

Challenges and Benefits of Including Institutionalized, Cognitively Impaired, and Unable-to-Respond Individuals in a Representative Survey of the Very Old

Roman Kaspar¹, Thomas Brijoux¹, Andrea Albrecht¹, Jaroslava Zimmermann¹, Judith Wenner^{1,2}, Jonas Fey¹, Marcella Reissmann¹, Michael Wagner^{1,3}, and Susanne Zank^{1,2}

¹University of Cologne, Cologne Center for Ethics, Rights, Economics, and Social Sciences of Health

²University of Cologne, Faculty of Human Sciences

³University of Cologne, Faculty of Management, Economics and Social Sciences

Institutionalization, cognitive impairment, and the inability to conduct an interview due to health impairment are among the top exclusion criteria for most large-scale social and aging surveys. Reservations about targeting vulnerable groups result from economic or legal restrictions of recruitment and concerns regarding research ethics or the validity of the data obtained. However, failure to include these individuals may lead to substantial bias. Metadata showed that privileged data access and checks against nursing home repositories prevented the undercoverage of institutionalized individuals. Measures to include difficult-to-survey groups led to a marked increase in response rates. Individuals with health impairments substantially contributed to the representativity of the sample. Nonresponse bias was cut in half when compared with a less inclusive study protocol. From a Total Survey Error perspective, reductions in nonresponse bias, low item-nonresponse, and evidence of measurement invariance across self-reports and proxy reports for key outcome variables show significant benefits of including difficult-to-survey groups in estimating characteristics of this population.

Keywords: quality of life; proxy report; total survey error; nonresponse bias; measurement invariance

1 Introduction

Very old individuals have been described as a difficult-to-survey part of the population (Feskens, 2009). Institutionalization, cognitive impairment, and lack of communicative ability to conduct an interview or questionnaire are among the top reasons for exclusion or nonresponse of the oldest old¹ in most large-scale social surveys (Sakshaug, 2022; Schanze, 2017) and many aging studies (e.g., Davies et al., 2010; Deeg et al., 2002; Schiel et al., 2021). Much of the reservation against targeting more vulnerable clients is well substantiated by economic or legal restrictions and concerns about research ethics or the validity of data obtained. For example, community registers in Germany are prohibited by law from including addresses of nursing home residents (NHR) in samples drawn for commercial survey institutes (BMG §52). Davies et al. (2010) report in the Newcastle 85+ study that increased time was necessary to engage with

family and gatekeepers. Lower contact rates have consistently been reported for NHR (Gaertner, Koschollek, et al., 2019; Wagner et al., 2019). Based on data from the Survey of Health, Ageing and Retirement in Europe SHARE, Schanze (2021) found the data quality was lower in individuals with lower socioeconomic and health status. Simultaneously, failure to include those with more prominent health issues or in nonprivate dwellings may lead to a substantial underestimation of, for example, dependencies in the activities of daily living (Kelfve et al., 2013; Schanze & Zins, 2019) or health inequalities (Kelfve, 2019). Vulnerable individuals represent a non-negligible portion of the population aged 80 years and older. Hence, their inclusion is essential in surveying the older population. The current lack of experience with including the most vulnerable in survey research is not limited to questions of recruitment and the mode of the interview but also pertains to identifying valid ways to integrate the collected information into the analysis and subsequent reporting. This study exemplifies how including difficult-to-survey groups of older adults could reduce nonresponse bias and

Contact information: Roman Kaspar, Cologne Center for Ethics, Rights, Economics, and Social Sciences of Health, University of Cologne, Albertus-Magnus-Platz, 50931 Cologne, Germany. (E-mail: roman.kaspar@uni-koeln.de)

¹There is no consensus definition of very old age or the oldest old. In this study, we use the cut-off of 80 years or older, also used in the Survey of Health, Ageing and Retirement in Europe SHARE (Andersen-Ranberg et al., 2005).

provide better population estimates from survey responses.

2 Elements of Survey Quality

The Total Survey Error (TSE) framework gradually arose from the work of scholars wanting to optimize survey design concerning different threats to survey quality (for an overview, see Groves & Lyberg, 2010). An early example is Groves (1987), who jointly considered both nonobservation error (i.e., coverage, sampling, and nonresponse errors) and observation (i.e., measurement) error. In this study, we consider the quality of a survey as a function of the coverage of the targeted population in the sampling frame, the definition of inclusion probabilities for individuals drawn from the sampling frame, the share and selectivity of respondents, and the degree to which the response measures the true value of respondents' characteristics. Specific challenges have been reported regarding most, if not all, of these steps when studying very old age (Schanze & Zins, 2019; Schanze, 2019).

2.1 Sampling Errors

Concerning *adequate coverage of the population of the very old*, a first issue is conceptual and arises from the need to a priori define very old age. While various definitions of very old age have been suggested in different research fields (Degnen, 2007; Foo et al., 2019), very limited official register data exist to define the population of the very old other than by chronological age (e.g., mostly 80 years or older). Even a simple demographic definition (i.e., more than half of one's birth cohort has passed) is subject to historical change and jeopardized by gender-specific mortality rates. Second, community residential register data in this age segment have a higher probability of being outdated (e.g., due to the delayed implementation of death records or relocation), incomplete (e.g., retaining the original address after relocating to a nursing home), or inaccessible. Regarding the latter, legal restrictions may limit information access on particularly vulnerable parts of the population, such as NHR. In Germany, data security laws afford statistic bureaus to mask sociostructural information from residential registries to disable potential reidentification, particularly in small communities with only a few very old individuals. Scherpenzeel et al. (2017) have described differences in sampling frames for social surveys across European countries.

Concerns have also been raised regarding *the definition of inclusion probabilities for individuals drawn from the sampling frame*. Because there are many more women in this age segment today than men and the number of individuals declines rapidly with increasing chronological age, simple random sample data will be limited in its potential to allow for reliable comparison of subpopulations in very old age. For example, oversampling of males is required to conduct meaningful subgroup analyses across gender. However, disproportionate sampling schemes compromise sampling effi-

ciency, because they need to be accounted for in the analysis by using design weights (Aust & Gilberg, 2016). Moreover, they require detailed information on the (age) structure of the targeted population to compute selection probabilities adequately. First, and related to problems of retaining detailed information on the age structure, official data on the number of inhabitants above the age of 80 may be difficult to obtain, particularly in small communities, depending on regional data protection laws or default register procedures. For example, standard reporting, from the state administration to the Federal Bureau of Statistics in Germany, excludes more finegrained age or birth cohorts beyond age 80. Second, the relative scarcity of the oldest old at the level of the primary sampling unit (e.g., communities) poses problems for delivering the required number of oldest-old individuals. For any expected total number of participants, these circumstances effectively increase the number of necessary primary sampling units (PSUs) (hence, reducing the design effect), result in synthetic sampling points or reduce the gross sample because communities may not be able to deliver the required number of oldest-old individuals. They also effectively limit the disproportionality that can be accomplished in a sampling design to safeguard subgroup comparison. Specific and sequential mixed-mode designs have been suggested to survey very old adults (Gaertner, Lüdtke, et al., 2019; Gaertner, Koschollek, et al., 2019). Additionally, separate sampling frames for subpopulations, such as nursing home registers, have been considered (Schneekloth & Müller, 1998). Integrating alternative data sources into multi-frame sampling has been suggested (Sand, 2014). However, such schemes may share the problems reported above for residential register data and challenges related to selecting random samples in dynamic and unique entities, such as nursing homes (preselection, high intraclass correlation).

2.2 Nonresponse and Measurement Errors

Specific challenges have been reported regarding *the number of very old individuals willing to and capable of participating* in survey research and whether this group is *representative of the general population of very old adults*.

First, aging survey response rates have decreased over the last decades. In the German Aging Survey, base sample participation dropped from 50.3% in 1996 to 27.1% in 2014 (Vogel et al., 2020). Mistrust of strangers and lack of interest in issues not affecting them were presumed reasons for the lower motivation of older adults to participate (Kühn et al., März 1999). Wagner et al. (2019) have identified specific dropout mechanisms for subgroups of the very old. More specifically, individuals in different settings (e.g., private vs. institutional) and with different levels of functioning drop out of the study at different points of the recruitment process. For example, low contact rates in NHR indicate difficulty in accessing some of these target persons. However, significantly

lower refusal rates have been found in NHR that could ultimately be contacted, corroborating the importance of gatekeepers for surveying vulnerable individuals (Davies et al., 2010; Hall et al., 2009).

Second, very old and oldest-old age is characterized not only by more years to live but also by a substantial proportion of time spent with illness and loss of independence (Tesch-Römer & Wahl, 2017). Even aging studies targeting very old adults sometimes fail to include some of the most vulnerable segments of this population, often for practicality or skepticism concerning the quality of the data obtained (Schanze, 2021). Exclusion criteria (e.g., unable to conduct an interview, nonprivate setting, or cognitive impairment) not only limit the generalizability of findings but also cause difficulty in defining eligibility and the computation of response rates with potential adverse effects of the initial overcoverage of the target population. For example, individuals who died between sampling and recruitment are often considered non-eligible cases in younger samples. In contrast, deceased target individuals may be considered eligible and nonrespondents in a population segment characterized by increased age-associated risk of mortality.

