that most indicators of response behavior and data quality are worse in institutions compared to private households. However, when controlling for sociodemographic and health-related variables, differences get very small. These results suggest the importance of health for the data quality irrespective of the housing situation.

¹ GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany

Corresponding Author:

Jan-Lucas Schanze, GESIS—Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, 68072 Mannheim, Germany.

Email: jan-lucas.schanze@gesis.org

Keywords

hard-to-interview, institutionalized population, retirement and nursing homes, data quality, cognitive impairment, SHARE

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 and they are hard-to-reach because the access to the institutions can be restricted by gatekeepers (Schanze and Levinson 2019; 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, Aaron Van Den, and Karel Van Den 2012; Lang 2014; Luppá 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 of the 21 researchers (66.7 percent) agreed that residents of long-term care institutions are at least somewhat or even very hard to interview (Schanze and Levinson 2019).¹

Nevertheless, the European Union–financed Synergies for Europe’s Research Infrastructures in the Social Sciences 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 aging, and they are conducted at a national level like the French *Enquêtes Capacités, Aides et REssources des seniors*, the U.S. *Health and Retirement Survey*, and the *UK Household Longitudinal Study* or with a cross-national design like the *European Health Interview Survey* 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, Rips,

and Rasinski 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 is 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 it 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, Diani, and Streiner 2007; Knäuper et al. 2016); and (4) the impact of interviewer–respondent interactions (Beullens, Loosveldt, and Vandenplas 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 article. 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 differ 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 sociodemographic characteristics and health condition?

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

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, nondifferentiation, 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, Klaus, and Simonson 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 (Koch and Blohm 2009; Colsher and Wallace 1989; Fuchs 2009; Knäuper 1999; Knäuper et al. 1997; 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, Herzog, and Andrews 1988). This finding was confirmed in a recent analysis by Kroh, Winter, and Schupp (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 depending 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 a standardized but artificial interview process to a more conversational approach resembling everyday conversations (cf. Beullens et al. 2018; 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 is 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 is not sufficient indicator 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 (Salt-house 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; Luppá 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 (Lang 2014). 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. 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 sociodemographic and health-related variables. Given the absence of previous literature on those research questions, it remains to be seen whether differences will completely disappear.

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, nondifferentiation, extreme responses, replies to trap question, or correlations of several variables (Roberts et al. 2019). This article 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, and 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.

Data Preparation and Variables Used

I use waves 5, 6, and 7 of the 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, 2019b, 2019c). 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 5–7 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 data set 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 percent 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 the interview” (Munich Center for the Economics of Aging 2019:13). Even in cases of hearing impairment or Alzheimer’s disease, some parts of the SHARE interview are nonproxy modules and need to be completed by the respondent, if possible.

As the second and third indicator of hard-to-interview respondents, I analyze how interviewers rated respondents in terms of requiring additional clarifications and their willingness to answer questions.^{5,6} 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 (Revilla, Saris, and Krosnick 2014). 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, Olson, and Smyth 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 (von dem Knesebeck et al. 2005). In SHARE, the CASP inventory consists of 12 items measuring four different dimensions, namely, control, autonomy, self-realization, and pleasure (CASP). 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 four-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 five-level Likert-type scale. Both item batteries are in a nonproxy 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 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 nonattitude by selecting the midpoint (Raaijmakers et al. 2000; Sturgis, Roberts, Smith 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 end points 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, Ryser, and Gonçalves 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, Goldberg, and Borg 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 one to five, the mean of items within a dimension is expected to be 3.0. A deviation from this pattern within dimensions and across the five dimensions, in other words any values higher than 3.0, indicates a tendency to agree to mutually exclusive statements to a larger extent than expected by theory. This kind of acquiescence was labeled as weak satisficing and is a sign of a worse data quality. 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.⁸

Propensity Score Matching (PSM)

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 researchers. 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 PSM is commonly used to overcome the selection bias in nonexperimental 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 predefined 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 and 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 article). Age is grouped into four 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 International Standard Classification of Education (ISCED) scheme that aims to generate a comparable measurement of educational achievements across countries. SHARE uses the seven-level ISCED-97 scale, which ranges from 0 (*preprimary education*) to 6 (*second stage of tertiary education*).

