The Nature of Missing Data in Longitudinal Surveys with Face-to-Face Data Collection

Inauguraldissertation zur Erlangung des akademischen Grades einer Doktorin der Sozialwissenschaften der Universität Mannheim

vorgelegt von

Sabine Friedel

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

Betreuer/in:
Prof. Annelies G. Blom, Ph.D.
Associate Prof. Dr. Peter Lugtig

Gutachter/innen:
Prof. Annelies G. Blom, Ph.D.
Associate Prof. Dr. Peter Lugtig
Prof. Dr. Edith de Leeuw

Tag der Disputation:
14.01.2021

Table of Contents

Tables

Figures

Abbreviations

AAPOR	American Association for Public Opinion Research
APC	age-period-cohort
BHPS	British Household Panel Survey
CAPI	Computer Assisted Personal Interview
cont.	continued
DE	Germany
ELSA	English Longitudinal Study of Ageing
ERIC	European Research Infrastructure Consortium
EU	European Union
Euro-D	European Union initiative to compare symptoms of depression
HH	household
IADL	instrumental activities of daily living
ICC	intraclass correlation
ID	identity
IT	Italy
km	kilometer
MAR	missing at random
max.	maximum
MCAR	missing completely at random
min.	minimum
MNAR	missing not at random
n	sample size
PSID	Panel Study of Income Dynamics
RR	response rate
SD	standard deviation
SE	Sweden
s.e.	standard error
SHARE	Survey of Health, Ageing and Retirement in Europe
SI	Slovenia
SIPP	Survey of Income and Program Participation
TSE	Total Survey Error
vs.	versus
W	Wave

Acknowledgement

The completion of this undertaking would not have been possible without the support of so many great and inspiring people. You know about my deep appreciation and I gratefully acknowledge your assistance. I am glad to have you in my life. Thank you!

In particular, I would like to express gratitude to my supervisors, thesis reviewers, and members of my defense committee: Prof. Annelies Blom, Ph.D., Associate Prof. Dr. Peter Lugtig, Prof. Dr. Edith de Leeuw, and Prof. Dr. Florian Keusch. Moreover, I would like to give special thanks to Martina Börsch-Supan, Ph.D. and Prof. Dr.h.c. Axel Börsch-Supan, Ph.D., and my mentors Dr. Julie Korbmacher and Dr. Annette Scherpenzeel.

1. General Introduction

Why What You Don't Know Matters

— David J. Hand (2020)

The great value of longitudinal surveys lies in the opportunity not only to take a snapshot of society at one point in time but to take a video of how society changes, growths and develops at both, the level of the individual and at the level of society as a whole (Davies 1994). For this reason, longitudinal surveys are very valuable and, over the last 30 years, many resources have been invested to establish or to improve longitudinal surveys (e.g., Fienberg and Tanur 1986; International Conference of Methodology of Longitudinal Surveys 2006; Toepoel and Schonlau 2017). One persistent problem in longitudinal surveys is that people may not participate or may not answer every single question and this behavior will result in missing data (Groves et al. 2002).

Missing data is the absence of information that was intended to be gathered in data collection (Allison 2009), and missing data occurs commonly in longitudinal survey data collection (e.g., Laird 1988; Yan, Curtin, and Jans 2010; de Leeuw, Hox, and Luiten 2018). Therefore, understanding, preventing, and reducing missing data in longitudinal surveys is essential in order to allow for precise and unbiased estimates of the surveyed population.

For instance, the Survey of Health, Ageing and Retirement (SHARE) (Börsch-Supan et al. 2013) provides long-term data to study the link between social, economic, and health aspects in the lives of Europeans aged 50 plus. However, the analysis of one longitudinal dataset of SHARE, that contains records from 21,227 initially recruited individuals aged 50 or older from 2004 to 2016, revealed that this dataset contains a large amount of missing data. After 12 years, about 60% of the initially recruited individuals of the study had dropped out (see chapter 2). This missing data may seriously impact the results of studying the links among these data and the conclusions researchers may draw for the overall target population. First, the decrease in number of observations in the data set induces an increase in the variance of estimates, and thereby, lowers the precision of estimates (Särndal and Lundström 2005). Second, there may be a systematic difference between the group of individuals who participate in the survey continuously and those who drop out. Nonresponse, resulting from dropping out systematically, can lead to a severe bias of the estimates. If this is the case, researchers have cause to doubt whether the results of their study are applicable to the overall target population (Little and Rubin 2020).