Finally, a large body of evidence exists that describes potential differences between age groups regarding *the quality of information obtained from very old and oldest-old adults in standard survey data collection procedures* and suggests an adaptation to procedures and materials (Isaksson et al., 2007). Some of these differences pertain to the specific historical background of the oldest old and differences in socialization, language, norms, education, or life experience (e.g., war) that may result in potential misfit with assessment procedures or concepts generally developed more recently with and for a younger audience (Chan et al., 1999). Examples of standard survey assessment where the current practice may not fully map onto the realities of very old individuals are the coding of educational or vocational backgrounds using the current ISCED standardization. Educational systems and access to higher education have changed over time, making it difficult to compare and classify the educational qualifications of today's oldest old. Similarly, assessing income or occupational prestige among mainly retired or institutionalized individuals is difficult and might ultimately lead to their exclusion from the analysis. Simultaneously, overcomplex assessments result in high item-nonresponse or information bias. Measures of social status are nonetheless indispensable for assessing social and health inequalities among oldest old individuals (Darin-Mattsson et al., 2017). Adapting standard instruments, e.g., inquiring about the last (former) occupation instead of the current employment status for coding social class, is a practical and straightforward solution (Grundy & Holt, 2001).

Other differences in responses to survey questions by very old individuals compared with younger age groups pertain

to developmental change. Research into measurement invariance across the lifespan has just begun acknowledging that changes in how survey questionnaires are understood and responded to may offer a unique window into understanding intra-individual developmental processes (e.g., response shift, see Edwards & Wirth, 2009; Kaspar et al., 2019). Evidence for a developmental change in late life has been reported for key characteristics such as cognition (Hülür et al., 2015), personality (Mueller et al., 2017), and values (Borg, 2019; Reissmann et al., 2021). Measures of functional health, such as activities of daily living (ADL) or instrumental activities of daily living (IADL) scales exhibit some degree of dependency on age and context (e.g., living at home vs. nursing homes) at the item level (Fleishman et al., 2002; LaPlante, 2010). However, the combined scales were rather stable over time, setting, and age, making ADL/IADL scales an important and comparable measure for functional health among the oldest old (Finlayson et al., 2005).

A third large group of concerns regarding the quality of survey responses obtained from very old individuals refers to compromised sensory, physical, or cognitive abilities more prevalent in this age segment (Schanze, 2021). On the most general level, challenges are reported regarding interview length, attention span, fatigue, or the need for breaks (Davies et al., 2010). Similarly, evidence shows that not all assessment modes are equally suited for very old adults, suggesting that face-to-face (f2f) interviews are the mode most responsive to challenges posed by sensory loss or limited cognitive capacity (Farmer & Macleod, 2011; Isaksson et al., 2007). Concerning obtaining retrospective biographical information, the potential combination of a very rich and long biography and memory impairment may result in inconsistent data (El Haj et al., 2015). This may even be true within reasonable timeframes such as "the last 12 months." These are extensively used in assessing activity, well-being, or social engagement for good reasons, including, but not limited to, expected seasonal variation, celebrations, or transient states such as short-term illness. Findings on the positive and negative affect schedule (PANAS) have consistently shown invariant measurement properties across different age groups and in very old individuals (Kercher, 1992; Mackinnon et al., 1999). However, a modified version of the PANAS that used dichotomized items reported low reliability of the scale in NHR with dementia (Gerritsen et al., 2007).

Nevertheless, the population of the very old includes a substantial number of individuals for whom adaptations of the f2f interview will not succeed in surpassing communication impairment due to substantial physical or cognitive decline. The group of individuals *unable to conduct the interview* (UCI) themselves for health reasons varies as a function of the expected interview burden (e.g., length and complexity). We argue that obtaining a proxy report for these cases may provide helpful information for many aspects of

the life of the targeted individual and offer a window to explore (the lack of) specific knowledge (e.g., introspection) or motivation (e.g., response style) of the proxy informant as a substantive-matter research question (Vazire, 2010). Maybe even more importantly, it allows scholars to test empirically to what extent such data could validly be integrated into population estimates.

2.3 Research Aims

This study weighs the advantages and disadvantages of including three difficult-to-survey subgroups (i.e., institutionalized, UCI, and cognitively impaired individuals) in a representative survey on quality of life (QoL) in very old age. More specifically, we first evaluated the success of sampling and recruitment measures taken to better address NHR, individuals with varying degrees of cognitive impairment, and information on those no longer able to conduct the 90-min-long interview themselves (i.e., UCI via proxy interviews). We expected an improvement in the absolute number of realized interviews and response rates relative to studies that adhere to the same inclusive definition of the target population but refrain from going the extra mile to address hard-to-survey subgroups during sampling and recruitment. Moreover, we expected increased representativity of the realized sample, thus reducing the risk of nonresponse bias relative to such survey protocols that fail to specifically address these difficult-to-survey subgroups of older adults. Next, we evaluated the extent to which integrating responses from these difficult-to-survey groups altered substantive-matter conclusions about socioeconomic resources, health resources, and well-being outcomes of the very old population. We expected that subject-matter responses for these groups might increase insecurity about population parameter estimates, particularly when options to establish measurement invariance (MI) or estimate response bias in subgroups (e.g., proxy interviews) were limited. However, from the viewpoint of the TSE framework, we assumed that the benefits of representing difficult-to-survey subgroups of very old adults in the sample (i.e., reduced nonresponse bias) could outweigh concerns about data quality (e.g., measurement bias from proxy interviews or inconsistent responses from cognitively impaired individuals) when including more vulnerable individuals.

3 Methods

3.1 Participants and Procedures

The data are from a study on QoL and well-being of very old adults conducted in Germany's most populous state, North Rhine-Westphalia (Wagner et al., 2018). A multistage sampling design was employed to define the sample. First, communities were drawn as PSUs based on the number of inhabitants aged 80 years or older (i.e., proportional-to-size selection). A total of 120 PSUs was selected, with large

communities contributing multiple PSUs. Next, community offices drew random samples of 400 individuals (i.e., secondary sampling units) per PSU. From the resulting sampling frame of more than 48,000 individuals, a disproportional gross sample of 8,040 individuals was drawn that should result in approximately 1,800 realized interviews. Persons 85 years and older and men were oversampled to allow for robust subgroup analyses according to a priori power analyses. Details about the computation of survey weights are provided in Appendix A. A total of 1,863 computer-assisted personal interviews (CAPI) were conducted at participants' homes to assess a wide array of individual QoL resources (e.g., economic, health) and subjective QoL outcomes (e.g., well-being). The study protocol also included objective testing of handgrip strength and mild cognitive impairment. The mean age of the realized sample at the time of the interview was 87.0 years (SD = 4.5; range: 80.1 to 102.9 years). The ethical board of the medical faculty at the University of Cologne approved this study (Protocol #: 17-169).

3.2 Major Design Decisions

The expected length of an interview is a critical determinant of study participation. Results from a pilot study (Brix et al., 2016) indicated that interviews with very old individuals should not exceed 90 min on average. Potential proxy interviews for UCI were not limited a priori to specific groups of informants (e.g., partners and children) or specific content (such as "facts" or easily observable characteristics of the target persons) to prevent over-exclusiveness and test the limits of data collection in this population. However, informants have been explicitly instructed to choose "refuse" or "do not know" categories for questions they felt uncomfortable or unable to answer. Additionally, all interviews conducted with individuals willing to participate and capable of understanding and answering the questions are included in the study, irrespective of their screening test results for mild cognitive impairment and suspected dementia.

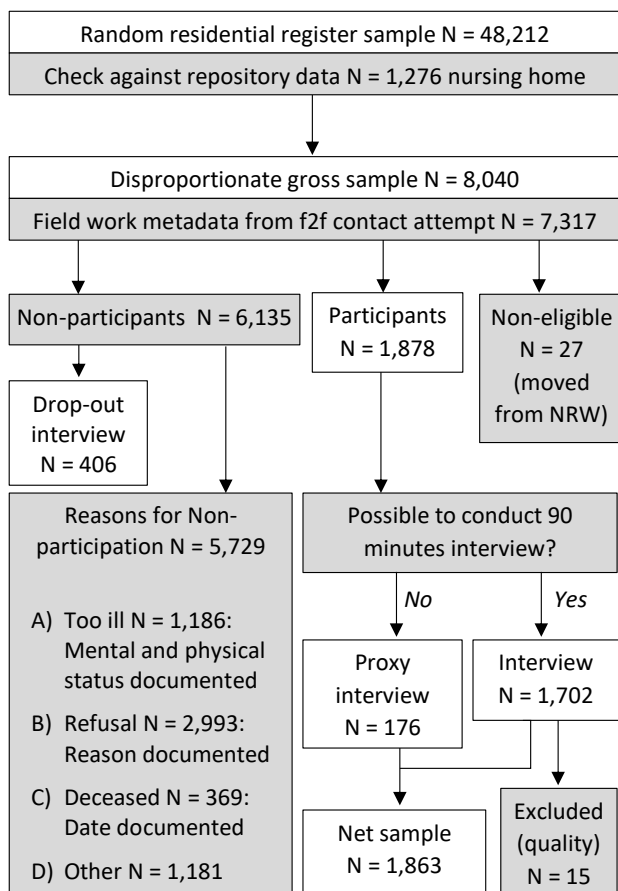
3.3 Fieldwork Metadata

During the initial steps of obtaining the study sample and contacting potential respondents, metadata (e.g., communicative abilities, health status, living conditions) were generated that identified difficult-to-survey individuals, leading to tailored approaches and inclusion strategies (Figure 1).