As the third step, I add three health-related variables to the abovementioned set of sociodemographic variables in the PSM (*full PSM*). In previous studies, a combination of sociodemographic 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 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 (ADLs) measures the ability of respondents to cope with six different tasks of daily living: 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; and using the toilet, including getting up or down. 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 PSM, 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 abovementioned 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 (*SD*) 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, which is why I excluded countries with 15 or less 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 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 respondent 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, c.f.). Only in the Netherlands, a little less than 20 respondents in private households were available as a potential match for each institutionalized respondent. 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 ($SD = .056$) and between 0.001 and 7.5 in the full PSM ($SD = .11$).

Exploration of the PSM

Table 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

Table 1. Overview of Institutionalized Cases Used in Propensity Score Matching (PSM) and Reasons for Losing Cases in the PSM.

	PSM Condition	Total Inst. Pop.	Missing Values	Beyond Caliper	Matched Cases
Unique cases in waves 5–7	Basic PSM: Sociodemographic variables	2,002	33	3	1,966
	Full PSM: Sociodemographic + health variables		113	44	1,845

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 predefined caliper. Depending on the type of matching, 33 institutionalized respondents in the sociodemographic 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 three institutionalized respondents for the basic sociodemographic characteristics. While matching with sociodemographic 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 show 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 dropouts, 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 dropouts are not at random.

PSM 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 institutionalization. For

the sociodemographic matching only age (odds ratio [OR] = 1.10, $p = .003$) and widowhood ($OR = 0.88$, $p = .023$), for the sociodemographic and health-related matching only age ($OR = 1.11$, $p = .004$) had a very small, nearly absent but statistically significant explanatory power.

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 et al. (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, Tourangeau, and Presser 2019). Checking the panel experience of respondents in my unmatched data set, 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 data set, 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.

Results

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 2. Proxy Interviews in Three Waves of the Survey of Health, Ageing and Retirement in Europe, Comparing Private Households and Institutions.

PSM Condition	Proxy Interviews	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: Tests of statistical significance conducted with bivariate logistic regressions. n.s. = not significant ($p > .05$); PSM = propensity score matching.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 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 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 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 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 sociodemographic variables and continue to decline when matching on health-related variables in addition to the sociodemographic variables. Health seems to be the stronger driver of more difficult interviews compared to age, gender, education, or widowhood. Even though the gap between

Table 3. Respondents Asking for Clarification, Comparing Private Households and Institutions Across Three Waves of the Survey of Health, Ageing and Retirement in Europe.

PSM Condition	Clarifications	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: Tests of statistical significance conducted with bivariate ordinary least squares regressions with institutionalization as explanatory variable. n.s. = not significant ($p > .05$); PSM = propensity score matching.

* $p < .05$. ** $p < .01$. *** $p < .001$.

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 showed a positive impact on the need for clarification for unweighted data.⁹ However, in both PSM conditions, the origin of respondents did not have a statistically significant influence on the dependent variable next to institutionalization.

As the last indicator of hard-to-interview respondents, I consider interviewer ratings regarding respondents' willingness to answer (see Table 4). In

Table 4. Respondents' Willingness to Answer Questions, Comparing Private Households and Institutions Across Three Waves of the Survey of Health, Ageing and Retirement in Europe.

PSM Condition	Willingness	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: Tests of statistical significance conducted with bivariate ordinary least squares regressions with institutionalization as explanatory variable. n.s. = not significant ($p > .05$); PSM = propensity score matching.

* $p < .05$. ** $p < .01$. *** $p < .001$.

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 of the 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 PSM show, sociodemographic variables and health are the true explanatory factors for respondents' willingness to answer, whereas institutionalization only has a marginal impact on the dependent variable.

Table 5. Item Nonresponse in 12 Items of the CASP Inventory in Three Waves of the Survey of Health, Ageing and Retirement in Europe.

