Survey Research Methods (2026)
Vol. 20, No. 1, pp. 43-58
https://doi.org/10.18148/srm/2026.v20i1.8500
European Survey Research Association

Addressing Answer Consistency of Residents in Residential Nursing Homes: Experiences from a Mixed Method Study

Marie-Kristin Döbler[1] · Katrin Drasch[2]
[1]Institut für Sozialwissenschaftliche Forschung, ISF
[2]Institute of Sociology, Friedrich-Alexander University Erlangen-Nürnberg

This paper addresses two methodological debates: first, it examines the consistency of responses in mixed-method designs; second, it does so with a particular focus on older people, both with and without cognitive impairments, living in care institutions. The empirical basis consists of 135 semi-structured interviews with individuals aged 50 to 92, conducted in residential homes for the elderly. We compare responses to open, narrative questions with those to closed questions on three topics: quality of life, decision to move, and cognitive decline. We conclude that responses on the same topic are generally consistent, although declining cognitive ability reduces this consistency. With this article we contribute to the discussion on the feasibility of surveying older (institutionalised) people with and without cognitive impairment, as well as the potential and benefits of combining qualitative and quantitative data.

Keywords: answer consistency; cognitive decline; institutionalised living; mixed methods; older people

1 Introduction

In 2021, 17% of older people in need of long-term care in Germany lived in residential nursing homes (Destatis 2024a). With residential nursing homes we subsume three different types of homes: assisted living facilities (Betreutes Wohnen), retirement homes (Altenheim) and nursing homes (Altenpflegeheim). Residents of these homes usually have physical and/or cognitive impairments, often age-related, and are generally perceived as (very) old, although there is no clear definition of this term (Kaspar et al. 2023). There are two main reasons for this: first, the homes we study are dedicated to "old people," as reflected in names such as "Altenheim" (literally translated: home for old people), and "being old" was originally a prerequisite for residency. While younger people requiring care also sometimes live in these facilities due to a lack of places in more age-appropriate institutions, chronological age still tends to serve as a primary selection criterion. Second, "being old" is often defined in terms of declining capabilities, with residents sometimes categorised as being in the third or fourth age, or as "frail old" (Kydd et al. 2018). In either case, these age groups are defined by the presence or absence of vitality, physical and mental abilities that constitute their health status. Thus, the chances of living in a residential nursing home depend on the age as well as the health status. Residents of such homes receive assistance with everyday tasks or are entirely relieved of these responsibilities, as they become the home's duty. The type and intensity of care required of course varies. In sum, this leads to the perception that residents of such homes are a homogenous group of "frail old".

Associations with decay and decline are reflected in ongoing methodological debates concerning studies with older people in residential nursing homes (Kaspar et al. 2023;

Supplementary Information The online version of this article (https://doi.org/10.18148/srm/2026.v20i1.8500) contains supplementary material, which is available to authorized users.

Corresponding author: Marie-Kristin Döbler, Institut für Sozialwissenschaftliche Forschung, ISF, München, Germany (Email: marie-kristin.doebler@fau.de)

Schanze 2023), which ultimately limits empirical research. Regardless of age, it is often assumed that people living in institutions are difficult to reach (e.g., due to infrastructural barriers) and challenging to study (e.g., due to special needs and limited capacities). As this would require tailored research methods (Dillman et al. 2014), they are often excluded from general population surveys (Kelfve et al. 2013; Schanze 2023). Additionally, studies conducted in institutions are frequently viewed with scepticism (Kelfve et al. 2013; Krosnick 1991; Tyler et al. 2010). However, there is little empirical evidence supporting assumptions such as that the data is less reliable, non-reproducible, and more susceptible to social desirability bias (Kelle et al. 2014).

To address the consistency of responses or data, comparisons can be made between (1) aggregated and individual results; (2) a participant's responses to the same question at different points in time; or (3) a participant's responses to different questions on the same topic using varied approaches (e.g., closed- and open-ended questions). We use the third type of comparison to assess the consistency of data in research with older people living in institutions. Drawing on a mixed method study conducted in Bavarian residential nursing homes (cf. Sect. 3.1), we explore the methodological question: Do responses from the quantitative and qualitative parts of an interview lead to similar conclusions? In other words, are responses to different types of questions on the same topic consistent? For this purpose, we developed a two-step analysis strategy (cf. Sect. 3.3). First, we converted (narrative) answers to open-ended questions into the same form as the answers to the closed-ended questions. Second, we compared the quantified qualitative data with the original quantitative data. To the best of our knowledge, up until now no studies have utilised such analyses. With this paper, we are therefore contributing to two debates: on mixed method designs as well as on the possibility to survey older (institutionalised) people with and without cognitive impairment. Overall, we conclude that responses on the same topic are to some extent consistent and that older institutionalised living people can participate in research, although deteriorating cognitive abilities reduce this consistency and make interviews more demanding.

We begin by considering the methodological challenges of researching older people in residential settings, along with the potential and limitations of mixed methods (2). Next, we describe the empirical background of this article and the study design, outlining the project context, sample, and analysis strategies (3). We then assess the consistency of responses to qualitative and quantitative questions in our study by examining three exemplary measures in detail (*quality of life, decision to move*, and *cognitive decline*) (4), draw conclusions about the validity of our assumptions, and discuss the limitations of our study (5). Finally, we provide recommendations for future research within the institutional context of residential nursing homes for older adults, and critically reflect on this paper (6).

2 Conducting Research with Older People (in Institutions): Methodological Problems and Solutions

In the following, we will discuss aspects that are perceived as obstacles to research on older people, people living in institutions, and in particular older people living in institutions. Following Fisk and Wigley (2000), these can be categorised as (a) problems with access and (b) problems with participation.

2.1 Getting Access and Finding Participants

Two key aspects for problems with access are related to institutional structures. First, conducting research in a facility requires obtaining permission to access the facility and to contact its members, especially its residents. This entails convincing institutional gatekeepers, i.e., in our case those responsible for the care homes, usually the managers, and sometimes also the umbrella organisations (e.g., in Germany: AWO, Caritas, Diakonie) in an often time-consuming process.

Second, there are gatekeepers within the institutions. Sometimes these gatekeepers are the same individuals who regulate access to the institutions (e.g., home managers). However, researchers who have gained access often encounter other gatekeepers, such as ward or shift managers, care staff, or service staff. These individuals may select or introduce "volunteers" to the researcher, permit or restrict access to potential respondents (Kelle and Niggemann 2002; Kelle et al. 2014; Motel-Klingebiel et al. 2019), shape the initial contact, and prepare prospective interviewees. Consequently, multiple gatekeepers limit researchers' control over the sampling of respondents, potentially undermining or necessitating—just like the prevention of general access to the homes by gatekeepers—changes to sampling strategies and leading to "biased samples".

Empirical research on older individuals (e.g., Kaspar et al. 2023) has identified a third group of gatekeepers who are not specific to any institution: family members and relatives or, more broadly, those responsible for the older person's affairs. Even if they do not officially have custody, they often feel responsible or are required to act as guardians. As a result, they can also prevent older people from participating in surveys (Hall et al. 2009).

Sometimes researchers may need the cooperation of these relatives, as they provide informed consent on behalf of the older people in question. This may be ethically necessary, especially when dealing with vulnerabilities

and cognitive impairments (Kühn and Porst 1999; Motel-Klingebiel et al. 2019; Wenger 2003), or legally required in cases of custody. However, from an ethical perspective, this is only the second-best approach for gaining access to older participants. The considerations here are similar to those in proxy research, where caregivers, family members, and relatives are interviewed as proxies for the intended respondents who are deemed unable to respond themselves. Yet, as Kaspar et al. (2023) indicate, this approach has limited benefits. On the one hand, using proxies affects the representativeness of the data positively; on the other hand, it restricts the scope of the questionnaire and alters the meaning of the answers. Proxies report indirectly on experiences or conditions, assessing, for example, the quality of life in old age through their own perspectives (Smith et al. 2020; Wenger 2003). While this may sometimes be the only viable option (e.g., in the case of comatose patients), it is often chosen simply because it is easier, i.e., for pragmatic reasons.