Detailed reasons for nonparticipation were documented for 2,993 individuals who actively refused to participate. Only in cases when targeted individuals were too ill to conduct the 90-minute interview themselves (i.e., UCI) was an attempt made to conduct a proxy interview, and specific reasons have been documented if no proxy interview could be realized ($N = 1,186$).

Figure 1

Flowchart of potential study participants, fieldwork meta-data, and interview data.



Institutionalization

Random residential register samples from 94 selected communities were received between May and September 2017. In the beginning, data were collected by a commercial survey institute. However, population registration offices were hesitant to provide data on NHR to a commercial survey institute because NHR had just been granted additional protection by novel data protection legislation at that time. In response, the academic staff obtained privileged access to register data directly from the population registration offices under the applicable data protection laws. Sample quality control included comparing the most recent official register data and screening for clear deviations from random sampling (e.g., demographic structure, name- or streetwise selection). A comprehensive repository of 1,276 addresses of nursing homes and care facilities in NRW was used to identify individuals in the samples living in an institution. The share of identified NHR was compared with the community-

level census data on the prevalence of institutionalization in the target population. If the data suggested that NHR were systematically excluded, a new and unrestricted random sample was requested.

Cognitive status

Extended documentation of mental or physical health conditions was sought in case the target person was unable to conduct the interview him-/herself due to health reasons (i.e., UCI).

Proxy interviews

Documentation of nonparticipation was also extended to include information on why no proxy interview could be conducted in UCI. More specifically, the lack of potential informants and the refusal of available informants were documented. In this study, we defined the difficult-to-survey subgroup of UCI as all individuals who were unable to conduct the interview him/herself due to health reasons and who therefore had to be included in the study via proxy interviews.

3.4 Data from Survey Interviews

Institutionalization

Respondents were asked to report whether they received formal or informal care. Additionally, interviewers rated the respondents' dwellings according to different categories of private, supported, or institutionalized housing. The time between drawing the register samples and interviewing was kept to a minimum (average 112 days, range 12 to 208 days). The community register information on nonprivate dwellings was validated before contacting potential respondents (see the section on fieldwork metadata). Nevertheless, some individuals initially classified as private-dwelling were finally interviewed in institutional settings. Details on the definition of the difficult-to-survey subgroup of very old individuals in nursing homes (NHR) in this study are reported in Appendix B.

Cognitive status

The DemTect has been developed as a brief screening tool for mild cognitive impairment and the early stages of dementia (Kalbe et al., 2004). It comprises subtests targeting immediate/delayed word recall, digit span, number transcoding, and verbal fluency. Favorable psychometric and diagnostic properties for identifying mild cognitive impairment (MCI) have been reported compared with alternative screening tools, such as the Mini-Mental State Examination (Kalbe et al., 2013).

Because most of the empirical evidence regarding the psychometric properties and clinical utility of the DemTect has been reported for clinical populations, a rigorous test of the tool for use in large-scale surveys in the general old age

population has been conducted. In a comprehensive feasibility study, the DemTect was implemented in CAPI interviews conducted in a random register sample of participants ($N = 291$) aged 80 years or above who were living privately and in long-term care facilities. All interviewers participated in a half-day test administration and scoring training. During the assessment, 49 participants were unable to answer the number transcoding task for reasons unrelated to cognitive problems (i.e., problems with reading/vision or writing 16.8%). In these cases, subtest scores were estimated based on the available partial responses in this subtest and performance in all other subtests to minimize the punishment for noncognitive impairment of test-takers. We used age-specific score transformation and derived an ordinal diagnostic group variable (i.e., age-adequate function, MCI, early dementia) based on the cut-off values suggested for persons aged 80 and over (Kessler et al., 2014). In the feasibility study, 26.8% of participants declined to take the test or some subtests, precluding the interpretation of their diagnostic group membership. The DemTect has been evaluated multiple times (Kalbe et al., 2004; Kessler et al., 2010; Perneczky, 2003), and a joint estimator using a Reitsma function (Doebler, 2020) yields a sensitivity of 94.1% and specificity of 91.4%. Based on the published age- and sex-specific prevalence rates (Doblhammer et al., 2012), 19.5% of the feasibility study sample can be expected to have dementia. A comparison with the DemTect classification yielded a positive predictive value of 72.5% and a negative predictive value of 98.5% of the screening.

In the survey reported here, an identical procedure was followed for the test administration and scoring in interviews with respondents. In the case of proxy interviews, cognitive status was reported with the Global Deterioration Scale (GDS, Reisberg et al., 1982) in seven stages, from 1 (*no cognitive impairment*) to 7 (*most severe*). More recently, Reisberg and colleagues aligned GDS stage 3 to correspond to a clinical presentation of MCI (Reisberg et al., 2011).

For the current analysis, all respondents whose cognitive function was labelled as MCI or dementia according to the diagnostic screening were considered members of the difficult-to-survey subgroup of very old adults with cognitive impairment.

Socioeconomic status

The International Socioeconomic Index of Occupational Status (ISEI-08, Ganzeboom et al., 1992; Ganzeboom & Treiman, 2003) was used to measure socioeconomic status (SES). The ISEI-08 is a metric measure (range 10–90) building mainly on the occupation of the respondents according to the International Standard Classification of Occupations (ISCO-08, Ganzeboom, 2010). The measure also considers the required educational qualification. Given that most participants were retired, we took the last occupation as a basis

for the ISEI-08. In case the targeted individual reported no previous occupation, the last occupation of the partner was used instead. ISEI-08 scores in this sample varied between 10 and 89.

Independence in activities of daily living

Adults' self-reported performance on the basic ADL (Katz et al., 1963) and IADL (Fillenbaum, 1988; Lawton & Brody, 1969) was used to measure everyday functioning. In this paper, we use five items representing ADL (e.g., getting dressed, walking) and seven items representing IADL (e.g., preparing meals, handling finances) with response options 0 (*impossible without help*), 1 (*some help needed*), or 2 (*no help needed*). The reliability of the ADL and IADL scales in the current sample was high (MacDonald's $\omega = 0.92$ and 0.93 , respectively).

Subjective well-being

The positive affect subscale of the short form of the PANAS (Kercher, 1992) was used as an indicator of subjective well-being (SWB). The frequency of positive affective states (e.g., "enthusiastic", "excited") across the past 12 months was reported on a 5-point response scale from 1 (*never*) to 5 (*very often*). Favorable psychometric properties of this very brief instrument have been reported for age-diverse and old-old samples (Hilleras et al., 1998; Kercher, 1992; Mackinnon et al., 1999). The scale consistency in our sample was satisfying (MacDonald's $\omega = 0.88$).

3.5 Plan of Analysis

We adopted the TSE framework to discuss advantages and disadvantages of including three groups of particularly difficult-to-survey individuals in a representative survey of QoL in the very old. First, we used sampling and fieldwork metadata (e.g., share of nursing home addresses, reasons for nonparticipation) to evaluate the success of tailored strategies to include NHR, individuals with (beginning) cognitive impairment, and UCI and to estimate their effect on representing the population 80 years or older. Here, based on random samples of individuals from 94 communities, we investigated the potential threat of undercoverage of the institutionalized population in the sampling frame due to legal restrictions on data access and the lack of information on nonprivate dwellings in the register data. Based on fieldwork metadata, we compared the reasons for the refusal of residents in private dwellings versus institutional settings and the reasons why no proxy interview could be conducted for most UCI. We then used response rates and representativity (R) indicators (Schouten et al., 2009) to estimate differences in nonresponse bias resulting from the exclusion of specific subgroups of respondents. The response rate was defined

based on the AAPOR definition of RR1 to include all eligible cases (The American Association for Public Opinion Research, 2016; Wagner et al., 2019). R-indicators estimate the representativity of a sample by predicting the nonresponse propensity by auxiliary variables known for the whole population (Schouten et al., 2009). Hence, they represent a measure of the deviation from weak representativity regarding theoretically meaningful threats to survey participation. The population R-indicator is defined as 1 minus two times the standard deviation of the response probability in the population ρ and ranges from 0 to 1, with higher values indicating higher representativity. Because response probabilities are typically unknown, they need to be estimated by regressing observed (non)responses on auxiliary variables. In this study, we used age, sex, community size, and NHR status as proxies for selection mechanisms (e.g., health and social status restrictions). The design-weighted (s_i/π_i) response-based R-indicator is defined by Schouten et al. (2009) as follows:

$$\hat{R}(\rho) = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{i=1}^N \frac{s_i}{\pi_i} (\hat{\rho}_i - \hat{\rho})^2} \quad . \quad (1)$$

We use the coefficient of variation of response propensities (CV) as a measure of the maximum potential nonresponse bias. The corresponding estimate acknowledges that nonresponse bias is a function of both the relative share of individuals participating in the study (i.e., response rate RR1) and the degree to which the realized sample is weakly representative of the target population:

$$CV(X) = \frac{1 - R(X)}{2RR1} \quad . \quad (2)$$

The CV represents a conservative worst-case estimate of nonresponse bias assuming that nonresponse correlates maximally with selected auxiliary variables X (De Heij et al., 2015).