PSM Condition	Missing Values	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: Tests of statistical significance conducted with bivariate logistic regressions and negative binomial regressions. n.s. = not significant ($p > .05$); PSM = propensity score matching.

* $p < .05$. ** $p < .01$. *** $p < .001$.

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 10 and 11 in the Online Appendix (which can be found at <http://smr.sagepub.com/supplemental/>) 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 percent). 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 11 in the Online Appendix, which can be found at <http://smr.sagepub.com/supplemental/>).

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 5 and 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.

In institutions, roughly every third respondent did not reply to a least one item in the CASP item battery. On average, 3.3 items 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 sociodemographic variables and health-related variables. The share of respondents in private households with a least one missing value nearly doubles when matching for sociodemographic 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 dropout 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.

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 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 sociodemographic 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 four more missing items in private households ($p = .00$) and 2.5 more missing items in institutions ($p = .00$). Respondents with a better rating of their willingness abstain less often from replying in both item batteries (coeff. = $-.89$, $p = .00$ 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. = $.43$, $p = .00$ 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

Table 6. Item Nonresponse in 11 Items of the Big-Five Inventory in Wave 7 of the Survey of Health, Ageing and Retirement in Europe.

PSM Condition	Missing Values	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: Tests of statistical significance conducted with bivariate logistic regressions and negative binomial regressions. n.s. = not significant ($p > .05$); PSM = propensity score matching.

* $p < .05$. ** $p < .01$. *** $p < .001$.

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.

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 7 and Table 12 (in the Online Appendix, which can be found at <http://smr.sagepub.com/supplemental/>) 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.

Institutionalized respondents pick the middle response option more often than respondents living in private households (see Table 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 sociodemographic variables, before shrinking when adding health-related variables to the set of sociodemographic variables in the PSM. In the full PSM condition, the difference between the two types of housing is not statistically significant any longer ($p = .188$).

Table 7. Share of Big-Five Items Answered With Middle Response Options of All Substantive Replies (Excluding Missing Values) in Wave 7 of the Survey of Health, Ageing and Retirement in Europe.

PSM Condition	Response Pattern	Private Households		Institutions		Diff.
		Percentage	SD	Percentage	SD	
All respondents	Middle resp. option	15.4	.16	17.4	.18	2.0**
	Total N	50,038		703		
Basic PSM	Middle resp. option	15.2	.16	17.4	.18	2.2**
	N (weighted)	1,122		695		
Full PSM	Middle resp. option	16.2	.18	17.4	.18	1.2 ^{n.s.}
	N (weighted)	841		666		

Note: Tests of statistical significance conducted with bivariate ordinary least squares regressions. n.s. = not significant ($p > .05$); PSM = propensity score matching.

* $p < .05$. ** $p < .01$. *** $p < .001$.

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 12 in the Online Appendix, which can be found at <http://smr.sagepub.com/supplemental/>). 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. = .07, $p = .00$) and less often with strong agreement (coeff. = $-.03$, $p = .00$). 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. = $-.06$, $p = .00$) and institutions (coeff. = $-.05$, $p = .00$). 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.

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 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 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 article, acquiescence does *not* change when running through the two stages of PSM. Significant differences between institutionalized respondents and respondents living in private households do not get smaller when matching for sociodemographic 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 ADLs. 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

Table 8. Acquiescence in Various Latent Dimensions of the Big-Five Inventory and across All Big-Five Items in Wave 7 of the Survey of Health, Ageing and Retirement in Europe.