Moreover, the examined data set revealed that about 13% out of these 21,227 initially recruited individuals did not report their income in the first year of data collection in 2004. The remaining sample 12 years later showed that about 6% did not report their income in 2016. However, income is one of the most important determinants researcher use to study links between social, economic, and health aspects. For instance, how much individuals earn over their lifetime determines their health behavior and retirement decisions (e.g., Gough, Adami and Waters 2008; Knoll 2011; Laaksonen et al. 2003; Pampel, Krueger and Denney 2010). Again, if the number of observations on the income data is reduced, the variance of the estimates increases. In the worst case, this missing data caused by income item nonresponse may bias estimates (Rubin and Little 2020). Any conclusion drawn for the target population based on the survey statistics of only those with complete data is potentially wrong (e.g., mistaken predictions for disposable income, inappropriate retirement reforms, with serious consequences for the older population such as poverty in old age).

The good news is that there are methods to deal with missing data, e.g., weighting strategies and or imputation methods, which allow to increase precision of estimates and reduce bias (Graham and Hofer 2000; Little and Vartivarian 2005). The database SHARE for instance provides various weights (Bergmann, De Luca and Scherpenzeel 2017; Klevmarken, Swensson and Hesselius 2005) and imputed values for missing income or applications for imputing missing monetary and non-monetary values (De Luca 2018a, 2018b). However, generating weights and imputations and applying these methods is not always straightforward (Ciol et al. 2006; de Leeuw 2001; Haziza and Beaumont 2017) and highly dependent on the researchers' assumption (de Leeuw 2001; Kalton and Flores-Cervantes 2003). Therefore, preventing missing data from the beginning seems the most obvious solution to minimize the risk of biased estimates.

Besides preventing and reducing missing data, people may use other, more complete, data sources to provide precise and unbiased estimates for the target population, such as for example administrative data (Majcen 2017). Administrative data sets provide rich data but primarily for purposes of registration, transaction and record keeping, such as births and deaths records, or pensions taxations (Groves and Schoeffel 2018). To answer questions such as "How much do people earn, and why?" or "How much do or can people spend on health care expenditures?" or "Why do people retire?", in other words, to understand the links between social, economic, and health aspects, researchers need to gather much more information than just hard numbers of the composition of the target population (Groves and Schoeffel 2018). This additional information can be provided by survey data only.

Two types of surveys are commonly used to collect data in order to study links in our society. The first type is a single *cross-sectional* survey, which collects information from sample elements at only one point in time (Kesmodel 2018). The second type is a *longitudinal* survey, which "collects data from the same sample elements on multiple occasions over time" (Lynn 2009, 1). The term *longitudinal survey* is also often used in the literature to refer to one specific type of longitudinal surveys, namely a *panel survey*.

Thus, in my thesis, the terms *longitudinal survey* and *panel survey* are used interchangeably.

Data, that represent the target population, collected by a single cross-sectional survey can be useful to determine the status quo (Kesmodel 2018), and if repeated periodically but as independent samples or a series of independet samples (e.g., repeated cross-sections), they can be useful to detect trends as well (Verbeek 2008). However, data, that represent the target population, collected by a longitudinal survey offers a broad range of additional benefits which are at least: i) observing individual changes over time, ii) identifying cause-effect chains, iii) distinguishing between age, period and cohort (APC) effects, and iv) analyzing macro changes and their effect on individual behavior (Lynn 2009). All these benefits can be used to study links e.g., between social, economic, and health aspects. In the following, I will provide an example for each one.

One example that shows the benefit of observing individual changes over time is a study based on longitudinal survey data that represents the older target population in Australia. The study has shown that people who experienced a health shock were very likely to drop out from the labor market and thus, retired earlier than those who did not experience a health shock (Zucchelli et al. 2010). This finding means for instance, that it may be worthwhile putting resources into facilitating continued work for people with health shocks and to overcome the problem of an increasing share of people that do not sufficiently contribute to the public pension fund (Jones, Rice and Roberts 2010).

This benefit of observing individual changes over time is closely linked to another benefit of longitudinal survey data, the identification of cause-chain effects (Lynn 2009). For instance, the relationship between health and labor force participation is a complicated one because causality can go both ways: health affects employment status (better health leads to greater participation in the labor force market and poor health leads to lower participation in the labor force market) and labor force participation also affects health (lower paid jobs and higher job insecurity increase the risk of chronic diseases) (Carter et al. 2013). By collecting data from the same individuals over time we can observe the temporal relationship between health and labor force participation and identify cause-chain effects. Thus, it might be worthwhile not to put only resources into facilitating continued work for people with health shocks, but to also invest into health prevention programs especially for low wage income people (Bull et al. 2014; Cunningham, Green and Braun 2018).