In the second step, substantive-matter responses from the realized sample of 1,863 individuals (of which NHR: 211, screened cognitively impaired: 504, and UCI: 176) were considered to estimate potential bias using a multi-group approach to measurement invariance and mode effects. In this study, we restricted our focus to socioeconomic status, independence in ADL/IADL, and SWB as prominent markers of welfare and well-being research characterized by and using distinct approaches to their measurement. To empirically estimate potential bias introduced by including responses from proxy reports, we used a multi-group factor analytic (MGCFA) approach to test MI of multi-item scales (i.e., PANAS and ADL/IADL) as suggested by Meredith (1993). Cross-group equality constraints regarding factor loadings or/and item intercepts were used to test for metric and scalar MI, respectively. Violations of scalar MI may introduce systematic response bias into estimating population means unless such group-specific deviations are adequately addressed

in estimating factor scores. If the observed or latent group means can be validly compared, bias resulting from excluding proxy information can be directly estimated. The benefit of including difficult-to-survey subgroups of very old individuals was estimated relative to their contribution to representing the target population and the potential nuisance introduced to the assessment of key QoL indicators.

Analyses were performed using SAS 9.4 (SAS Institute, Cary, NC), Mplus 8.6 (Muthén & Muthén 1998–2021), and the *RISQ 2.1 functions* (De Heij et al., 2015) for R software (R Core Team, 2017).

4 Results

4.1 Including the Institutionalized Population

Coverage in the sampling frame

In the random register samples, the proportion of NHR identified was almost consistently smaller than expected from the best available census data (Table 1). The overall percentage of NHR in the gross sample obtained from community registers was 8.1%, with 12.0% expected from the census data. However, even in communities that explicitly excluded individuals flagged as NHR in their databases, the obtained samples included 62.8% of all institutionalized individuals expected from the census data. Most remarkably, even in communities that provided ample evidence for using the full sampling frame, we noted a discrepancy of -2.3 percentage points, most likely attributable to a broader definition of institutional settings in census data compared with institutions listed in the nursing home register. Hence, the margin of uncertainty that arises from diverging definitions of the institutionalized population and procedures to identify NHR in the register data is even greater than the estimated degree of potential undercoverage remaining without it (-1.6 percentage points).

Survey participation

Based on data from a feasibility study, mechanisms of survey nonresponse have been reported to be different for individuals living in private dwellings versus institutional settings (Wagner et al., 2019). Particularly, potential respondents in nursing homes were less easily accessible (i.e., lower contact rates) due to health impairments. However, once contacted, they show much lower refusal rates than individuals in private dwellings. In this study, reasons for refusing to participate were documented for 2,993 individuals (Figure 1). The main reasons for these were similar in private dwellings and NHR (Table 2). No interest (more than two out of three) and a general refusal to participate in surveys (one out of four or five) were the most common reasons for both groups of potential participants. In NHR, the length of interviews (8.4% vs. 6.8%) and inability to comment on the survey's subject

Table 1*Undercoverage of the institutionalized population in population register samples.*

Origin of information	Communities with explicit				All communities			
	inclusion		exclusion					
	<i>N</i> ^a	% NHR ^b	<i>N</i> ^a	% NHR ^b	<i>N</i> ^a	<i>N</i> individuals	<i>N</i> NHR	% NHR
Census	41	12.3	13	12.1	88	560,221	67,242	12.0
Sample	41	10.0	13	7.6	88	45,809	3,701	8.1
Difference		-2.3		-4.5				-3.9

^a Census data are not available for 6 communities. ^b NHR = nursing home residents.

Table 2*Reasons for individuals living in private housing or institutions refusing to participate.*

	% Total	% Private	% NHR ^a
Not interested ^b	69.4	69.6	63.5
General refusal	25.3	25.6	18.8
Not allowed by others	8.2	8.2	8.4
Interview too long	6.8	6.8	8.4
Subject too intimate	7.4	7.4	5.8
Lack of knowledge	4.4	4.4	6.0
Privacy concerns	3.7	3.7	3.4
No time	3.6	3.7	1.5
Participated too often	0.5	0.4	2.1
<i>N</i>	2,993	2,861	132

Design-weighted data.

^a Nursing home residents.

^b Multiple reasons for refusal to participate could be given.

(6.0% vs. 4.4%) were more frequent reasons for refusal than they were for private-dwelling individuals.

In the group of participants, the living status may have changed since the sample generation. A total of 151 addresses of participants had been flagged as institutions based on the residential register or nursing home repository data before fieldwork. Based on all available information related to dwellings at the time of the interview, 211 participants were identified as NHR. In some cases, information from different sources was contradictory. For example, 13 participants listed as private-dwelling according to the residential register lived in a nursing home at the time of first contact or at the time of the interview.

4.2 Including Those Unable to Conduct the Interview Via a Proxy Interview

Of all 8,040 individuals contacted, 1,362 were UCI themselves for health reasons (Figure 1). Of these, it was possi-

ble to conduct 176 proxy interviews. The design-weighted gross sample estimated that this corresponded to 15.1% of the noncommunicative, very old population that could, in principle, be included in this measure. For 38.4% of the remaining cases, the inability to provide informed consent due to mental illness precluded a proxy interview. In cases where informed consent could have been obtained, the most frequent reasons for not conducting a proxy interview were proxy refusal (24.0%), the target person opposing a proxy interview (23.7%), and no identifiable proxy (12.6%). However, in many cases, the reasons why the proxy interview could not be realized remain unknown (35.6%). The number of contacts needed to obtain a proxy interview in this population was estimated to be significantly higher than for self-reports (Mean = 3.0, 95% CI = [2.6, 3.5] vs. Mean = 2.4, 95% CI = [2.3, 2.5]).

About half (46.2%) of the interviews conducted with a proxy informant were with the children of the target persons, mostly daughters (64.0%). About one in five proxy interviews was conducted with spouses (18.7%). Proxy informants were predominantly female (70.2%). The percentage of self-reports decreased across age groups (Table 3). However, a similar number of proxy interviews were conducted in all three age groups. UCI were more often living in institutions, female, and less healthy than self-reporting target persons. For example, UCI were treated for significantly more health conditions. The possibility of conducting proxy interviews appears particularly relevant for representing NHR, as their share is four times higher in proxy reports than in self-reports (43.3% versus 11.0%). In line with our expectation, the availability of information differed between self- and proxy reports for different reasons. Proxy informants, for example, reported more often that they did not know the answer to a question than respondents in self-reports. However, refusal rates regarding specific questions were substantially higher in self-reports than in proxy interviews. Despite these plausible limitations, information on more than 90% of all questions posed is available to investigate the QoL of UCI.

4.3 Including Individuals with Cognitive Impairment in the Survey

Regarding the cognitive status of the participants, 3.7% of respondents (self-reports) declined to participate in the cognitive screening test, and 14.3% of participants did not complete all DemTect subtests. GDS-Scale scores were available for 96.1% of individuals represented by proxy interviews. Few proxy informants answered “do not know” or refused to rate the cognitive status of the target person (1.2% and 2.8%, respectively). Overall, cognitive status could be assessed and interpreted for 83.2% of participants. The assessment of cognition was more likely to be available in the youngest age group (85.5%, $\chi^2 = 7.03$, $df = 2$, $p = 0.030$) and in proxy interviews (96.1%, $\chi^2 = 9.09$, $df = 1$, $p = 0.003$). In contrast, the availability of information on cognitive status was independent of gender ($\chi^2 = 1.90$, $df = 1$, $p = 0.168$) and living situation ($\chi^2 = 1.28$, $df = 1$, $p = 0.258$).

The cognitive function of most individuals capable of self-report was screened as age-adequate (73.7%), whereas 16.4% had MCI and 9.8% had suspected dementia (Table 3). In UCI, however, the most common cognitive status was suspected dementia (73.5%), 7.4% were classified as MCI, and 19.1% had age-adequate cognitive functioning. Thus, in line with our expectations, including proxies as informants led to a better representation of people with dementia in the data. Overall, estimates from the total sample yielded a prevalence of 16.3% for dementia and 15.5% for MCI in the population 80 years or older.

4.4 Consequences for Representativity and Unit-Nonresponse-Bias

Failure to allow for proxy interviews in case the target persons were willing to participate but unable to conduct the interview themselves would have “lost” 176 participants and led to a decrease of 2.2 percentage points in the response rate. Similarly, the inclusion of the 211 individuals interviewed in nonprivate settings accounted for a 2.7 percentage point increase in the response rate. Restricting the sample to the 1,359 screened participants with uncompromised cognitive function (i.e., using the MCI classification as an exclusion criterion) would have resulted in a 6.3 percentage point drop in the response rate. The measures taken in this study to assure the best-possible inclusion of these difficult-to-survey groups increased the response rate from 15.3% to 23.2%.