		Private Households		Institutions		Diff.
		Mean	SD	Mean	SD	
All cases	Extroversion	3.36	.70	3.28	.71	-.07**
	Agreeableness	3.20	.63	3.17	.61	-.03n.s.
	Conscientiousness	3.13	.63	3.09	.65	-.04n.s.
	Neuroticism	3.07	.67	3.03	.69	-.03n.s.
	Openness	3.06	.82	3.00	.88	-.07*
	Mean <i>N</i> (<i>SD</i>)	49,819 (61.8)		670 (8.9)		
	Big-Five, all items	3.16	.35	3.11	.37	-.05***
<i>Total N</i>		49,961		688		
Basic PSM	Extroversion	3.38	.70	3.28	.70	-0.10***
	Agreeableness	3.19	.61	3.17	.61	-.01*
	Conscientiousness	3.08	.63	3.09	.65	.01n.s.
	Neuroticism	3.08	.68	3.03	.69	-.05***
	Openness	3.01	.83	3.00	.88	-.02n.s.
	Mean weighted <i>N</i> (<i>SD</i>)	1,110 (2.9)		662 (8.4)		
	Big-Five, all items	3.15	.35	3.11	.38	-.04***
<i>N</i> (weighted)		1,116		680		
Full PSM	Extroversion	3.37	.70	3.27	.71	-.10***
	Agreeableness	3.21	.62	3.17	.61	-.04***
	Conscientiousness	3.10	.65	3.09	.65	-.01*
	Neuroticism	3.09	.70	3.04	.69	-.05***
	Openness	3.05	.83	3.00	.88	-.05***
	Mean weighted <i>N</i> (<i>SD</i>)	821 (3.3)		638 (7.8)		
	Big-Five, all items	3.16	.35	3.11	.38	-.05***
<i>N</i> (weighted)		829		656		

Note: Tests of statistical significance conducted with bivariate ordinary least squares regressions. PSM = propensity score matching; n.s. = not significant ($p > .05$).

* $p < .05$. ** $p < .01$. *** $p < .001$.

item nonresponse and response patterns. Respondents with proxy interviews in other parts of the interview achieve a lower level of acquiescence (coeff. = $-.05$, $p = .00$), while respondents with a better rating of their willingness have a slightly stronger tendency for acquiescence (coeff. = $.04$, $p = .00$). The less pronounced motivation in institutions (see Table 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.

= $-.02$, $p = .00$). 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 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.

The analysis of the factor structure of the Big-Five items sheds a light at the impact of acquiescence on the survey data. Table 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 PSM.

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 four 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 five 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 five factors for private households. Table 9 shows the factor loadings for the Big-Five items that are not corrected for acquiescence. In both populations, only the extroversion 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

Table 9. Principal Component Factor Analysis (Varimax Rotated) of Big-Five Items (Not Corrected for Acquiescence).

PSM Condition	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	
	Percentage 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	
Percentage 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			

(continued)

Table 9. (continued)

PSM Condition	I See Myself as sb. Who ...	Private Households			Institutions						
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
	Percentage 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		

Highest loadings are set in bold.

Note: PSM = propensity score matching.

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 PSM 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 PSM. The factor structure as such does not undergo drastic changes. The total share of explained variance remains very stable (65.8 percent with all cases compared to 66.1 percent in the full PSM). Like for the analysis of acquiescence presented in Table 7, 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 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' *SD*. 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 four instead of five factors should have been extracted. However, the factor loadings of items do not show any distinct advantage of the theory in institutions as compared to private households.

Conclusion

This article 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 article 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 non-response 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 PSM 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 ADLs 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 sociodemographic variables less than 2 percent of institutionalized dropped out. The dropout 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 dropouts 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 (Bergmann et al. 2019). 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 article 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. Sociodemographic characteristics can only explain a small fraction of data quality issues. When comparing institutions and private households while controlling for sociodemographic *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 PSM 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 of the 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 and 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.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Jan-Lucas Schanze  <https://orcid.org/0000-0002-1520-9398>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Schanze and Levinson (2019) only report aggregate results for the entire institutionalized population. The abovementioned 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.

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.
5. Asking for clarification is measured on a six-level scale. I merged the levels “almost never” and “never” resulting in a five-level scale.
6. Willingness to answer is measured on a four-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).
8. Nevertheless, in a robustness check, it did not lead to substantive changes in the results presented in Table 8.
9. 0 = born in country of interview and 1 = not born in country of interview.