Another benefit of longitudinal survey data is that they allow us to separate components of three conflicting effects that all lead to change: effects of age, period, and cohort (APC) (Lynn 2009). In order to separate them, the longitudinal survey data needs to cover a long time span and include different cohorts at a minimum. Considering APC effects is important because changes in people's life may not only expected because they are getting older, i.e., by age; changes in people's life may also depend on the time period they live in, and their birth cohort (Palmore 1978). For instance, income is likely to vary by age (age effect). Young people who are likely to have recently entered the labor market and older people who are likely to have left the labor market probably have a lower income than those who are employed and are between 25 and 65 years old (Eurostat 2020). However, in times of crisis, income is likely to drop; this decrease in income potentially affects all age groups of the working population at the same time, thereby limiting saving options for everyone (period effect). In contrast, if people of a particular birth cohort are generally saving more money than others because they were socialized to do so (e.g., Baby Boomers after World War II vs. Generation X and Y as cohort effects), this saving behavior is present independent from time period or age (DeVaney and Chiremba 2005).

Last but not least, longitudinal survey data allows for ex-ante predictions and ex-post evaluations of macro changes and their effect on individual behavior and decisions

(Lynn 2009). For instance, to cope with future potential state insolvency, a pension reform introduced in 1999 in Germany abolished early retirement for woman born after 1952 at the age of 60. After this reform, women, born after 1952, could not retire before the age of 63 at the earliest. A study of the older population has shown that the introduction of this pension reform in 1999 lead to an increase in their expected retirement age (Etgeton, Fischer and Ye 2019). This finding is further supported by another German study which showed that individuals postpone their retirement on average by approximately 13 months if pension benefits are reduced for each year of retirement (Giesecke 2018). Interestingly the study by Etgeton and his colleagues (2019) showed further, that the individuals on the labor market tend to cope with losses of pension generosity by working longer rather than saving more (Etgeton, Fischer and Ye 2019). This finding may call for putting resources into facilitating continued work for older people in general.

One crucial determinant of missing data in longitudinal surveys that collect data face-to-face are interviewers (Groves and Couper 1998): They are data collectors for the researchers and thus, can impact the data collection process in many ways (Olson et al. 2020). I will explain this relationship in more detail in chapter 1.5.

In order to improve the quality of longitudinal survey data mentioned above, this thesis aims to understand missing data processes. This understanding enables us to implement measures that prevent missing data. To do so, my thesis answers the following research questions:

1. How many initially recruited individuals for a longitudinal survey drop out over 12 years of data collection and do those who drop out differ systematically from those who do not?
2. To what extent do interviewers contribute to the occurrence of missing data in income and can we explain this link between income item nonresponse and interviewers?

3. Does missing data caused by unit nonresponse and income item nonresponse have common causes that can be located with interviewers collecting the data?

Of course, missing data caused by individuals who did not participate in the survey (unit nonresponse) or did not answer certain survey items (item nonresponse) are only two error sources that can compromise the inference from the survey statistic (Groves et al. 2009). Further error sources are entailed in the framework of the *Total Survey Error* (TSE). To place nonresponse into the context of survey errors in general, I will briefly summarize the TSE in the following and will explain how these two errors (unit nonresponse and item nonresponse) are placed within the TSE.

1.2 Evaluating the Quality of Survey Statistics with the *Total Survey Error (TSE) framework*

The *Total Survey Error (TSE) framework* accumulates potential survey errors that may arise in the survey design, the data collection, the processing of data, and the analysis of survey data (see Figure 1.1). This makes the TSE a tool for comprehensive evaluation of survey statistics (Groves et al. 2009).

The *TSE framework* groups various sources of survey errors into two classes: measurement and representation (Groves et al. 2009). Missing data can occur with any source of survey errors. In my thesis I focus on missing data that is caused by nonresponse errors and/or partially by measurement error (see Figure 1.1).