Although an increase in survey participation represents an important step toward unbiased population estimates, a higher response rate is insufficient. It should be accompanied by increased representativity of the realized sample to reduce survey nonresponse bias efficiently. The analysis of response propensities for the sample showed only a minor predictive value of age, gender, living in an institution, or regional characteristics (BIK). Thus, a resulting R indicator

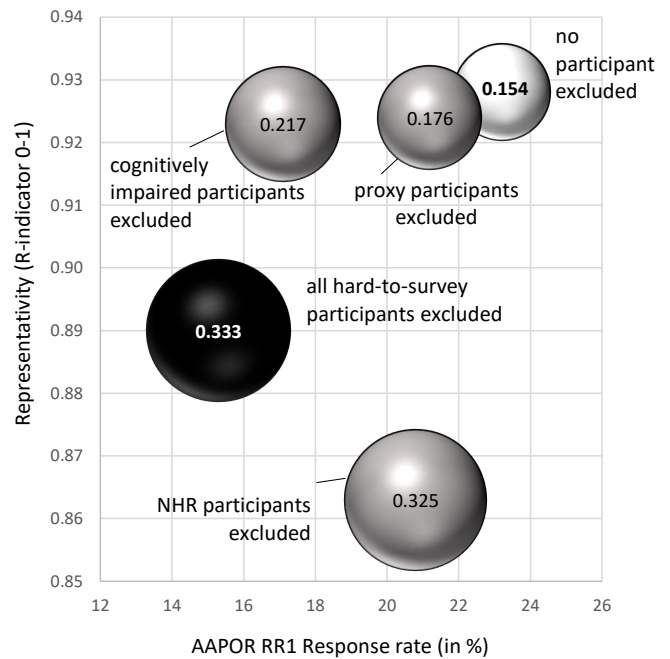


Figure 2

Maximum nonresponse bias (CV) resulting from failure to include difficult-to-survey groups of very old individuals. CV = Coefficient of Variation. Bubble size is proportional to the size of the CV. Bubble location represents the unique combination of the response rate and representativity of the retained sample.

of 0.928 with a 95% confidence interval [0.908 to 0.948] underscored the high representativity of the realized sample regarding these potential threats to participation. Additionally, an estimate of 0.154 [0.112 to 0.196] for the coefficient of variation suggested that even under a worst-case scenario regarding the characteristics of very old individuals not included in the study, population estimates will not exhibit more than a maximum of 15.4 percent nonresponse bias (Figure 2). In contrast, the failure to include the 640 individuals who showed signs of cognitive decline, lived in institutions, or were unable to conduct the interview themselves for health reasons would have resulted in a lower response rate, significantly worse representativity of the realized sample ($R = 0.89$, 95% CI = [0.881, 0.905]), and more than twice the insecurity regarding the maximum potential nonresponse bias ($CV = 0.333$, 95% CI = [0.296, 0.370]) in population parameter estimations.

4.5 Consequences for Substantive-Matter Conclusions

Population estimates for socioeconomic and health resources and SWB showed substantial differences across difficult-to-survey subgroups of very old individuals (Table

Table 3*Characteristics of respondents able versus unable to answer the interview themselves.*

	Self-report			Proxy-report			Test ^a			
	% / Mean	95% CI		% / Mean	95% CI		χ^2/F	df	p	
		Lower	Upper		Lower	Upper				
Gender (%)										
Male	37.0	34.5	39.6	28.3	20.6	36.1	3.6	1	0.057	
Female	63.0	60.4	65.5	71.7	63.9	79.4				
Age group (%)										
80-84 yrs	56.3	53.7	59.1	33.0	23.7	42.3	44.9	2	< 0.001	
85-89 yrs	30.6	28.1	33.0	32.4	24.1	40.8				
90 yrs or older	13.1	11.5	14.5	34.6	26.8	42.4				
Setting (%)										
Private dwelling	89.0	87.1	90.9	56.7	47.5	65.9	66.9	1	0.001	
Institution	11.0	9.1	12.9	43.3	34.1	52.5				
Treated health conditions (0-19) ^b	3.4	3.2	3.6	4.2	3.7	4.8	11.7	1	< 0.001	
Cognitive status (%) ^c										
Age-adequate function	73.7	71.0	76.5	19.1	12.5	25.7	153.6	2	< 0.001	
MCI	16.4	14.1	18.8	7.4	2.9	11.8				
Suspected dementia	9.8	7.9	11.7	73.5	66.0	81.1				
Response (%)										
Don't know	1.5	1.3	1.8	5.8	4.7	6.9	58.1	1/119	< 0.001	
Refuse to answer	2.0	1.6	2.4	0.4	0.2	0.7	40.8	1/119	< 0.001	
n	1687			176						

Calibration-weighted data.

^a Wald test for predicting proxy interview by participant characteristics or F-test for predicting levels of “do not know” answers and refusals by interview type. All tests consider the complex sampling design and use unbiased standard errors (Taylor linearization).^b A list of 19 currently treated health conditions (e.g., heart disease, hypertension, respiratory or lung disease, diabetes).^c A classification based on DemTect scores from self-reports and Global Deterioration Scale ratings from proxy reports.

4). Hence, not including specific subpopulations may lead to systematic bias in estimates for the population 80 years or older.

Socioeconomic status

In difficult-to-survey subgroups, ISEI-08 scores could not be computed for 5.6% of cognitively impaired individuals, 6.0% of those represented by proxy interviews, and 8.2% of those living in institutions. In contrast, missing rates for SES ranged between 2.1% and 3.1% in the remaining participants. Estimates for SES were substantially lower in all three subgroups (Table 4). In individuals with limited cognitive functioning and UCI, ISEI-08 scores were significantly (as judged from nonoverlapping confidence intervals) lower than those in the respective self-report and age-adequate cognition groups were. Hence, a substantial bias suggesting better socioeconomic resources for the very old population would result from failing to include these difficult-to-survey groups.

ADL/IADL independence

The data availability regarding functional status was high (92.8% to 100%) in both self- and proxy reports. Interestingly, more “do not know” answers were encountered with more complex tasks of everyday living (e.g., housework) in self-reports and private settings. This suggests that sharing chores was more common, leaving respondents unsure if they could perform this task when needed. Unsurprisingly, “do not know” answers were more common in individuals screened as cognitively impaired, although missing rates did not exceed 2.6% for any item in this subgroup.

The estimates for functional health were reported to be significantly lower in UCI than in interviewed participants (0.50 versus 1.6, Table 4). However, loglikelihood ratio tests for MI showed that proxy- and self-reports were not comparable already at the level of metric invariance ($\Delta_{-2LL_{corr}} = 22.05$, $\Delta_{df} = 2$, $p < 0.001$). Therefore, the degree of poten-

Table 4*(Sub-)Population estimates of QoL resources and outcomes in very old age.*

	Socioeconomic Status (ISEI, 10–90)			Functional health (I/ADL, 0–2)			Well-being PANAS, 1–5)		
	95% CI ^a			95% CI ^a			95% CI ^a		
	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
Full sample	40.84	39.38	42.30	1.54	1.50	1.58	3.26	3.16	3.34
Setting									
Private dwelling	41.41	39.88	42.94	1.65	1.61	1.68	3.31	3.17	3.34
Institution	37.08	33.82	40.33	0.88	0.80	0.96	2.91	2.73	3.08
Cognitive status									
Age-adequate	44.10	42.28	45.93	1.73	1.70	1.76	3.41	3.32	3.50
MCI or suspected dementia	35.38	32.89	37.86	1.20	1.11	1.28	2.94	2.80	3.08
Mode of survey participation									
Self-report	41.31	39.78	42.83	1.64	1.61	1.67	3.32	3.22	3.41
Proxy report	35.89	32.48	39.31	0.52	0.44	0.60	2.60	2.42	2.78

Calibration-weighted data.

^a Confidence intervals are based on standard errors that consider the complex sampling design (Taylor linearization).

tial response bias resulting from proxy informants answering items differently cannot be estimated empirically. The comparison of group means and even the inclusion of UCI in covariance structure analysis are severely limited.

Subjective well-being

The availability of information on affective states was comparable in self-reports and UCI participants on all five items and ranged between 91.7% (“alert”) and 96.6% (“enthusiastic”). However, there was more item-level nonresponse due to refusal in self-reports (1.3% to 1.4%) compared with less than four per thousand for all items in proxy reports. Refusal rates were slightly higher on average in NHR (1.8%) and those with beginning cognitive impairment (1.0%). However, “do not know” answers were responsible for most of the item-nonresponse in all respondents.

The tests for metric and scalar MI found no evidence of a) a different conceptual understanding of positive affect in self- compared with proxy reports or b) systematic differences between groups in the interpretation of the 5-point response scale, respectively (Table 5). Thus, full scalar MI was established for the PANAS scale, allowing for uncompromised integration of information on UCI in population estimates. The level of positive affect was significantly lower in proxy interviews than in self-reports (Mean = 2.60 and Mean = 3.32, respectively; Table 4). Thus, failure to include information on UCI would have led to a positively biased estimate of affective well-being in the 80+ population. However, a mean population value of 3.26 [3.16 – 3.34] was estimated for positive affect in the combined sample, suggest-

ing that the magnitude of bias due to the exclusion of UCI would be minor (i.e., 0.06 scale points). Similar risks apply when failing to include individuals living in institutions and those with (beginning) cognitive impairment, as levels of well-being were also significantly lower in these subgroups. Although proxy interviews appear to hold merit in representing individuals with pronounced cognitive decline, questions may linger regarding the validity of information collected via self-reports in less cognitively impaired individuals (e.g., MCI or suspected dementia). However, the “known truth,” against which responses might unequivocally be validated, is generally limited in most social science surveys. In this sample, discrepancies between reported and registered birth dates were observed to a similar degree in those with age-adequate cognitive function compared with those classified MCI or suspected dementia (6.0% and 5.8%, respectively).

5 Discussion

This study drew on metadata from sampling and fieldwork. It also used the responses from a large-scale survey on QoL to weigh the advantages and disadvantages of targeting and including difficult-to-survey subgroups of very old individuals. After comprehensive piloting, a study protocol was employed to maximize the inclusion of those with limited cognitive function, NHR, and UCI to optimize the representativity of the data and the precision of estimates. The results show that the failure to include these individuals would have resulted in much lower response rates and worse representativity. Hence, a potential maximum nonresponse bias is more than two times the size achieved with the full sam-

Table 5

Nested model comparison of configural, metric, and scalar measurement invariance (MI) for the PANAS scale in self- and proxy report subgroups

Degree of invariance (equality constraint)	Absolute model fit			Relative fit indices				Difference test ^a		
	χ^2	df	<i>p</i>	RMSEA	90% CI		<i>p</i>	$\Delta_{-2LL_{corr}}$	Δ_{df}	<i>p</i>
					Lower	Upper				
Configural MI (factor model)	79.0	10	< 0.001	0.087	0.070	0.106	< 0.001			
Metric MI (loadings)	91.4	14	< 0.001	0.078	0.063	0.094	0.001	2.44	4	0.656
Scalar MI (loadings and intercepts)	103.2	18	< 0.001	0.072	0.059	0.086	0.003	13.00	8	0.112

Calibration-weighted data.