In conclusion, these issues contribute to a significant underrepresentation of individuals residing in care facilities in general population surveys. This results in limited research findings on people living in these institutions and biased comparisons between residents and non-residents (Kelle et al. 2014; Motel-Klingebiel et al. 2019; Wagner et al. 2019). In the following, we shift our focus from (a) problems with access to (b) problems with participation.

2.2 Participation

People living in institutional settings such as retirement and nursing homes (as well as clinics or prisons) are often perceived as a hard-to-survey population (Hall et al. 2009; Lundström 1987; Schanze and Levinson 2019; Schanze 2023). However, recent findings suggest that the difficulty of researching this population is not necessarily due to their place of residence (once access is granted). Instead, multiple factors often contribute to an individual's decision not to participate in research (Fisk and Wigley 2000), with health identified as the primary indicator of non-participation (Wagner et al. 2019; Schanze 2023). It is therefore important to distinguish between those who are hard-to-reach and those who are hard-to-interview (Schanze and Levinson 2019; Schanze 2022), and to avoid categorising all institutional residents indiscriminately as hard-to-interview (Schanze 2023). Among older people, specific challenges include physical and cognitive problems (Kühn and Porst 1999; Motel-Klingebiel et al. 2019; Wenger 2003), ranging from hearing impairments or acute illnesses and pain to cognitive decline. Some of these challenges can be overcome, e.g., by adapting the interview material and the interview design to the specific characteristics of this population. Using larger fonts for better legibility (Motel-Klingebiel et al. 2019), instructing interviewers to speak slowly and clearly, to repeat questions, allow pauses, and offer breaks can limit or overcome problems, such as dropouts due to exhaustion, distress, or annoyance. Thus, while the interview location for those living in institutions can hardly be controlled by researchers, making this population hard to reach and potentially resulting in specific interview settings, other elements can be designed to facilitate the interview which makes it easier to survey the hard-to-interview population and increasing the likelihood of responses.

Similarly, keeping questions and possible answers as simple as possible, limiting the length of the questionnaire, avoiding the feeling of being tested and fear of poor performance can decrease the likelihood that respondents stop answering. Whereas the accuracy and completeness of self-reports decline as questions become more difficult (Knäuper et al. 2016; Schanze 2023), answer and data quality benefit from less complexity and shorter questionnaires. Additionally, the nature of potential answers must be taken into account. On the one hand, it has been reported that especially older people with cognitive impairments struggle with questions involving a long recall period, frequency estimations, or complex reference points, e.g., in time or in comparison to other persons (Knäuper et al. 2016; Schanze 2023). On the other hand, Likert-type questions are challenging and can create cognitive overload because they require respondents to understand the items and respond repeatedly on an abstract scale (Isaksson et al. 2007; Knäuper et al. 2016; Reuband 2006). Furthermore, Likert-type questions are commonly utilised to evaluate latent attitudes or estimations that present a greater cognitive challenge to respondents than factual questions.

Standardised research interviews with older people in residential nursing homes also face a unique challenge: the interview's structure, type, and often its content can resemble the assessment procedure conducted by the Medical Service. In Germany, the Medical Service assesses eligibility for benefits and allocates them based on supposedly objective medical criteria, while also monitoring quality aspects in residential facilities. Being surveyed by the Medical Service is a regular experience for residents, but this process often leads to biased answers, as both residents and homes benefit financially from high degrees of care needs and favourable institutional ratings. This dynamic may tempt, push, or even pressure residents to provide particular answers. Consequently, it is debatable whether standardised questionnaires can objectively and reliably measure resident satisfaction and quality of care services (Kelle et al. 2014), and whether they generally carry a higher risk of producing socially desirable answers.

It therefore remains unclear whether standardised interviews are suitable for questioning older people in institu-

tions (as Lang et al. 2007 suggest) or not (as Kelle et al. 2014 argue). At the same time, there are increasing efforts to overcome challenges associated with quantitative research and standardised interviews, many of which involve combining these methods with qualitative approaches (Hall et al. 2009). Nonetheless, both quantitative and qualitative empirical research involving older respondents in institutional settings remains relatively scarce, and comparatively little is known about this growing group of people.

2.3 Mixed Method

Quantitative and qualitative research are often perceived as fundamentally different approaches within the social sciences. They operate with distinct epistemologies (Hathaway 1995), follow different research strategies, utilise diverse methods for data collection and analysis—ranging from numerical data in quantitative research to textual or narrative data in qualitative research—and employ different criteria for evaluating research (Bryman et al. 2008; Hammersley 2023). This has led to a perceived divide between these two dominant methodologies (Kim 2019).

However, there have been numerous efforts to integrate quantitative and qualitative methods at various stages of the research process (Knappertsbusch 2023; Krause 2002). Despite scepticism about the possibility and value of combining qualitative and quantitative methods, as well as debates over whether mixed methods constitute a "third way" or "third paradigm" (Kelle and Reith 2023), the combination of methods has become increasingly popular in contexts where a single method is insufficient to address the full spectrum of a research topic, or when the resources of either researchers or participants are limited.

The reasons for combining methods vary widely (Collins et al. 2007; Morgan 2019), and mixing methods can offer several advantages. Some researchers aim to overcome perceived limitations or shortcomings of a single method, such as lack of representativity (Benítez et al. 2022; Tanner 2023). Others seek to provide a more comprehensive understanding of a phenomenon by examining it from different perspectives (Eickelmann and Burzan 2023; Howe 2012). For example, open-ended questions in a survey can elicit more detailed and meaningful responses, revealing the reasoning behind answers to closed-ended questions or shedding light on the respondents' thought processes through think-aloud statements. Others use exploratory qualitative research as a foundation for formulating hypotheses, which are then tested using quantitative surveys (Akremi and Zanker 2023; Kim 2019; O'Halloran et al. 2018), or conversely, start with quantitative research to identify trends and patterns and then use qualitative data to explain or deepen these statistical findings through illustrative case studies (O'Halloran et al. 2018). Method combinations are also commonly used to confirm or validate findings (Benítez et al. 2022; Hesse-Biber 2012; Howe 2012). Furthermore, some approaches employ both methods iteratively or in parallel (Barbosa Neves and Baecker 2022; Benítez et al. 2022; Collins et al. 2007; Kelle et al. 2014). Additionally, less standardised interviews can create a more interactive setting, which can be beneficial where standardised questioning feels like an evaluation or assessment, such as those conducted by the Medical Service (as mentioned earlier). An example is Kelle et al. (2014), who integrate qualitative interviews into a standardised questionnaire.

However, employing mixed methods also necessitates careful consideration of the opportunities, limitations, and potential shortcomings of each method (Collins et al. 2007). Researchers must ask whether there is any added value in combining methods (McKim 2017) and establish criteria for assessing the quality of mixed methods research (Hammersley 2023). They must also address potential conflicts between qualitative and quantitative findings (Arnon and Reichel 2009; Slonim-Nevo and Nevo 2009), verify whether these methods are genuinely addressing the "same thing", and ensure methodological coherence (Collins et al. 2007). So far, assessing data quality has not yet led to the establishment of a standardised way to address aspects of reliability, validity, and consistency. Instead, each study focuses on its specific shortcomings, problems that are specific to certain methods or study designs, while others appear to be more general but are nevertheless approached from a particular methodological perspective.

Against the backdrop of the methodological problems discussed above in connection with research on older people, institutionalised living people, as well as advantages and limitations of mixing methods, we conducted our research described in the next section. At this point, however, we want to emphasise that this article has a methodological focus: We deal with the consistency of answers given on the same topic in open- and closed-ended questions, not with the content or the respective results. Thus, we address the feasibility and usefulness of a parallel mixed methods design in the form of integrated qualitative interview questions within a standardised questionnaire, assuming that the overall conclusions derived from the quantitative and qualitative parts should ideally be similar.