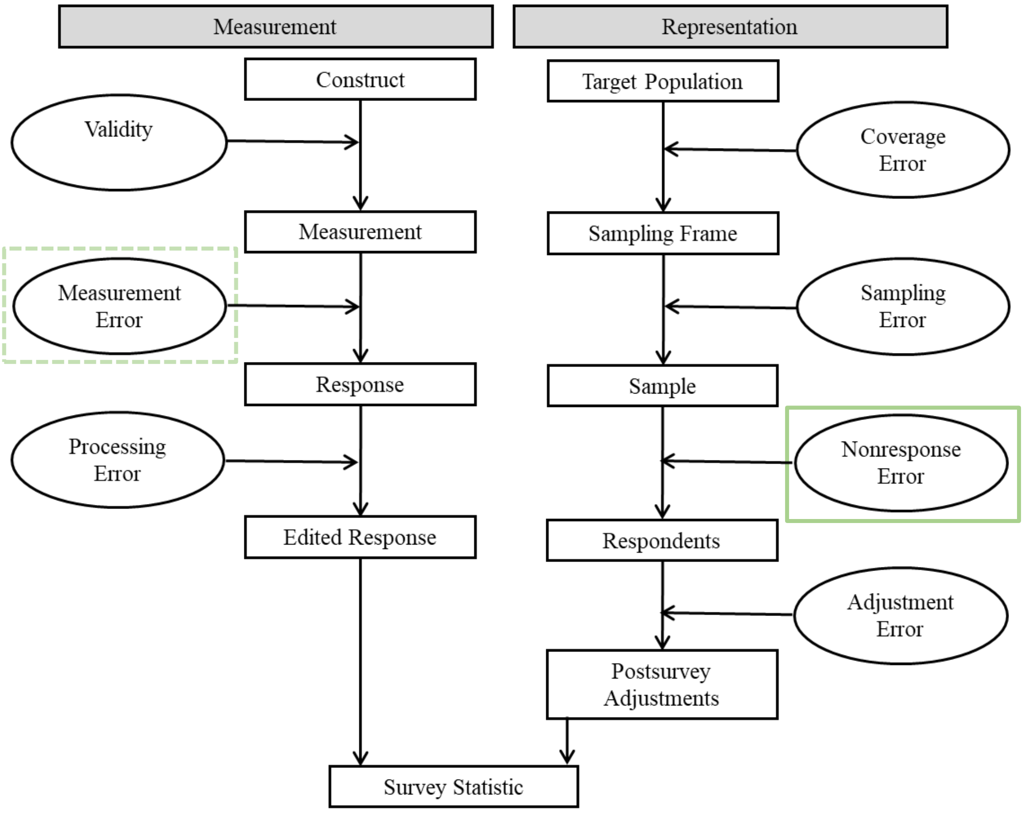

Figure 1.1 Total Survey Error framework. Adapted from Groves et al. (2009, 48).

Nonresponse error belongs to the error class of representation and is defined as sampled subjects not responding to the survey (Groves et al. 2009). For instance, if we are interested in the income distribution of the target population, we need to ensure that all kind of income levels of the target population are represented in the data accordingly. Although characteristics that are associated with nonresponse have varied across studies and target populations, people from low occupational classes and with low educational level have usually been associated with a low response rates (Bianchi and Biffignandi 2019; Lepkowski and Couper 2002; Watson and Wooden 2009). Since a low level of occupational class and education is often linked to a low level of income, we may fail to collect entire survey data from people with low income because they are less likely to participate in surveys (Fitzgerald, Gottschalk and Moffitt 1998; Uhrig 2008), and even people with low income participate in surveys, they are less likely to report it (Bollinger et al. 2019). Nonresponse to single survey items is known as item nonresponse (Dillman et al. 2002).

Measurement error belongs to the error class of measurement. This error describes the difference between the 'true value' of the measurement and the value provided (Groves et al. 2009). Measurement error can for instance occur when respondents adapt their answers in accordance with social norms. Therefore, respondents tend to overreport socially desirable characteristics and tend to underreport socially undesirable characteristics (Krumpal 2013). Other possible respondent reactions are to claim not to know the answer to a question or to refuse to answer (Tourangeau, Rips and Rasinski 2000). Such item nonresponses occur typically with respondents that have a very low or a very high income (Bollinger et al. 2019). They are less likely report their 'true' income than respondents with an average income. Those with a very low or high income may feel uncomfortable reporting their 'true' income and therefore, adapt their answer or do not provide a substantive value.

Any errors of either class can compromise the inference from the survey statistic and, thus, may result in wrong conclusions drawn for the target population. Whether missing data caused by nonresponse and item nonresponse affects estimates is determined by its missing data mechanism which I will explain in the following chapter.