^a Likelihood-ratio model comparison for robust maximum likelihood estimation used scaling correction factors.

ple. Increasing response rates may be critical to guarantee that there is enough power for subgroup comparison, given the very uneven proportion of men and women or different birth cohorts (i.e., oldest-old) in this population segment and the accumulated risk to survey participation due to communication or care needs. Results regarding NHR showed that the omission of a relatively small difficult-to-survey group had minor adverse effects on response rates but a substantial impact on potential nonresponse bias. While this finding may be surprising upfront, it is a direct consequence of the fact that this characteristic was deemed a particularly important aspect of representativity analysis and has been explicitly considered a predictor of unit nonresponse in this study. Because of the scarcity of characteristics of the oldest old available from the register data for representativity analysis, our promising results regarding the identification of NHR in community register data, and the expectation that NHR status serves as a good proxy for threats to participation because of gatekeepers and poor health status, we think this emphasis is well-justified.

The findings of this study underscore the need to include difficult-to-survey subgroups, such as the institutionalized and communication impaired, to represent the population of individuals aged 80 years or older. Undercoverage of NHR in random register samples may be avoided using privileged data access of public-law institutions (e.g., universities) and dedicated quality control of received samples. This assertion is confirmed by more recent experiences from the second wave of the study. Here, all register samples were obtained directly by the university, and the proportion of identified nursing home addresses was even closer to the expected figure. According to the census data, the percentage of NHR in these communities was an expected 11.7%, while 8.6% were identified in the sample. Acknowledging that a difference of about 2.3 percentage points could be attributed to a broader definition of this group in the census, only a minor potential undercoverage of -0.8 percentage points results from the sampling procedure employed. Thus, we expect

that alternative strategies, such as dual-frame sampling, may be challenged by considerable uncertainty in defining the institutionalized subpopulation and a loss of efficiency from overlapping target populations in the nursing home and community register samples.

The cognitive status could be measured for 83.2% of the sample and resulted in a prevalence estimation of 163 per thousand, a slight underestimation of the actual dementia prevalence of 194 per thousand known from the health insurance data. This prevalence is also achieved by including persons with dementia (PWD) via proxy informants. Studies not including UCI via proxy interviews, such as the SHARE study, report an even lower prevalence, of 104 per thousand, in an even older (85+ years) population (Ferreira et al., 2020). Although proxy interviews for UCI come with difficulties in the comparability of measurement, they represent a keystone for the representation of PWD in survey samples. Our results also showed that substantive-matter (here: socioeconomic status, functional health, and SWB) research could profit from including difficult-to-survey subgroups in multiple ways. First, we presented evidence that information on the well-being of UCI can validly be integrated, allowing for a less biased population estimate where subgroup levels differ from those generally observed in less inclusive studies. Second, differences observed in difficult-to-survey subgroups regarding socioeconomic status or independence in ADLIADL agree with findings from studies targeting more specific subpopulations or studies with similar inclusive approaches (Kelfve et al., 2013). The identified tendency for lower socioeconomic status in these groups underlines the importance of including them in research on social and health inequalities.

5.1 Limitations and Future Directions

Some limitations of the current study warrant attention. First, while consulting both interview data and fieldwork metadata may ultimately result in the best-possible characterization of participants' situation at the time of the inter-

view (e.g., concerning NHR status) and benefits data interpretation, such updates may not be easily or timely integrated into the computation of survey weights. In this study, using updated information on NHR status instead of outdated residential register information may have contributed to more efficient calibration weights. Second, the assessment of cognitive function in UCI relied on proxy GDS ratings instead of objective testing. We acknowledge that although some degree of age-adequate functioning and MCI may be compared across instruments, this option may not extend to more severe levels of cognitive impairment (i.e., higher GDS scores). Third, and more conceptually, the usefulness of the R indicator approach critically depends on selecting relevant predictors of participation. Considering institutionalization as a proxy for adverse conditions in very old age may allow for a rather strict test of the risks of noninclusion of function-impaired groups to the representativity of our old age sample. When exploring survey error related to the exclusion of other difficult-to-survey subgroups (e.g., migrants), different predictors of response propensity (e.g., language proficiency) may be deemed relevant. In this situation, identifying bias (e.g., cultural bias) in survey responses using MI testing may be warranted. Therefore, we expect that this example of considering both sampling errors and measurement errors related to the inclusion of difficult-to-survey older adults will be useful beyond research on aging. Finally, the measures taken to reduce unit nonresponse in the difficult-to-survey subpopulations studied here relied on structural resources (e.g., detailed structural data available from registers, communities able/willing to draw samples, institutions granting access to NHR) or social resources such as the availability of proxy informants. Because such resources are certainly not randomly distributed in the population of 80+, generalization of findings concerning the characteristics of those living in institutions, those unable to answer the interview, and those with impaired cognitive functioning remain limited. Similarly, although the difficult-to-survey groups of very old adults studied here are very likely to show lower survey participation also in other countries, their prevalence and characteristics may diverge, and different structural (e.g., digital infrastructure, data protection law, see Scherpenzeel et al. (2017) for an overview of sampling frames for European social surveys) as well as social resources (e.g., caring relatives) may be available for inclusive survey research.

5.2 Conclusion

Taken together, this study provided ample evidence that the benefits of tailored measures to reduce unit nonresponse in difficult-to-survey groups and the integration of responses obtained from NHR, the cognitively impaired, and for UCI via proxy interviews may outweigh the undisputed challenges along this road. However, evaluations should be increasingly based on more substantiated research hypotheses

that may well go beyond the estimation of population point parameters focused on in this study and consider bias in the covariance structure.

The strong discrepancy in reported levels of I/ADL independence between self- and proxy reports, with evidence of the lack of MI across informants, warrants further investigation. The authors hope that the availability of such information may help shed light on mechanisms that could ultimately promote independence in daily living, particularly for those very old individuals who already receive support and assistance or those less aware of existing threats to their independence.

Acknowledgements

The authors declare that they have no conflicts of interest. Survey materials and a de-identified dataset of survey responses are available as scientific use files at the GESIS – Leibniz Institute for the Social Sciences data repository (Zank et al., 2020). The de-identified metadata and analytical code used in this study are available at the GESIS – Leibniz Institute for the Social Sciences data repository (Kaspar et al., 2023). The project is part of the Key Research Area “Aging and Demographic Change” at the Cologne Center for Ethics, Rights, Economics, and Social Sciences of Health (ceres) at the University of Cologne. The members of the project board are Michael Wagner and Susanne Zank. This work was supported by a grant from the Ministry of Culture and Science of the German State of North Rhine-Westphalia (323-8.03-125240).

References

- Andersen-Ranberg, K., Petersen, I., Robine, J.-M., & Christensen, K. (2005). Who Are the Oldest-Old? In A. Börsch-Supan (Ed.), *Health, ageing and retirement in Europe* (pp. 35–40). Mannheim Research Institute for the Economics of Aging (MEA).
- Aust, V., & Gilberg, R. (2016). *Überlegungen zur Gemeindestichprobe der geplanten Erhebung NRW80+: Kurzbericht*. infas Institut für Sozialforschung.
- Borg, I. (2019). Age- and gender-related differences in the structure and the meaning of personal values. *Personality and Individual Differences*, 138, 336–343. <https://doi.org/10.1016/j.paid.2018.10.013>
- Brix, J., Steinacker, G., Schneekloth, U., & Stadler, M. (2016). Machbarkeitsstudie NRW80+: Expertise zur Erreichbarkeit einer Personenstichprobe Hochaltriger.
- Chan, K.-Y., Drasgow, F., & Sawin, L. L. (1999). What is the shelf life of a test? The effect of time on the psychometrics of a cognitive ability test battery. *Journal of Applied Psychology*, 84(4), 610–619. <https://doi.org/10.1037/0021-9010.84.4.610>