3 Data & Methods

Drawing on a mixed method study conducted in Bavarian residential nursing homes, we explore the methodological question: Do responses from the quantitative and qualitative parts of an interview lead to similar conclusions? Thus, we

examine the consistency of our data by exploring whether the answers to open- and closed-ended questions on the same topic are consistent.

To this end, we selectively analyse three topics addressed using both open-ended questions and Likert scale items in semi-structured interviews with the same individuals conducted at the same time. The simultaneous collection of qualitative and quantitative data on the same topic offers an excellent opportunity for comparison. We will present the background of our project, describe the data, outline the strategies used to translate qualitative data into numerical form for comparison, assess the consistency of responses, and formulate expectations that guide the presentation of results.

3.1 Project & Data

This article is based on the project "Quality of Life in Bavarian Residential Nursing Homes" (Lebensqualität in Altenpflegeheimen), which was a part of the interdisciplinary research network "ForGenderCare". This network was funded by the Bavarian State Ministry of Education and Cultural Affairs, Science, and the Arts. Therefore, the study concentrates on the German federal state of Bavaria. The project aimed to explore the experienced quality of life among older people living in residential nursing homes. To achieve this, we adopted a parallel mixed method design (Lundström 1987; Maas et al. 2002). Different than usual, the qualitative and the quantitative parts did not address distinct research questions, but rather investigated whether institutional and organizational factors influence the residents' experienced life quality (Döbler 2021). This created some amount of overlap with information available from open-ended as well as closed-ended questions.

We designed our research against the backdrop of the previously discussed methodological considerations concerning the feasibility of interviewing older people in institutions. Thus, we aimed to create an interview setting and instrument that was compatible with the physical and cognitive conditions of residents. In accordance with the recommendations outlined in the literature (Kühn and Porst 1999; Lang et al. 2007; Schanze 2023), we used face-to-face personal semi-structured interviews.

We conducted three forms of pretests: 1) group discussions with experts (sociologists, qualitative social scientists, nursing staff) to review the interviewer guidelines, become familiar with the order of questions and answer categories, and identify key aspects for the interviewer training; 2) cognitive interviews to explore how interviewees respond to questions and to identify misunderstandings, ambiguous wording, or unintended interpretations; 3) a pilot study in one home to test the entire data collection process under real conditions and to prepare for practical matters such as interview structure and interviewer instructions. The first and second pretest resulted in repeated revisions and improvements of the interview guidelines, the third impacted the timing and structure of the interviews. Drawing on our pretest experiences and the understanding that older people may need more time and prompting to answer questions (Isaksson et al. 2007), interviews were scheduled to last between 45 and 90 min. This longer duration allowed interviewers to socialize to some extent with the respondents. Moreover, we aimed to be flexible with study procedures and to give room to further thoughts, our approach broke away from the strict sequence of closed-ended questions and gave the interviews a more conversational character (Kühn and Porst 1999; Schanze 2023; Zermansky et al. 2007). Thus, interviewers were required not to interrupt digressions, enabling respondents to elaborate on topics that were important to them. Overall, this strategy proved effective, as the interviewees seemed to appreciate the opportunity to discuss a broader range of topics in a format different from the rigid structure of standardised questionnaires with closed-ended answer options (Kelle et al. 2014).

Assuming that the older people were capable to form opinions about their present lives, questions were primarily oriented towards the current circumstances of the respondents, with a deliberate effort to minimise the inclusion of questions pertaining to attitudes and retrospective perspectives. On the assumption that simple Likert scale items can be answered even by individuals who have experienced some cognitive decline, the response options were often limited to "yes", "partially", and "no".

Two of the three measures of interest for this paper were first addressed with an open-ended question, followed by closed-ended questions on the same topic. We adopted this order because it allows interviewees to respond in their own words and to present their own relevance structures without being influenced by pre-given answer categories or limited to a particular answer universe. Nevertheless, ordering effects may occur, e.g., when interviewees try to give consistent answers and therefore match closed-ended responses on the same topic with their previously presented narrative. The reasons for combining open-ended questions and closed-ended questions on the same topic were to reduce the feeling of being interrogated, which can easily arise in a fully standardised interview (Kelle and Niggemann 2002), and to meet the older people's desire to narrate, which was identified as pronounced in the pretest. Although our experiences showed that alternating between closed and open-ended questions occasionally caused confusion for some interviewees, they generally managed the questionnaire and its alternating formats effectively.

The third measure of interest for this paper is a proxy-rating on the interviewees' cognitive abilities. Thus, in the

following we rely on how the interviewers assessed the interviewees right after the interview. The research team intentionally chose not to include standardised cognitive impairment tests, such as the MMST (Folstein et al. 1975) or DemTec (Kelle et al. 2014), as cognitive impairment was not the study's primary focus. The fact that we avoided these tests also helped to reduce the similarity with the assessments of the Medical Service, which could otherwise have influenced the answers. Instead, the interview process was designed to allow interviewers to gauge the interviewees' cognitive status in more subtle ways. For example, interviewers were encouraged to ask interviewees to repeat the purpose of the interview before it officially began. This served a dual purpose: it ensured informed consent by confirming that interviewees understood the interview's general aim and their role in the study, and it provided a preliminary assessment of their cognitive state. Likewise, questions about the age of the interviewees' children were included not only to obtain information about their families, but also to assess whether the interviewees were living in the "here and now". The cognitive assessment was further supported by the face-to-face interaction and the extended duration of the interviews.

3.2 Sampling & Sample

Data collection took place from September 2016 to September 2019. Initially, a stratified random sample of all Bavarian residential nursing homes was planned, based on administrative districts within Bavaria and the type of provider (commercial, non-profit, public). However, this approach proved to be unfeasible due to access issues, mainly related to home managers acting as gatekeepers (see Sect. 2 for further discussion). Home managers were frequently reluctant to participate due to concerns about potential consequences of their consent. Many did not respond at all, while others cited lack of time to accommodate research because of the existing bureaucratic burdens from health or care insurances, staff shortages, and generally tight schedules. This necessitated modifying the sampling strategies at the level of individual residential nursing homes, resulting in a sample which is no longer random or representative. Instead, we adopted a theoretical sampling method (Dimbath et al. 2018) to highlight contrasts in home characteristics such as location (urban/rural), provider, and size (<50, 50–100, >100 residents). Ultimately, data were collected in 12 different residential nursing homes, with at least one home per administrative district in Bavaria. Our final sample shows a similar distribution across the different home sizes as all Bavarian residential care homes (cf. Table B1 in the online appendix). However, due to the aforementioned gatekeeper problems and the small sample size at the home level, we were unable to achieve comparable distributions with respect to either sponsorship or administrative district.

Home managers were first contacted by mail and provided with general information about the project, and telephone follow-up enquiries. Encountering continued problems with getting access, we started to attach a letter of support by the Bavarian state government in our first-contact emails. If home managers agreed to participate, interview dates were arranged, considering the typically tight schedules of these institutions, limiting the "time window for an interview" (Zermansky et al. 2007), e.g., through fixed meal times, nap time and leisure activities. Consequently, interviews were planned for the time between lunch and dinner, in some cases also between breakfast and lunch.

The overall team of 27 interviewers consisted mainly of two groups: experienced scientific staff familiar with both quantitative and qualitative interviewing, and sociology students (B.A. or M.A. level) who had personal or professional experience with older people and had received specialised training. This training covered general interviewing techniques and specific issues related to interviewing older people, such as potential physical and cognitive impairments. Additionally, the GDS description of states of cognitive impairment was discussed and interviewers were provided with detailed guidance on how to evaluate and grade the cognitive status of the interviewees. The interviewers received instructions on interpretation criteria and were given examples to ensure consistency and reliability.

On the scheduled interview days, about five interviewers arrived on site at noon to conduct interviews with residents, who had been informed in advance. Managers had been asked to distribute the provided information sheets to residents (and their relatives), outlining the project and the visit by interviewers in simple language. This was to enable residents to make an informed decision about their participation and to discuss it with a trusted person if they wished.