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, NJ: John Wiley & Sons.
- Alwin, Duane and Jon Krosnick. 1991. “The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes.” *Sociological Methods & Research* 20(1):139-81.
- 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): 403-10.
- Bergmann, Michael, Thorsten Kneip, Giuseppe De Luca, and Annette Scherpenzeel. 2019. “Survey Participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1-7.” *SHARE Working Paper Series* 41:1-52.
- 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. doi: 10.1093/ijpor/edy031.
- 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+.” FORS Working Paper Series 2015-4:1-9. Lausanne (CH): FORS.
- Börsch-Supan, Axel. 2019a. “Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5.” SHARE-ERIC. doi: 10.6103/SHARE.w5.700.
- Börsch-Supan, Axel. 2019b. “Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6.” SHARE-ERIC. doi: 10.6103/SHARE.w6.700.
- Börsch-Supan, Axel. 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.” *DIW Discussion Papers* 485: 1-29.
- 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):105-19.
- Dehejia, Rajeev H. and Sadek Wahba. 2002. “Propensity Score-matching Methods for Nonexperimental Causal Studies.” *Review of Economics and Statistics* 84(1): 151-61.
- 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* 30:468-81.
- Fuchs, Marek. 2009. “Item-nonresponse in einer Befragung von Alten und Hochbetagten—Der Einfluss von Lebensalter und kognitiven Fähigkeiten.” Pp. 333-49 in *Umfrageforschung—Herausforderungen und Grenzen*, edited by Martin Weichbold, Johann Bacher, and Christof Wolf. Wiesbaden (DE): VS Verlag für Sozialwissenschaften.
- 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, Bartram Dave, Inceoglu Ilke, and van de Vijver Fons. 2014. “Response Styles and Personality Traits: A Multilevel Analysis.” *Journal of Cross-Cultural Psychology* 45(7):1028-1045.

- Hyde, Martin, R. Wiggins, P. Higgs, and D. 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):186-94.
- Kieruj, Natalia 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):817-46.
- Knäuper, Bärbel. 1999. "The Impact of Age and Education on Response Order Effects in Attitude Measurement." *Public Opinion Quarterly* 63:347-70.
- 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):181-99.
- 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):185-93.
- 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):914-42.
- Krosnick, Jon. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5(3):213-36.
- Kühn, Konstanze and Rolf Porst. 1999. "Befragung alter und sehr alter Menschen: Besonderheiten, Schwierigkeiten und methodische Konsequenzen. Ein Literaturbericht." *ZUMA-Arbeitsbericht* 99(03):1-40.
- Laferrère, Anne, Heede Aaron Van Den, and Bosch Karel Van Den. 2012. "Entry into Institutional Care: Predictors and Alternatives." Pp. 253-64 in *Active ageing and solidarity between generations in Europe*, edited by Axel Börsch-Supan, Martina Brandt, Howard Litwin, and Guglielmo Weber. Berlin/Boston: De Gruyter.
- Lang, Gert. 2014. "Zur Befragung und Befragbarkeit von kognitiv eingeschränkten und demenziell veränderten Menschen in Alten- und Pflegeheimen." Pp. 207-29 in *Das erzwungene Paradies des Alters?* edited by Anton Amann and Franz Kolland. Wiesbaden (DE): Springer.
- Lechner, Clemens and Beatrice Rammstedt. 2015. "Cognitive Ability, Acquiescence, and the Structure of Personality in a Sample of Older Adults." *Psychological Assessment* 27(4):1301-11.
- Lenzner, Timo. 2012. "Effects of Survey Question Comprehensibility on Response Quality." *Field Methods* 24(4):409-28.