- Darin-Mattsson, A., Fors, S., & Kåreholt, I. (2017). Different indicators of socioeconomic status and their relative importance as determinants of health in old age. *International Journal for Equity in Health*, 16(1). <https://doi.org/10.1186/s12939-017-0670-3>
- Davies, K., Collerton, J. C., Jagger, C., Bond, J., Barker, S. A. H., Edwards, J., Hughes, J., Hunt, J. M., & Robinson, L. (2010). Engaging the oldest old in research: lessons from the Newcastle 85+ study. *BMC geriatrics*, 10, 64. <https://doi.org/10.1186/1471-2318-10-64>
- De Heij, V., Schouten, B., & Shlomo, N. (2015). RISQ manual 2.1 Tools in SAS and R for the computation of R-indicators, partial R-indicators and partial coefficients of variation. <http://hummedia.manchester.ac.uk/institutes/cmist/risq/RISQ-manual-v21.pdf>
- Deeg, D. J., van Tilburg, T., Smit, J. H., & de Leeuw, E. D. (2002). Attrition in the Longitudinal Aging Study Amsterdam. *Journal of Clinical Epidemiology*, 55(4), 319–328. [https://doi.org/10.1016/S0895-4356\(01\)00475-9](https://doi.org/10.1016/S0895-4356(01)00475-9)
- Degnen, C. (2007). Minding the gap: The construction of old age and oldness amongst peers. *Journal of Aging Studies*, 21(1), 69–80. <https://doi.org/10.1016/j.jaging.2006.02.001>
- Doblhammer, G., Schulz, A., Steinberg, J., & Ziegler, U. (2012). *Demografie der Demenz* (1. Aufl.). Huber.
- Doebler, P. (2020). mada: Meta-analysis of diagnostic accuracy. <https://CRAN.R-project.org/package=mada>
- Edwards, M. C., & Wirth, R. J. (2009). Measurement and the Study of Change. *Research in Human Development*, 6(2-3), 74–96. <https://doi.org/10.1080/15427600902911163>
- El Haj, M., Antoine, P., Nandrino, J. L., & Kapogianis, D. (2015). Autobiographical memory decline in Alzheimer's disease, a theoretical and clinical overview. *Ageing research reviews*, 23(Pt B), 183–192. <https://doi.org/10.1016/j.arr.2015.07.001>
- Farmer, M., & Macleod, F. (2011). *Involving disabled people in social research: Guidance by the Office for Disability Issues*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/321254/involving-disabled-people-in-social-research.pdf
- Ferreira, R. G., Brandão, M. P., & Cardoso, M. F. (2020). An update of the profile of older adults with dementia in Europe: findings from SHARE. *Ageing & Mental Health*, 24(3), 374–381. <https://doi.org/10.1080/13607863.2018.1531385>
- Feskens, R. C. W. (2009). *Difficult Groups in Survey Research and the Development of Tailor-Made Approach Strategies* (Dissertation). University of Utrecht. Utrecht.
- Fillenbaum, G. G. (1988). *Multidimensional Functional Assessment of older Adults: The Duke Older Americans Resources and Services Procedures*. Duke University Center for the Study of Aging and Human Development.
- Finlayson, M., Mallinson, T., & Barbosa, V. M. (2005). Activities of daily living (ADL) and instrumental activities of daily living (IADL) items were stable over time in a longitudinal study on aging. *Journal of Clinical Epidemiology*, 58(4), 338–349. <https://doi.org/10.1016/j.jclinepi.2004.10.008>
- Fleishman, J. A., Spector, W. D., & Altman, B. M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 57(5), S275–84. <https://doi.org/10.1093/geronb/57.5.S275>
- Foo, H., Mather, K. A., Thalamuthu, A., & Sachdev, P. S. (2019). The many ages of man: diverse approaches to assessing ageing-related biological and psychological measures and their relationship to chronological age. *Current Opinion in Psychiatry*, 32(2), 130–137. <https://doi.org/10.1097/YCO.0000000000000473>
- Gaertner, B., Koschollek, C., Grube, M. M., Lütke, D., Fuchs, J., Scheidt-Nave, C., Gößwald, A., & Wetzstein, M. (2019). Including nursing home residents in a general population health survey in Germany. *Survey Methods: Insights from the Field (SMIF)*. <https://doi.org/10.13094/SMIF-2019-00010>
- Gaertner, B., Lütke, D., Koschollek, C., Grube, M. M., Baumert, J., Scheidt-Nave, C., Gößwald, A., Fuchs, J., & Wetzstein, M. (2019). Effects of a sequential mixed-mode design on participation, contact and sample composition – Results of the pilot study “IMOA – Improving Health Monitoring in Old Age”. *Survey Methods: Insights from the Field (SMIF)*. <https://doi.org/10.13094/SMIF-2019-00016>
- Ganzeboom, H. B. G. (2010). *A New International Socio-Economic Index [IseI] of Occupational Status for the International Standard Classification of Occupation 2008 [Isco-08] Constructed with Data from the Issp 2002-2007*. 2010. [http://www.harryganzeboom.nl/Pdf/2010%5C%20-%5C%20Ganzeboom-ISEI08-ISSP-Lisbon-\(paper\).pdf](http://www.harryganzeboom.nl/Pdf/2010%5C%20-%5C%20Ganzeboom-ISEI08-ISSP-Lisbon-(paper).pdf)
- Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56. [https://doi.org/10.1016/0049-089X\(92\)90017-B](https://doi.org/10.1016/0049-089X(92)90017-B)
- Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three Internationally Standardised Measures for Compara-

- tive Research on Occupational Status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison* (pp. 159–193). Springer US. https://doi.org/10.1007/978-1-4419-9186-7_9
- Gerritsen, D. L., Steverink, N., Ooms, M. E., de Vet, H. C. W., & Ribbe, M. W. (2007). Measurement of overall quality of life in nursing homes through self-report: the role of cognitive impairment. *Quality of Life Research*, 16(6), 1029–1037. <https://doi.org/10.1007/s11136-007-9203-7>
- Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly*, 51(4 PART 2), S156–S172. https://doi.org/10.1093/poq/51.4_PART_2.S156
- Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, 74(5), 849–879. <https://doi.org/10.1093/poq/nfq065>
- Grundy, E., & Holt, G. (2001). The socioeconomic status of older adults: how should we measure it in studies of health inequalities? *Journal of Epidemiology and Community Health*, 55(12), 895–904. <https://doi.org/10.1136/jech.55.12.895>
- Hall, S., Longhurst, S., & Higginson, I. J. (2009). Challenges to conducting research with older people living in nursing homes. *BMC geriatrics*, 9, 38. <https://doi.org/10.1186/1471-2318-9-38>
- Hansen, S., Kaspar, R., Wagner, M., Woopen, C., & Zank, S. (2021). The NRW80+ Study: Conceptual background and design decisions. *Zeitschrift für Gerontologie und Geriatrie*, 54, 76–84. <https://doi.org/10.1007/s00391-021-01970-z>
- Hilleras, P. K., Jorm, A. F., Herlitz, A., & Winblad, B. (1998). Negative and Positive Affect among the Very Old. *Research on Aging*, 20(5), 593–610. <https://doi.org/10.1177/0164027598205003>
- Hülür, G., Ram, N., Willis, S. L., Schaie, K. W., & Gerstorf, D. (2015). Cognitive dedifferentiation with increasing age and proximity of death: Within-person evidence from the Seattle Longitudinal Study. *Psychology and Aging*, 30(2), 311–323. <https://doi.org/10.1037/a0039260>
- 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>
- Kalbe, E., Calabrese, P., Fengler, S., & Kessler, J. (2013). DemTect, PANDA, EASY, and MUSIC: cognitive screening tools with age correction and weighting of subtests according to their sensitivity and specificity. *Journal of Alzheimer's disease: JAD*, 34(4), 813–834. <https://doi.org/10.3233/JAD-122128>
- Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A. P., Brand, M., & Bullock, R. (2004). DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International journal of geriatric psychiatry*, 19(2), 136–143. <https://doi.org/10.1002/gps.1042>
- Kaspar, R., Brijoux, T., Albrecht, A., Zimmermann, J., Wenner, J., Fey, J., Reissmann, M., Wagner, M., & Zank, S. (2023). Replication data: Challenges and benefits of including institutionalized, cognitively impaired, and unable-to-respond individuals in a representative survey of the very old. <https://doi.org/10.7802/2536>
- Kaspar, R., Gabrian, M., Brothers, A., Wahl, H.-W., & Diehl, M. (2019). Measuring awareness of age-related change: Development of a 10-item short form for use in large-scale surveys. *The Gerontologist*, 59(3), e130–e140. <https://doi.org/10.1093/geront/gnx213>
- Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *JAMA: The Journal of the American Medical Association*, 185, 914–919. <https://doi.org/10.1001/jama.1963.03060120024016>
- Kelfve, S. (2019). Underestimated Health Inequalities Among Older People—A Consequence of Excluding the Most Disabled and Disadvantaged. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 74(8), 125–134. <https://doi.org/10.1093/geronb/gbx032>
- 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>
- Kercher, K. (1992). Assessing Subjective Well-Being in the Old-Old: The PANAS as a Measure of Orthogonal Dimensions of Positive and Negative Affect. *Research on Aging*, 14(2), 131–168. <https://doi.org/10.1177/0164027592142001>
- Kessler, J., Calabrese, P., & Kalbe, E. (2010). DemTect-B: ein Äquivalenztest zum kognitiven Screening DemTect-A. *Fortschritte der Neurologie-Psychiatrie*, 78(9), 532–535. <https://doi.org/10.1055/s-0029-1245452>
- Kessler, J., Fengler, S., Kaesberg, S., Müller, K., Calabrese, P., Ellwein, T., & Kalbe, E. (2014). DemTect 40- und DemTect 80+: Neue Auswertungsroutinen für diese Altersgruppen. *Fortschritte der Neurologie-Psychiatrie*, 82(11), 640–645. <https://doi.org/10.1055/s-0034-1385278>