Depending on the size of the facility, resident composition, and the preparatory work of the manager and geriatric nurses, between 5 and 21 interviews were conducted per home. In practice, we employed what we termed "spontaneous on-site sampling" of residents rather than the initially planned random sampling strategy. This was necessary because, firstly, due to data protection, home managers often refused to provide a list of residents from which a random sample could be drawn. Secondly, and more importantly, staff identified only a portion of residents as both able and willing to participate (see Sect. 2 for a discussion of gatekeeping and other methodological issues). However, some arranged interviews could not take place because residents were temporarily unavailable, changed their minds, or had a "bad day". Conversely, interviews were sometimes con-

ducted with residents who had not been initially selected by managers or caregivers, but who expressed spontaneous interest in participating. On the one hand, this might have increased effects of positive self-selection on side of interviewees—it can be assumed that the likelihood of self-initiated voluntary participation is higher among "fitter", "healthier" home residents. On the other hand, this could have helped to minimise the issue of positive pre-selection by caregivers. Additionally, it complies with our aim to interview all residents within a home who were willing and able to be interviewed. However, we can only reflect on possible effects of positive self- and pre-selection as we have no comparison data—neither in form of random samples within homes nor of non-institutionalised older people.

We approached a total of 134 residents (an average of 11.2 per home). All residents with whom an interview was started had informed the care staff that they were willing to be interviewed and had agreed to take part in the study. However, eight of the initial interviewees withdrew their consent to participate after the interview had started or did not agree to be recorded. We presume that the number of drop outs, especially in face-to-face contacts, is caused by older people's fear of saying anything "wrong" and their dislike of being tested. For both, ethical and methodological reasons, our analysis sample includes only those who consented to be interviewed and agreed to have their interviews recorded. Taken together, the sample, for which both quantitative and qualitative data can be compared, amounts to a maximum of 126 interviews.

However, we have to deal with lower numbers of valid cases in our analysis. On the one hand, there is item nonresponse to the closed-ended questions. On the other hand, not all interviews could be (fully) transcribed due to low-quality recordings, or because there was nothing to be transcribed for selective measures, e.g., as interviewees did not narrate but rather answered in monosyllables or diverged too much from the questions' topic. Some brevity of responses might have been caused by the interview's structure, i.e., the alternation of open- and closed-ended questions, or the residents' orientation on their experience with surveys, e.g., those by the Medical Service. Qualitative data may therefore be marked by a potential positive selection, which we will address in the discussion section of the paper.

Our interviewees are aged between 51 and 100 years, with an average age of 83.1 (standard deviation: 9.3 years). About three quarters of our interviewees were female and one-quarter male. Our sample thus reflects the age structure and gender distribution typical of residential nursing home populations (69 % female in Germany in 2021) (Destatis 2024b). However, compared to the Bavarian population in need for care (Statistisches Landesamt Bayern 2014) there are some differences regarding the care degrees of the interviewees: on average they were in a better state of health (see online appendix Table B2). This difference was most likely caused by pre-selections mentioned above, such as carers' evaluations of residents' health, (i.e., ability to participate in the survey), as well as the fact that only fitter residents volunteered. The average interview duration was 42 min (standard deviation: 22 min), ranging from 3 to 120 min.

3.3 Analysis Strategy

In line with the Total Survey Error (TSE) framework (Groves and Couper 1998; Kaspar et al. 2023), we consider non-observation errors (including coverage, sampling, and nonresponse errors) and measurement errors from observations as joint indicators of survey quality. Although it is not possible to compare responses with a "known" truth—as such a truth often does not exist—we suggest that greater deviations indicate a higher degree of unreliability in the responses. In this paper, we focus on the consistency of answers. To do this, we compare the answers that a person gives to standardised questions on a specific topic with their qualitative-narrative statements on the same topic in the context of the same interview.

In a sense, we translated qualitative data into quantitative data by quantifying narratives. For this, interview transcripts were first imported into MAXQDA[1], a software for analysing qualitative and mixed methods data (MAXQDA 2022). The transcripts were then coded blindly (i.e., without reference to corresponding quantitative data) using a coding scheme (see online appendix A) that mirrored the closed-ended questions' answer categories. Online Appendix A provides an illustrative example of the coding procedures for the three different measures described below (Tables A2–A4). The coding was supervised for accuracy, and checked by the supervisor for completeness and consistency. The coded qualitative data set was then merged with the quantitative data set generated from the initial closed-ended questions, and the combined data were analysed using SPSS 29 (IBM 2023) and Stata 18 (StataCorp 2023). This process allowed for several comparisons of responses on the same topic in both open- and closed-ended formats.

Our analysis focuses on three distinct measures: two self-ratings on the perceived *quality of life* and the *decision to move into a home*, and one third-party rating on the *degree of cognitive decline*.

The first measure refers to the first topic addressed in our semi-standardised interviews, the *quality of life*. After gaining participants' consent, we started the interview. We used a combination of a statement and open-ended question to build a good relationship with the interviewer to facilitate

[1] We used MAXQDA 2022 (VERBI Software 2021) for data analysis.

participants' engagement with the interview process and to encourage narratives: "Our study deals with the topic of quality of life. In this context, I would first like to ask you: What things are important to you in life?" Thus, without asking directly about the quality of life—which would have been difficult to put into an open-ended question—we addressed it. We then followed up on it with the closed-ended question: "How would you rate your quality of life in general?", as suggested by Brod et al. (1999). Responses were operationalised using a global self-reported rating on a five-point scale. The maximum scale anchor was labelled "very good" (1) and the minimum "very poor" (5), with intermediate labels of "good" (2), "medium" (3), and "poor" (4). It was not uncommon for interviewees to add shorter or longer narratives related to the overall topic of this first measure when answering this and subsequent closed-ended questions. Therefore, all passages dealing with aspects of quality of life, even if scattered throughout the interview, were taken into account during coding and an overall estimation was given (see Table A5 in online appendix A).

The second measure, the *decision to move* into the residential nursing home, was raised in the middle of the interview. We asked about the circumstances and reasons for moving to the retirement home, without specifying answer categories, and then posed the follow up question: "It is not uncommon that the decision to move into a residential nursing home is not made alone. What was that like for you?" Interviewees could choose from three answer categories: "self-decided", "made together with others", and "others made the decision".

The third measure differs from the first two, as it is based on proxy rating. After conducting the interview, interviewers rated their general impression of the interviewees' cognitive functioning. To assess the "*degree of cognitive decline*", we used the Global Deterioration Scale (GDS, Reisberg et al. 1982)—a seven-point Likert scale, ranging from stage 1 ("no cognitive decline") to stage 7 ("very severe cognitive decline" or "Severe Dementia"). Reisberg and colleagues aligned GDS stage 3 with Mild Cognitive Impairment (MCI) (Reisberg et al. 2011).

Using the mentioned answer categories and five- or seven-point scales, the coder assessed the entire interview and assigned categories or points based on the transcript. Below, we compare the interviewees' and interviewers' ratings with the coder's subsequent assessment.

In general, we calculate the mean and standard deviation for each measure, where applicable; estimate bivariate correlations between measures derived from the standardised part of the interview and those assigned by the coder; and report valid measures to assess the selectivity. As we rely on a comparatively small sample size for quantitative empirical research, we did not apply tests for differences (e.g., t-tests). Instead, we rely on correlations that require much smaller sample sizes. To be more precise, the consistency of the quality-of-life instrument and the GDS are measured on an ordinal scale, so we use Kendall's Tau-b as a measure of rank correlation. However, by treating the measures as a quasi-metric scale, we also calculate Pearson's r. For both correlations sample sizes >30 are sufficient (Bonett & Wright 2000). Based on available tools for power calculations (G*Power) we conclude, and subsequently operate on the premises, that with an expected medium Pearson's correlation of 0.4 and an intended alpha error (α) $\leq$ 0.05, a sample size of 75 is sufficient to achieve acceptable power.

To assess whether deviations depend on the initial values, we test whether the correlation between the quantitative measurement and the absolute difference of the qualitative assessment is significant.