- Loosveldt, Geert. 1997. "Interaction Characteristics of the Difficult-to-interview Respondent." *International Journal of Public Opinion Research* 9(4):386-94.
- Lunt, Mark. 2014. "Selecting an Appropriate Caliper can be Essential for Achieving Good Balance with Propensity Score Matching." *American Journal of Epidemiology* 179(2):226-35.
- Luppa, Melanie, Tobias Luck, Elmar Brähler, Hans Helmut König, and Steffi 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 Gilbert, 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(4):333-48.
- Motel-Klingebiel, Andreas, Daniela Klaus, and Julia Simonson. 2014. "Befragung von älteren und alten Menschen." Pp. 781-86 in *Handbuch Methoden der empirischen Sozialforschung*, edited by Nina Baur and Jörg Blasius. Wiesbaden, Germany: Springer Fachmedien.
- Munich Center for the Economics of Aging. 2019. "Release Guide 7.0.0." Munich Center for the Economics of Aging, Munich (DE), pp. 1-83. www.share-project.org/fileadmin/pdf_documentation/SHARE_release_guide_7-0-0.pdf.
- 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* 21:145-63.
- 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, Anne van Hoof, Harm 't Hart, Tom Verbogt and Wilma Vollebergh. 2000. "Adolescents' Midpoint Responses on Likert-type Scale Items: Neutral or Missing Values?" *International Journal of Public Opinion Research* 12(2):207-15.
- 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):121-25.
- Revilla, Melanie, Willem Saris, and Jon Krosnick. 2014. "Choosing the Number of Categories in Agree-Disagree Scales." *Sociological Methods & Research* 43(1):73-97.

- Roberts, Caroline, Emily Gilbert, Nick Allum, and Léila 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):264-72.
- Salthouse, Timothy. 2009. "When Does Age-related Cognitive Decline Begin?" *Neurobiology of Aging* 30(4):507-14.
- Sangl, Judith, Joan Buchanan, Carol Cosenza, Shula Bernard, Susan Keller, Nancy Mitchell, Julie Brown, Nicholas Castle, Edward Sekscenski, and Deborah Larwood. 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." *Deliverable 2.17 of the SERISS project funded under the European Union's Horizon 2020 Research and Innovation Programme GA No: 654221*. <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*. <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." Pp. 17-43 in *Cognition, Aging, and Self-Reports*, edited by Norbert Schwarz, Denise Park, Bärbel Knäuper, and Seymour Sudman. Boca Raton, FL: Taylor & Francis Group.
- Schwarz, Norbert. 2007. "Cognitive Aspects of Survey Methodology." *Applied Cognitive Psychology* 21:277-87.
- Schwarz, Norbert and Bärbel Knäuper. 1999. "Cognition, Aging, and Self-reports." Pp. 233-53 in *Cognitive Aging: A Primer*, edited by Denise Park and Norbert Schwarz. Hove, England: Psychology Press.
- Soto, Christopher, Oliver John, Samuel 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):718-37.
- 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 & 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* 7:572-88.
- 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(2): 195-208.
- Tourangeau, Roger. 2014. "Defining Hard-to-survey Populations." Pp. 3-20 in *Hard-to-Survey Populations*, edited by Roger Tourangeau, Brad Edwards, Timothy Johnson, Kirk Wolter, and Nancy Bates. Cambridge, England: Cambridge University Press.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, England: 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.
- von dem Knesebeck, Olaf, Martin Hyde, Paul Higgs, Alexandra Kupfer and Johannes Siegrist. 2005. "Quality of Life and Well-being." Pp. 199-203 in *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*, edited by Axel Börsch-Supan, Agar Brugiavini, Hendrik Jürges, Johan Mackenbach, Johannes Siegrist, and Guglielmo Weber. Mannheim (DE): Mannheim Research Institute for the Economics of Aging.
- West, Brady, Conrad Frederick, Kreuter Frauke, and Mittereder Felicitas. 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.

Author Biography

Jan-Lucas Schanze holds a master of arts in political science from the Free University of Berlin. In January 2016, he joined GESIS—The Leibniz Institute for the Social Sciences as a doctoral researcher. For three years, he worked as a research associate in the European Union–funded Synergies for Europe’s Research Infrastructures in the Social Sciences project (<https://seriss.eu>) until the project came to an end in July 2019. Since August 2018, he worked for the Core Scientific Team of the European Social Survey, and joined the national team responsible for implementing the ESS in Germany in early 2021. His research interests are primarily focused on methodological questions on the inclusion of institutionalized populations in cross-national social surveys (coverage bias, data quality, etc.).