1.3 The Role of Missing Data in Longitudinal Surveys

Whether missing data can introduce bias depends on the causal mechanism underlying the missingness. Moreover, missing data in longitudinal surveys caused by nonresponse can take several forms. The following two sub-chapters will describe the missing data mechanism and types of nonresponse in longitudinal surveys.

1.3.1 Missing data mechanism

The literature distinguishes between three types of missing data mechanisms: *missing completely at random* (MCAR), *missing at random* (MAR), and *missing not at random* (MNAR) (see Figure 1.2). Whether missing data has an effect on the conclusions drawn from the survey statistics depends on the mechanism responsible for the production of the missing data (Little and Rubin 2020).

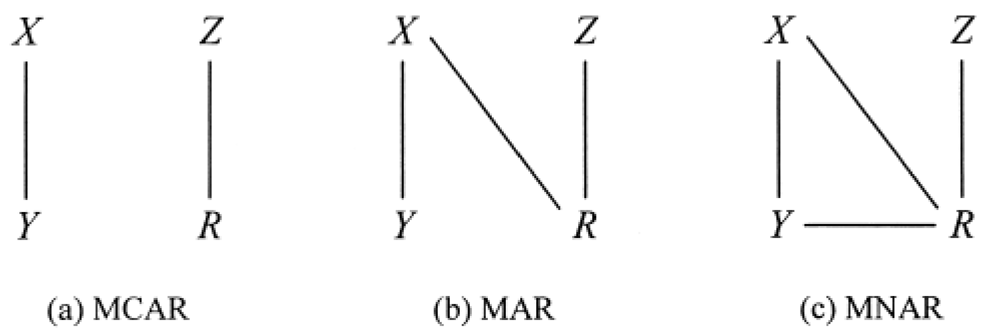

Figure 1.2 Missing data mechanism. Source: Schafer and Graham (2002, 152).

Image (a) in Figure 1.2 describes the MCAR mechanism. In this situation, variable Y is related to variable X and the values of X are known for all survey participants but the values of Y are missing for some. The missingness is named R. The variable Z causes missingness R and is independent of Y and X, and moreover, the missing values are unrelated to values of other observed or any unobserved variables (Schafer and Graham 2002). In other words, there are separate causes for the missing values in Y and for the relationship between X and Y. In that case, the data is described as *missing completely at random* (MCAR) (Little and Rubin 2020). For instance, if we want to study the effect of gender (X) on income (Y), but the income data could not be processed for an individual because of a one-time instrument failure, we may treat such missing data as being completely random: Some of the data is missing by coincidence. The survey estimates are precise and unbiased assuming that the sample size is not heavily reduced (de Leeuw, Hox and Huisman 2003). Therefore, MCAR is also called *ignorable nonresponse*. If such random breakdowns occur more often and lead to a large decrease in sample size, the variance increases which reduces precision and power of estimates and statistical tests (Little and Rubin 2020). The MCAR assumption is a very strong one and often unrealistic in practice (de Leeuw, Hox and Huisman 2003).

Image (b) in Figure 1.2 describes a situation similar to (a), with the important difference in (b) being that the missing values (R) in variable Y depend on variable X. In this case, the data is said to be *missing at random* (MAR) because the missingness depends on X, which is known for all survey participants, but is independent from Y. In other words, the missingness R in Y and the variable Y have a common cause that is dependent on X. MAR is also called *ignorable nonresponse* (Schafer and Graham 2002). MAR requires researchers to take the missing data mechanism into account when they analyze the effect of X on Y. If this requirement is met, survey estimates are precise and unbiased as well for sample sizes that are large enough (Little and Rubin 2020). Using the same example earlier, where we are interested in the effect of gender (X) on income (Y), let us assume this time that we observed that women are less likely to answer income questions than men (Yan, Curtin and Jans 2010). Thus, the probability that income information is missing varies according to the gender of the respondent. However, the probability of income information missing is the same for respondents of the same gender. If we observe and control or adjust for the respondent's gender, the data is MAR because the missingness is related to gender not to income itself (Little and Rubin 2020). Researchers can compare values of those with and without missing data to see if they differ systematically on a specific variable (Allison 2009). If they do, researchers can use appropriate methods after data collection (e.g., applying nonresponse weights or imputations, such as multiple imputations or full maximum likelihood) in order to achieve valid estimates (Edwards, Berzofsky, and Biemer 2017; Lang and Little 2018). It is worth noting, nevertheless, that the application of these methods is not always easy and straightforward (de Leeuw 2001; Kalton and Flores-Cervantes 2003).