Due to the low number of cases, we also report significances at the 0.1 level. Additionally, we collapsed GDS categories twice, allowing us to estimate Phi for 2 × 2 tables. First, we contrast category 1 with the artificially created category 2+, which summarises all the cases assigned to the categories 2 or higher. With this we intended to contrast cases without dementia with cases with at least mild symptoms of cognitive impairment. Second, we collapsed categories 1 and 2 and contrasted with the collapsed category 3+. This was to compare "non-dementia" and pre-dementia stage with "dementia". To assess the consistency of responses on the decision to move into a residential nursing home, which is measured on a nominal scale, we use Cramér's V and Lambda as measures of correlation.

Furthermore, we use Sankey plots to illustrate individual-level variation and response patterns. Sankey plots are flow diagrams that visually represent the distribution and quantity of flows between different categories using arrows with widths according to the aforementioned frequency of changes (Naqvi 2025).

In sum, our analysis is guided by three assumptions regarding answer consistency. First, we assume that, despite methodological challenges, answers to the same topic given in both closed- and open-ended questions will show some level of consistency, and we expect at least a moderate correlation between them. Second, we anticipate that responses to standardised questions may be slightly more positive due to social desirability, whereas qualitative answers, which require more elaboration, might be less controlled. Third, we expect that answer consistency decreases with declining cognitive function, leading to lower consistency among individuals with higher GDS scores.

4 Results

The results are organised as follows: We sequentially examine the three measures referred to in this article—*quality of*

life, decision to move, and *cognitive decline*. Each section begins with a presentation of the descriptive statistics and correlation measures in a dedicated table, which is then discussed in detail. Throughout, "assessment" refers to the coded data, while "rating" is used for the data collected during the interviews—namely, the interviewees' responses for the *quality of life* and *decision to move* measures, and the interviewers' ratings for the measure *cognitive decline*. In order to make possible selection processes more transparent, we report both, the univariate and the bivariate valid cases, and evaluate on selectivity by assessing measures of central tendency or distribution.

4.1 Quality of Life

Table 1 presents the means, standard deviations, and rank correlations for the *quality of life* measure, based on both the ratings and the assessment.

Table 1

Comparison of rating and assessment for "Quality of life" measure

	Mean	SD	*N*
Quality of life (rating)	2.16	0.846	125
Quality of life (assessment)	2.41	0.879	73
Difference (assessment-rating)	0.25		
Quality of life (rating)—valid	2.19	0.828	73
Quality of life (assessment)—valid	2.41	0.879	73
Difference (assessment-rating)—valid	0.22		
Correlation of within person answers			
Kendall-Tau-b		0.384***	
Pearson's r		0.425***	
Correlation of quality of life (rating) and absolute difference of assessment and rating			
Kendall-Tau-b		–0.039	
Pearson's r		0.070	
Assessment equals rating		48%	
Assessment better than rating		36%	
Assessment worse than rating		19%	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

The mean self-rated *quality of life* in the full available sample (N = 125 (valid); mean: 2.16) is nearly identical to the mean in the available transcripts (N = 73; mean: 2.19), suggesting low selectivity with respect to the *quality of life*-measure. Estimating the rank correlation between ratings and assessments yields a medium, positive, and highly significant correlation (Kendall's Tau-b: 0.384, $p < 0.001$; Pearson's r: 0.425, $p < 0.001$). This result indicates a medium correlation, which is not as high as expected. Without having any empirical proof of this, we suspect that this is a result of the scale itself because determining quality of life precisely on a 5-point Likert scale may have been challenging. Nonetheless, there is some correspondence between the two types of measurements.

Descriptively, about 48% of the ratings by interviewees and the assessments by the coder are consistent, i.e., the self-rating prompted by the question "How would you rate your quality of life in general?" and the coder's assessment of interviewees' narratives result in assigning the same point on the 5-point Likert scale. Approximately 36% of the answers given to the open-ended qualitative questions were (coded) more optimistic than the answers to the closed-ended question, usually with a one-point difference on the 5-point Likert scale, while 19% were (coded) more pessimistically. A pronounced difference of two scale points in either direction occurred in only 11% of the comparisons.

Moreover, the overall value of life quality does not affect answer consistency. The absolute difference between the qualitative assessment and the quantitative rating indicates no systematic bias depending on the absolute value of life quality correlation (Kendall's Tau-b: –0.039, $p = 0.772$; Pearson's r: 0.070, $p = 0.557$). Thus, answer consistency appears to function independently of the initial value of life quality as reported by the respondent. The Sankey plot in Table C1 in the appendix shows to some extent stability, although the coder tended to assess the quality of life more negative using open-answers than the interviewee's rating.

4.2 Decision to Move into a Residential Nursing Home

Table 2 shows the distribution of responses to the closed-ended question regarding who decided on the move into the residential nursing home, and how these responses were coded based on the open-ended interview question.

Significant descriptive differences can also be seen when the number of cases is kept constant by only using valid responses for both the open- and closed-ended questions (N = 52). Compared to the interviewees' responses in the standardised question format (N = 118 (valid)), their more detailed answers to the open-ended questions were more frequently coded as self-made decisions, and, as depicted in the Sankey plot in Table C2 in the appendix, especially as

Table 2

Comparison of rating and assessment for moving decision

	Self-decided (%)	Made together with others (%)	Others made the decision (%)	*N*
Decision to move				
... Rating	36	43	21	118
Rating (valid)	42	37	21	52
... Assessment	54	19	28	52
Difference (assessment-rating)	11	–18	7	52
Correlation of within person answers				
Cramér's V		0.565***		
Lambda		0.418***		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

the decision of others. Consequently, the narrative answers suggest less often a shared decision and more often an either/or allocation of decisions. On the one hand, narratives indicate a greater, on the other hand, a lower self-influence by the older people on the decision to move into the residential nursing home. Although we lack data to validate either interpretation, we propose the following explanations: self-perception or biased memory, and the impact of cognitive decline. We will now examine each difference between the frequency of the answer and the corresponding code:

– "Self-made": It is possible that older people perceive the relocation as something imposed upon them, while still acknowledging their involvement in a longer open-ended answer.
– "Made together with others": It may be more challenging, especially for those with declining cognitive skills, to recount stories with shared activity, perhaps due to difficulties in aligning perspectives.
– "Others made the decision": A desire for positive self-presentation, particularly when expressing complaints about the home or a negative self-rating of life quality, may lead to externalising responsibility. If the decision was made by others, it becomes more plausible to express dissatisfaction with the current situation.

Nonetheless, there are statistically significant moderate correlations between the qualitative and quantitative data (Cramér's V: 0.565, $p < 0.005$; Lambda: 0.418, $p < 0.005$), indicating at least a certain degree of consistency in the responses. We also consider the possibility that there might have been a higher consistency in the responses if the closed-ended question asked had been less complex, complying with the standard for avoiding double negatives. However, we do not consider the effects to be particularly serious. On the one hand, the interviewers could have clarified the questions in the interview, which had an interactional, dialogic character. On the other hand, the open-ended question had already established a basis for understanding this question, and the answer categories provided for the closed-ended question were straightforward: "I made the decision on my own." "I made the decision together with others." "Others made the decision for me."

4.3 Cognitive Decline

Table 3 presents the ratings and assessments of cognitive decline using the 7-point Likert scale based on the Global Deterioration Scale (GDS).

In this section, we compare two third-party ex-post evaluations of the interviewed residents' cognitive decline using the Global Deterioration Scale (GDS): the rating provided by the interviewer immediately after the interview ($N = 113$ (valid)) and the transcript assessment made by the coder ($N = 74$ (valid)). Descriptive results show broadly similar mean and standard deviations of GDS ratings and assessments. However, the slightly higher standard deviation in the full sample suggests that successful participation, particularly in the narrative portion of the interview, is associated with GDS ratings.

The third-party ex-post ratings exhibit an insignificant (but on the edge of significance) weak, positive rank correlation (Kendall-Tau-b: 0.182, $p = 0.100$), but a significant, positive and moderate linear correlation (Pearson's r: 0.441, $p < 0.005$). When simplifying the data into dichotomous variables (1 vs. 2+), the Phi correlation (0.462, $p < 0.01$) remains significant, but again only moderate. This suggests that precise assessments of cognitive decline based on transcripts alone, without access to the original interview interaction, is challenging.