- Kühn, K., Porst, R., & Zentrum für Umfragen, Methoden und Analysen -ZUMA- (Eds.). (März 1999). *Befragung alter und sehr alter Menschen. Besonderheiten, Schwierigkeiten und methodische Konsequenzen. Ein Literaturbericht*. ZUMA.
- LaPlante, M. P. (2010). The classic measure of disability in activities of daily living is biased by age but an expanded IADL/ADL measure is not. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65(6), 720–732. <https://doi.org/10.1093/geronb/gbp129>
- Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *The Gerontologist*, (9), 179–186.
- Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the Positive and Negative Affect Schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences*, 27(3), 405–416. [https://doi.org/10.1016/S0191-8869\(98\)00251-7](https://doi.org/10.1016/S0191-8869(98)00251-7)
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Mueller, S., Wagner, J., & Gerstorff, D. (2017). On the role of personality in late life. In J. Specht (Ed.), *Personality development across the lifespan* (pp. 69–84). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-804674-6.00006-5>
- Perneczky, R. (2003). Die Eignung einfacher klinischer Tests für die Erkennung der leichten kognitiven Störung und der leichtgradigen Demenz. *Aktuelle Neurologie*, 30(3), 114–117. <https://doi.org/10.1055/S-2003-38269>
- R Core Team. (2017). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Ed.). <https://www.r-project.org/>
- 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 A. Kumar, C. L. E. Katona, & M. T. Abou-Saleh (Eds.), *Principles and practice of geriatric psychiatry* (pp. 162–169). Wiley. <https://doi.org/10.1002/9780470669600.ch31>
- Reissmann, M., Storms, A., & Woopen, C. (2021). Individual values and spirituality and their meaning for affective well-being and engagement with life in very old age. *Zeitschrift für Gerontologie und Geriatrie*, 54(2), 85–92. <https://doi.org/10.1007/s00391-021-01974-9>
- Sakshaug, J. W. (2022). Reducing nonresponse and data linkage consent bias in large-scale panel surveys. *Forum for health economics & policy*, 25(1-2), 41–55. <https://doi.org/10.1515/fhep-2021-0060>
- Sand, M. (2014). *Dual-Frame-Telefonstichproben: Entwicklung, Handhabung und Gewichtung* (Vol. 2). GESIS – Leibniz-Institut für Sozialwissenschaften.
- Schanze, J.-L. (2017). Report on sampling practices for the institutionalized population in social surveys: Deliverable 2.1 of the SERISS project funded under the European Union’s Horizon 2020 research and innovation programme (GA No: 654221). www.seriss.eu/resources/deliverables
- Schanze, J.-L. (2019). Schwer befragbar und vernachlässigbar? Die Bevölkerung in Gemeinschaftsunterkünften. *Informationsdienst Soziale Indikatoren*, (62), 13–20. <https://doi.org/10.15464/isi.62.2019.13-20>
- Schanze, J.-L. (2021). Response behavior and quality of survey data: Comparing elderly respondents in institutions and private households. *Sociological Methods & Research*, 004912412199553. <https://doi.org/10.1177/0049124121995534>
- Schanze, J.-L., & Zins, S. (2019). Undercoverage of the elderly institutionalized population: The risk of biased estimates and the potentials of weighting institutionalized population and not the risk of biased estimates: A simulation study using cross-national survey data (Survey Methods: Insights from the field, Ed.). <https://doi.org/10.13094/SMIF-2019-00017>
- Scherpenzeel, A., Maineri, A., Bristle, J., Pflüger, S.-M., Mindarova, I., Butt, S., Zins, S., Emery, T., & Luijkx, R. (2017). Report on the use of sampling frames in European studies: Deliverable 2.1 of the SERISS project funded under the European Union’s Horizon 2020 research and innovation programme (GA No: 654221).
- Schiel, S., Ruiz Marcos, J., & Fahrenholz, M. (2021). *Methodenbericht Deutscher Alterssurvey (DEAS): Schriftliche Panelbefragung 2020*. infas - Institut für Angewandte Sozialwissenschaft GmbH.
- Schneekloth, U., & Müller, U. (1998). *Hilfe- und Pflegebedürftige in Heimen: Endbericht zur Repräsentativerhebung im Rahmen des Forschungsprojekts "Möglichkeiten und Grenzen selbständiger Lebensführung in Einrichtungen"* (2nd ed., Vol. 147.2). Kohlhammer.

- Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, 35(1), 101–113.
- Tesch-Römer, C., & Wahl, H.-W. (2017). Toward a More Comprehensive Concept of Successful Aging: Disability and Care Needs. *The journals of gerontology. Series B, Psychological sciences and social sciences*, 72(2), 310–318. <https://doi.org/10.1093/geronb/gbw162>
- The American Association for Public Opinion Research. (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.). AAPOR. http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf
- Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281–300. <https://doi.org/10.1037/a0017908>
- Vogel, C., Klaus, D., Wettstein, M., Simonson, J., & Tesch-Römer, C. (2020). German ageing survey (DEAS). In D. Gu & M. E. Dupre (Eds.), *Encyclopedia of Gerontology and Population Aging* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-319-69892-2_1115-1
- 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>
- Wagner, M., Rietz, C., Kaspar, R., Janhsen, A., Geithner, L., Neise, M., Kinne-Wall, C., Woopen, C., & Zank, S. (2018). Quality of life of the very old: Survey on quality of life and subjective well-being of the very old in North Rhine-Westphalia (NRW80+). *Zeitschrift für Gerontologie und Geriatrie*, 51(2), 193–199. <https://doi.org/10.1007/s00391-017-1217-3>
- Zank, S., Woopen, C., Wagner, M., Rietz, C., & Kaspar, R. (2020). Lebensqualität und Wohlbefinden hochaltriger Menschen in NRW (Repräsentativbefragung NRW80+). <https://doi.org/10.4232/1.13527>

Appendix A

Computation of Survey Weights

A computation of survey weights was done in two steps. First, design weights were calculated for all individuals ($N = 8,040$) from the gross sample of potential participants as the inverse of the inclusion probabilities. The inclusion probabilities for the individuals in the sampling frame were defined in the two-stage sampling design as the product of probabilities for selecting communities (proportional-to-size pps sampling of primary sampling units) and the probabilities for simple random selection of individuals within PSUs. The design weights also corrected for the deliberately disproportionate inclusion probabilities for the design groups defined by age group (80–84, 85–89, 90+ years) and gender (Table A1) in selecting the gross sample of potential participants to be contacted. Disproportionate sampling was necessary for two reasons. First, the older age groups and men represent small proportions of the very old population. Consequently, a simple random sampling would have resulted in only a few respondents in these design groups and compromised the statistical power for comparing population subgroups. Second, a feasibility study showed different response rates concerning the age and sex of the targeted participants. The analysis of the gender and age population structure also revealed that it would be practically impossible to sample enough individuals in the rare population groups (i.e., men 90+) to achieve equal cell sizes of $N = 300$ for all design groups in the projected realized (net) sample of approximately 1,800 study participants. Hence, a less extreme oversampling of older age groups and men was conducted that would still result in the high statistical power of the subgroup comparison. The total design weights were calculated by multiplying the probability of the community to be drawn into the sample, the probability of the person to be drawn into the community sample, and an adjustment that considers the deliberate oversampling of specific design groups during the last step of selecting individuals for the gross sample. The resulting design weights ranged from 0.278 to 1.738, and the efficiency of weighting was 84.7%.

Second, calibration weights have been computed for all individuals from the realized sample ($N = 1,863$) to correct for selective nonresponses. The recalibration of the design weights was conducted using an iterative process regarding the known population distributions of household size, institutional versus private dwelling, marital status, administrative district, community size and type (BIK-10), age, and gender. The resulting calibration weights ranged from 0.2 to 1.942, and the weighting efficiency was 72.4%. A comparison of sample and population distributions of key population characteristics at different levels of weighting is reported in (Hansen et al., 2021).

Table A1

Gross sample selection probabilities by design group.

Design group	Sampling frame abs.	Gross sample abs.	Prob. %
<i>Male</i>			
80–84 years	10,699	1,407	13
85–89 years	5,150	1,179	23
90 years or older	1,739	1,005	58
<i>Female</i>			
80–84 years	15,668	1,608	10
85–89 years	9,369	1,501	16
90 years or older	5,512	1,340	24
Total	48,137	8,040	100

Appendix B

Nursing Home Residents (NHR)

During the initial steps of obtaining samples of secondary sampling units (SSU) from community registers, nursing home residents (NHR) in the sample were identified based on their primary address of residence and a comprehensive repository of 1,276 addresses of nursing homes and care facilities in North-Rhine Westphalia (NRW).

Additional information on dwelling status and care was collected during contacting and interviewing to validate or update preliminary register-based information from sampling. More specifically, interviewers rated the housing situation during contact for all potential participants in the gross sample, using the categories *typical private apartment or house, nursing home, nursing home facility (e.g., hospice), residential care group, multigenerational house, senior residence, retirement home, and assisted living apartment or house*. In addition, interviewers assessed whether the address was likely to be part of an institution after an interview had been realized. During interviews, the interviewee (i.e., target or proxy person) answered questions about the time of relocation and the need for full inpatient care.

For this study, the definition of NHR at the time of

the interview considered the most current and consistent information from sampling, fieldwork, and interviews. To this end, a hierarchy of available information on institutionalization was defined. Priority was given to trained interviewers' assessments of the housing situation at the time of the interview. More specifically, NHRs were defined as all individuals living in a *nursing home, nursing home facility (e.g., hospice), or residential care group*. Only if this information was unavailable (e.g., due to interviews being conducted elsewhere), or the target person lived in a retirement home or a seniors' residence in which both private and institutional dwellings are often possible, information from residential registers and contact during fieldwork was considered. In the case of inconsistent information from residential registers and fieldwork, information on the recent relocation from the interview was considered. Fieldwork information from the time of first contact was prioritized over residential register information if the target person had relocated within the last 3 years. In case the target person did not move within the last 3 years or no information on relocation was available, interview information on receiving full inpatient care was considered to define NHR status.