Overall, our data indicates that cognitive decline is generally difficult to measure, and we assumed that challenges increase with greater cognitive limitations. We therefore examined the association of the initial interviewer rating and the assessment of the coder. This reveals a medium, positive, and significant correlation (Kendall-Tau-b: 0.432, $p < 0.001$; Pearson's r: 0.396, $p < 0.001$). The disparity between the interviewers' immediate ratings and the transcript-based assessments increases as cognitive decline worsens.

In nearly 50% of cases, interviewers and coders evaluate the cognitive conditions of interviewees equally, and in about 78% of cases the difference between their evaluations is no more than one point on the Likert scale. A closer look at these discrepancies (see Table C4 in online appendix C) reveals that coders tend to assess cognitive decline more

Table 3

Comparison of rating and assessment for GDS measure

	Mean	SD	*N*
GDS (rating)	1.86	1.308	113
GDS (assessment)	1.70	1.131	74
GDS (assessment-rating)	–0.09	–	–
GDS (rating)—valid	1.75	1.172	67
GDS (assessment)—valid	1.64	1.083	67
Absolute difference GDS (assessment-rating)—valid	0.11	–	–
Correlation of within person answers			
Kendall-tau-b		0.182	
Pearson's r		0.441***	
Phi (1(+2) vs. 2+(3+))		0.462*/0.574	
Correlation of GDS (rating) and absolute difference of rating and assessment			
Kendall-Tau-b		0.432***	
Pearson's r		0.396***	
Assessment equals rating		48%	
Difference assessment rating max+–1		78%	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

severely than interviewers. This is also illustrated by the Sankey plot in Table C3 in the Appendix. We hypothesise that this may be due to the interview context. During the face-to-face interaction interviewer and interviewees most likely collaborate to ensure smooth communication. This dynamic might result in interviewers underestimating the cognitive decline or providing a more positive rating, for instance, either due to the evaluation of the situation as a whole, taking the multimodal interaction and the setting into account, or out of a sense of loyalty.

Conversely, transcripts instead of audio or even video recordings not fully capture the nuances of the interview situation. Social interactions at the beginning and end as well as non-verbal cues are missed, leading to an incomplete picture. Additionally, errors in grammar and structure generally occur and become apparent in written form, whereas in conversation, these and other mistakes are often overlooked. Though, when individuals are suspected of cognitive impairment, natural occurrences such as pauses or word-finding difficulties may be over-interpreted and taken as confirmation of impairment.

5 Discussion

The increasing number of older people and the growing demand for long-term care in residential nursing homes make institutionalised older adults an important yet often overlooked research group. As the proportion of individuals requiring care rises with age—reaching nearly 80% for those over 90 years old (Destatis 2024c)—excluding them from research skews findings and assumptions about older people. Although surveys like SHARE and DEAS in Germany focus on the older population, they often omit institutionalised individuals due to the perception that they are difficult to interview. This exclusion is problematic. First, it prevents acknowledging important aspects of later life, and of long-term care, such as quality of life. Second, from a methodological point of view, it carries the risk of statistical errors, like the misinterpretation of reasons for living in a long-term care facility, where health status is often the core reason rather than age alone. Thus, it is not only important to include older (institutionalised) people, but also to ensure that research methods are appropriately adapted to their circumstances and the heterogeneity of this population. This is reflected in recent efforts to include institutionalised older adults in research aimed at avoiding undercoverage issues. It is thus recognised that research with older people living in institutions is generally possible, but requires more diligent planning, and specifically fine-tuned methods and adjusted tools; methods not only need to be appropriate for the research question, which is always the case (Dillman et al. 2014; Flick 2020), but rather to take the interview situation and interviewees' physical and cognitive capacities into account (cf. Sect. 2). To put it differently, inappropriate instruments, untrained interviewers, ineffective interactions between researcher and respondent, or the interview situation itself lead to non-participation, dropouts or item

non-response. Moreover, collecting valid and reliable data from people living in institutional settings requires specific strategies.

Our exploratory research contributes to this goal by offering insights relevant for larger-scale studies and providing useful lessons concerning sampling, response rates, and consistency testing. These insights are especially valuable in mixed methods designs, where the combination of qualitative and quantitative approaches allows for richer data and a more complete understanding of the population being studied (Maas et al., 2002), as well as for some kind of flexibility and spontaneous ad hoc and on-site adaptation of using instruments. We gave interviewees the opportunity to provide narrative responses in addition to answering closed-ended questions by integrating open-ended questions in an otherwise standardised questionnaire. This mixed method design allowed us to generate both quantitative and qualitative data from the same interviewees at the same time on the same topics, enabling a direct comparison and therefore addressing consistency.

Nevertheless, our study has four central limitations. First, we lack a control group of non-institutionalised older people or the general population. Thus, we cannot say whether inconsistencies are more prevalent within this "age group" or among institutionalised compared to non-institutionalised older people. However, because we rely on within-person comparisons, we can at least shed some light on whether different methods of data collection yield similar results. This suggests that a carefully designed and delivered quantitative survey may be a valid approach when carrying out research with this population also in large-scale studies.

Second, we focused on the consistency of only three exemplary measures, which might raise questions about the generalisability of our findings. Nonetheless, we believe these measures—covering subjective assessments, factual information, and interviewer ratings—are good examples of different types of data, illustrating the usefulness of our strategy for analysing the consistency of responses.

Third, due to the limited funding, the qualitative responses were coded by a single coder only. Despite being checked by a supervisor, we cannot address intercoder reliability. However, we believe the coding process was reliable because it was kept rather simple. Additionally, it seems necessary to control for order effects of answers given and coded as there was, as mentioned, some kind of divergence from the intended structure of questions in the concrete interview situation. Thus, it might be wise to check whether the consistency is higher when interviewees narrated about the respective topics after the closed-ended questions were asked and the answer options were presented.

Fourth, not all interviews were transcribed, particularly those with little narrative data. Nonetheless, by consistently reporting valid interview values, we believe that any bias introduced by this selectivity is minimal.

6 Conclusion

With the study presented here, we contribute to two methodological debates: first, the consistency of responses in mixed method designs, and second, the possibility to research older institutionalised living people. To assess this, we examined the consistency of answers to closed-ended and open-ended questions on two interview topics (*quality of life, decision to move*), and one overall evaluation of their respective cognitive decline. Our findings indicate an acceptable level of internal validity in mixed method designs, even in this group of so-called "hard to research people". Additionally, we present one possibility to address and evaluate the quality and validity of mixed methods research.

Our experiences beyond the central aim of this research papers point to the problem of access to those institutions for survey researchers and practitioners. Different sampling strategies are needed for institutions, in particular for institutions caring for older people (maybe also for other marginalised and as vulnerable perceived groups, such as people with disabilities), as well as for interviewees within these institutions. These strategies differ from those established for the general population because of the role of gatekeepers operating on different levels. Our plans to apply theoretical sampling for the institutions and then random sampling for individuals within the institutions turned out to be not feasible. While this aspect is of central importance in (larger) surveys with the aim to include institutionalised older people, this was acceptable to us as we did not aim to compare non-institutionalised and institutionalised living older people. Moreover, despite its small scale and limited representativeness, our study offers valuable recommendations for larger-scale surveys. We conclude that it is feasible to conduct research with older adults in residential care with results that are generally comparable across different interview techniques, and we encourage future research to include them in study designs.

In conclusion, our study addresses two increasingly relevant methodological debates: the efficacy of mixed method designs and the critical need to include institutionalised older adults in research. Our findings indicate an acceptable level of internal validity in mixed method designs, even in this challenging-to-research population. We highlight the importance of adaptability in research methodologies to meet the unique needs of institutionalised older adults. However, as this is only a preliminary exploratory attempt, further investigations are essential—particularly given our study's limitations. The absence of a control group constrains comparison, but our within-person anal-

yses provide insights into the comparability of data collection methods and enhances the use of secondary data also for qualitative research. Moving forward, researchers should prioritise rigorous designs and a deeper understanding of the populations being studied to enhance data quality and consistency in mixed methods research.

Acknowledgements We thank our student assistant Pauline Post for coding the interviews. We thank the participants of the session "Surveying Older People in Times of Crisis" at ESRA 2023 Milan. We thank the editor and reviewers for helpful comments to previous versions of this article.

Funding While this paper was written without any financial funding, we gratefully acknowledge the support from the ForGenderCare research network for the described project. This network examined the relationship between gender and care both theoretically and empirically from an interdisciplinary perspective. The network was funded by the State Ministry of Education and Culture, Science and the Arts and is part of the Bavarian Research Alliance BayFor between 2015 and 2019.

Data availability We cannot provide the transcripts of the qualitative interviews. First, they contain too much personal information about our interviewees which cannot be anonymised or deleted without distorting the meaning of the interviews. Second, the interviewees have not given their consent to secondary use of the qualitative interview data. However, we can provide the quantitative dataset used for the analyses upon request. Replication files are provided.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

Ethical standards At the time of planning and conducting this research, there was no Institutional Review Bord or Ethics committee at our faculty to approve this kind of research.

Author Contribution Both authors contributed equally to the work.

AI Use Disclosure During the preparation of the initial version of this paper, the authors used Chat Generative Pre-Trained Transformer (ChatGPT; OpenAI, San Francisco, CA, USA) to enhance readability and language, aiding in formulating and structuring content. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Akremi, L., & Zanker, D. (2023). Mixing standardized administrative data and survey data with qualitative content analysis in longitudinal designs: perceptions of justified pensions and related life courses. *Social Forum Qualitative Sozialforschung Forum: Qualitative Research*. https://doi.org/10.17169/fqs-24.1.4003.

Arnon, S., & Reichel, N. (2009). Closed and open-ended question tools in a telephone survey about the good teacher. *Journal of Mixed Methods Research*, *3*(2), 172–196. https://doi.org/10.1177/1558689808331036.

Barbosa Neves, B., & Baecker, R. (2022). Mixing methods and sciences: a longitudinal cross-disciplinary mixed methods study on technology to address social isolation and loneliness in later life. *Journal of Mixed Methods Research*, *16*(1), 88–113. https://doi.org/10.1177/1558689820977646.

Benítez, I., van de Vijver, F., & Padilla, J.L. (2022). A mixed methods approach to the analysis of bias in cross-cultural studies. *Sociological Methods & Research*, *51*(1), 237–270. https://doi.org/10.1177/0049124119852390.

Bonett, D.G., & Wright, T.A. (2000). Sample size requirements for estimating Pearson, Kendall and spearman correlations. *Psychometrika*, *65*, 23–28. https://doi.org/10.1007/BF02294183.

Brod, M., Stewart, A.L., Sands, L., & Walton, P. (1999). Conceptualization and measurement of quality of life in dementia: the dementia quality of life instrument (DQoL). *The Gerontologist*, *39*(1), 25–35. https://doi.org/10.1093/geront/39.1.25.

Bryman, A., Becker, S., & Sempik, J. (2008). Quality criteria for quantitative, qualitative and mixed methods research: a view from social policy. *International Journal of Social Research Methodology*, *11*(4), 261–276. https://doi.org/10.1080/13645570701401644.

Collins, K.M.T., Onwuegbuzie, A.J., & Jiao, Q.G. (2007). A mixed methods investigation of mixed methods sampling designs in social and health science researc. *Journal of Mixed Methods Research*, *1*(3), 267–294. https://doi.org/10.1177/1558689807299526.

Destatis (2024a). *Zahl der Pflegebedürftigen über 50 Jahre nach Art der Versorgung, Tab. 22421-0001 Pflegebedürftige: Deutschland, Stichtag, Geschlecht, Altersgruppen, Art der Versorgung von Pflegebedürftigen*. DeStatis. https://www-genesis.destatis.de; own calculations

Destatis (2024b). *Frauenanteil von Pflegebedürftigen über 50 Jahre, Tab. 22412-0004: Pflegebedürftige (stationär): Deutschland, Stichtag, Geschlecht, Altersgruppen*. DeStatis. https://www-genesis.destatis.de; own calculations

Destatis (2024c). *Anteil Personen mit Pflegegrad über 90 Jahre, 22421-0003: Pflegequote: Deutschland, Stichtag, Geschlecht, Altersgruppen*. DeStatis. https://www-genesis.destatis.de; own calculations

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2014). *Internet, phone, mail, and mixed-mode surveys. The*

tailored design method (4th edn.). New York: Wiley. https://doi.org/10.1002/9781394260645.

Dimbath, O., Ernst-Heidenreich, M., & Roche, M. (2018). Praxis und Theorie des Theoretical Sampling. Methodologische Überlegungen zum Verfahren einer verlaufsorientierten Fallauswahl. *FQS – Forum: Qualitative Sozialforschung Social Research*. https://doi.org/10.17169/fqs-19.3.2810.

Döbler, M.-K. (2021). Different by culture. Toward an analysis of age(ing) in care homes for older people. *GeroPsych*, *35*(2), 104–114. https://doi.org/10.1024/1662-9647/a000280.

Eickelmann, J., & Burzan, N. (2023). Challenges of multimethod and mixed methods designs in museum research. *Social Forum Qualitative Sozialforschung Forum: Qualitative Research*. https://doi.org/10.17169/fqs-24.1.3988.

Fisk, M., & Wigley, V. (2000). Accessing and interviewing the oldest old in care homes. *Quality in Ageing and Older Adults*, *1*(1), 27–33. https://doi.org/10.1108/14717794200000005.

Flick, U. (2020). *Introducing research methodology. Thinking your way through your research project*. Thousand Oaks: SAGE.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6.

Groves, R.M., & Couper, M.P. (1998). *Nonresponse in household interview surveys*. New York: Wiley. https://doi.org/10.1002/9781118490082.

Hall, S., Longhurst, S., & Higginson, I.J. (2009). Challenges to conducting research with older people living in nursing homes. *BMC Geriatrics*. https://doi.org/10.1186/1471-2318-9-38.

Hammersley, M. (2023). Are there assessment criteria for qualitative findings? A challenge facing mixed methods research. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*. https://doi.org/10.17169/fqs-24.1.3935.

Hathaway, R. (1995). Assumptions underlying quantitative and qualitative research: Implications for institutional research. *Research in Higher Education*, *36*(5), 535–562. https://doi.org/10.1007/BF02208830.

Hesse-Biber, S. (2012). Feminist approaches to triangulation. *Journal of Mixed Methods Research*, *6*(2), 137–146. https://doi.org/10.1177/1558689812437184.

Howe, K.R. (2012). Mixed methods, triangulation, and causal explanatio. *Journal of Mixed Methods Research*, *6*(2), 89–96. https://doi.org/10.1177/1558689812437187.

IBM Corp (2023). *IBM SPSS Statistics for Windows, Version 29.0.2.0*. Armonk: IBM Corp VERBI Software.

Isaksson, U., Santamäki-Fischer, R., Nygren, B., Lundman, B., & Åström, S. (2007). Supporting the very old when completing a questionnaire. *Research on Aging*, *29*(6), 576–589. https://doi.org/10.1177/0164027507305924.

Kaspar, R., Brijoux, T., Albrecht, A., Zimmermann, J., Wenner, J., Fey, J., Reissmann, M., Wagner, M., & Zank, S. (2023). Challenges and benefits of including the institutionalized, cognitively impaired and unable to respond in a representative survey of the very old. *Survey Research Methods*, *17*(2), 111–129. https://doi.org/10.18148/srm/2023.v17i2.8008.

Kelfve, S., Thorslund, M., & Lennartsson, C. (2013). Sampling and non-response bias on health-outcomes in surveys of the oldest old. *European Journal of Ageing*, *10*(3), 237–245. https://doi.org/10.1007/s10433-013-0275-7.

Kelle, U., & Niggemann, C. (2002). „Weil ich doch vor zwei Jahren schon einmal verhört worden bin …" – Methodische Probleme bei der Befragung von Heimbewohnern. In A. Motel-Klingebiel & U. Kelle (Eds.), *Perspektiven der empirischen Alter(n)ssoziologie* (pp. 99–131). Wiesbaden: VS. https://doi.org/10.1007/978-3-663-11027-9_5.

Kelle, U., & Reith, F. (2023). Strangers in paradigms!? Alternatives to paradigm-bound methodology and methodological confessionalism. *Social Forum Qualitative Sozialforschung Forum: Qualitative Research*. https://doi.org/10.17169/fqs-24.1.4015.

Kelle, U., Metje, B., & Niggemann, C. (2014). Datenerhebung in totalen Institutionen als Forschungsgegenstand einer kritischen gerontologischen Sozialforschung. In A. Amann & F. Kolland (Eds.), *Das erzwungene Paradies des Alters?* (pp. 175–206). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-02306-5_8.

Kim, Y. (2019). New opportunities for sociological research. *Journal of Asian Sociology*, *48*(3), 343–358. https://doi.org/10.21588/jas/2019.48.3.003.

Knappertsbusch, F. (2023). From paradigm wars to peaceful coexistence? A sociological perspective on the qualitative-quantitative-divide and future directions for mixed methods methodology. *Social Forum Qualitative Sozialforschung Forum: Qualitative Research*. https://doi.org/10.17169/fqs-24.1.3966.

Knäuper, B., Carrière, K., Chamandy, M., Xu, Z., Schwarz, N., & Rosen, N.O. (2016). How aging affects self-reports. *European Journal of Ageing*, *13*(2),

185–193. https://doi.org/10.1007/s10433-016-0369-0.

Krause, N. (2002). A comprehensive strategy for developing closed-ended survey items for use in studies of older adults. *Journal of Gerontology*, *57B*, 263–274. https://doi.org/10.1093/geronb/57.5.s263.

Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213–236. https://doi.org/10.1002/acp.2350050305.

Kühn, K., & Porst, R. (1999). *Befragung alter und sehr alter Menschen: Besonderheiten, Schwierigkeiten und methodische Konsequenzen; ein Literaturbericht*. ZUMA-Arbeitsbericht 1999/3. Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-200424

Kydd, A., Fleming, A., Gardner, S., & Hafford, T.-L. (2018). Ageism in the third age. In L. Ayalon & C. Tesch-Römer (Eds.), *Contemporary perspectives on ageism* (pp. 115–130). Cham: Springer. https://doi.org/10.1007/978-3-319-73820-8_8.

Lang, G., Löger, B., & Amann, A. (2007). Well-being in the nursing home—a methodological approach towards the quality of life. *Journal of Public Health*, *15*(2), 109–120. https://doi.org/10.1007/s10389-006-0082-0.

Lundström, F. (1987). Research design in total institutions: problems, pitfalls, and possible solutions. *Quality and Quantity*, *21*, 209–218. https://doi.org/10.1007/BF00134520.

Maas, M.L., Kelley, L.S., Park, M., & Specht, J.P. (2002). Issues in conducting research in nursing homes. *Western Journal of Nursing Research*, *24*(4), 373–389. https://doi.org/10.1177/01945902024004006.

MAXQDA (2022). Berlin: VERBI Software. maxqda.com

McKim, C.A. (2017). The value of mixed methods research. *Journal of Mixed Methods Research*, *11*(2), 202–222. https://doi.org/10.1177/1558689815607096.

Morgan, D.L. (2019). Commentary—after triangulation, what next? *Journal of Mixed Methods Research*, *13*(1), 6–11. https://doi.org/10.1177/1558689818780596.

Motel-Klingebiel, A., Klaus, D., & Simonson, J. (2019). Befragungen von älteren und alten Menschen. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 935–942). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-21308-4_66.

Naqvi, A. (2025). *"SANKEY: Stata module for Sankey diagrams (Version 1.9)," Statistical Software Components S459154*. Boston College Department of Economics. https://github.com/asjadnaqvi/stata-sankey

O'Halloran, K.L., Tan, S., Pham, D.-S., Bateman, J., & Vande Moere, A. (2018). A digital mixed methods research design: integrating multimodal analysis with data mining and information visualization for big data analytics. *Journal of Mixed Methods Research*, *12*(1), 11–30. https://doi.org/10.1177/1558689816651015.

Reisberg, B., Ferris, S.H., de Leon, M.J., & Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *The American Journal of Psychiatry*, *139*(9), 1136–1139. https://doi.org/10.1176/ajp.139.9.1136.

Reisberg, B., Jamil, I.A., Khan, S., Monteiro, I., Torossian, C., Ferris, S., Sabbagh, M., Gauthier, S., Auer, S., Shulman, M.B., Kluger, A., Franssen, E., & Wegiel, J. (2011). Staging dementia. In M.T. Abou-Saleh, C.L.E. Katona & A. Kumar (Eds.), *Principles and practice of geriatric psychiatry* (pp. 162–169). Chichester: Wiley. https://doi.org/10.1002/9780470669600.ch31.

Reuband, K.-H. (2006). Postalische Befragungen alter Menschen: Kooperationsverhalten, Beantwortungsstrategien und Qualität der Antworten. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, *59*, 100–127. https://www.ssoar.info/ssoar/handle/document/19836.

Schanze, J.-L. (2022). Hard-to-survey and negligible? The institutionalized population in Europe. *Survey Research Methods*, *17*(1), 91–109. https://doi.org/10.18148/srm/2023.v17i1.7830.

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.

Schanze, J.-L., & Levinson, M. (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*. https://seriss.eu/resources/deliverables/ Deliverable 2.17 of the SERISS Project funded under the European Union's Horizon 2020 Research and Innovation programme GA No: 654221″.

Slonim-Nevo, V., & Nevo, I. (2009). Conflicting findings in mixed methods research. *Journal of Mixed Methods Research*, *3*(2), 109–128. https://doi.org/10.1177/1558689808330621.

Smith, S.C., Hendriks, A.A.J., Cano, S.J., & Black, N. (2020). Proxy reporting of health-related quality of life for people with dementia: a psychometric solu-

tion. *Health and Quality of Life Outcomes*. https://doi.org/10.1186/s12955-020-01396-y.

StataCorp (2023). Stata: Release 18. Statistical Software. College Station, TX: StataCorp LLC.

Statistisches Landesamt Bayern (2014). *Verzeichnis Einrichtungen für ältere Menschen in Bayern 2014 [Directory of facilities for older people in Bavaria 2014]*. https://www.statistik.bayern.de/mam/produkte/veroffentlichungen/statistischeberichte/k8100c201451_848.pdf

Tanner, P. (2023). Nested sampling in sequential mixed methods research designs: comparing recipe and result. *Social Forum Qualitative Sozialforschung Forum: Qualitative Research*. https://doi.org/10.17169/fqs-24.1.4004.

Tyler, D., Shield, R., Rosenthal, M., Miller, S., Wetle, T., & Clark, M. (2010). How valid are the responses to nursing home survey questions? Some issues and concerns. *Gerontologist*, *51*(2), 201–211. https://doi.org/10.1093/geront/gnq095.

Wagner, M., Kuppler, M., Rietz, C., & Kaspar, R. (2019). Non-response in surveys of very old people. *European Journal of Ageing*, *16*(2), 249–258. https://doi.org/10.1007/s10433-018-0488-x.

Wenger, C. (2003). Interviewing older people. In J. Holstein & J. Gubrium (Eds.), *Inside interviewing: new lenses, new concerns* (pp. 111–128). Thousand Oaks, London, New Delhi: SAGE. https://doi.org/10.4135/9781412973588.n17.

Zermansky, A.G., Alldred, D.P., Petty, D.R., & Raynor, D.K. (2007). Striving to Recruit: the Difficulties of Conducting Clinical Research on Elderly Care Home Residents. *Journal of the Royal Society of Medicine*, *100*(6), 258–261. https://doi.org/10.1258/jrsm.100.6.258.