In contrast to the MAR mechanism in image (b), image (c) in Figure 1.2 describes the situation where variable Y is directly related to the missingness R. The data is described as *missing not at random* (MNAR). The differences between sampled subjects who provided information and those who did not, lead to biased estimates (Little and Rubin 2002). MNAR is also called *nonignorable nonresponse* (Schafer and Graham 2002). Take again the example of estimating income (Y). If sampled subjects with very high or very low income are less likely to report their income than those with an average income, we will inaccurately estimate income. This is because probability that income is recorded varies according to income, and thus, the missingness R in Y is related to Y. As a result, the survey estimates are biased (Little and Rubin 2020). In the case of MNAR, "no simple solution for treating missing data exists" (de Leeuw, Hox and Huisman 2003, 155).

In survey practice, it is very likely to encounter missing data, especially missing data that is not missing completely at random. In longitudinal surveys in particular, the MNAR mechanism is likely to account for some missing data, such as for income item nonresponse (Frick and Grabka 2014; Giusti and Little 2011) and unit nonresponse (Das, Toepoel and van Soest 2011; van den Berg, Lindeboom and Ridder 1994; van der Zouwen and van Tilburg 2001).

In general, researchers hope to identify and observe variables that render missingness ignorable by correcting or adjusting for the missingness after data collection. However, since the objects of study in surveys, namely humans, are complex social actors, it is difficult to model every missing data mechanism in every variable correctly. Moreover, correcting or adjusting for missing data is not always straightforward or easy (de Leeuw 2001; Kalton and Flores-Cervantes 2003). The use of imputation methods or weighting strategies relies on many assumptions and is not always effective in reducing bias (Graham and Hofer 2000; Little and Vartivarian 2005). Therefore, the prevention of missing data before data collection is a more desirable goal than reduction of missing data after data collection.

To conclude, it is important to study survey errors that are likely to induce bias into survey statistics so that they can be prevented. To prevent missing data caused by nonresponse, it is necessary to identify the different causes of missing data. Only if we understand where, how, and why missing data is generated, we can prevent its occurrence (de Leeuw 2001; Groves and Couper 1998).

1.3.2 Missing data caused by nonresponse

In longitudinal surveys, we can differentiate between different types of nonresponse (e.g., *initial nonresponse*, *wave nonresponse*, *attrition*, *item nonresponse*, *breakoff*, *non-consent*). I will describe each of them in turn.

Data can be missing because the sampled subject did not participate in the survey. In survey methodological literature, this phenomenon is called *unit nonresponse* (Dillman et al. 2002) and it occurs at the contact and cooperation stage. The literature distinguishes between three causes of unit nonresponse: the failure to deliver the survey request (e.g., non-contact or non-location), the refusal to participate, and inability to participate (Groves et al. 2002). In face-to-face surveys, for example, such missing data can occur at the contact stage when no contact with a sampled subject is established, e.g., because the person was never at home when contact attempts were made. At the cooperation stage, for example, missing data can occur when contact with the target person was established but he or she refused to participate, e.g., because he or she had was not interested in participating or had no time to participate.

In longitudinal surveys, unit nonresponse can be further differentiated into *initial nonresponse*, *wave nonresponse*, and *attrition* (Bethlehem, Cobben and Schouten 2011). In the case of *initial nonresponse*, the data of the sample subjects is missing because they were not recruited into the longitudinal survey at the beginning of the study. *Wave nonresponse* occurs when sampled subjects were recruited into the longitudinal survey, but the panel respondents missed one or more subsequent waves (Lepkowski and Couper 2002). *Attrition* occurs if the panel respondents dropped out from a longitudinal survey, i.e., the sampled subjects initially responded to the survey but stopped participating (Bethlehem, Cobben and Schouten 2011). For instance, panel respondents may drop out of the survey because they have moved and cannot be relocated. Another possible reason for attrition, one that is likely to occur in longitudinal surveys of older populations, is death.

Another cause for missing data is the participation of a sampled subject in a survey, but their refusal or inability to answer one or more questions. This is commonly called *item nonresponse* (Dillman et al. 2002). The literature distinguishes between three sources of item nonresponse: the failure to obtain an answer because the respondent does not know the answer or does not want to respond, the failure to obtain an answer because the response is not codable within the response options, and the failure to obtain an answer due to technical problems that prevent from entering the answer in the system or processing the response data (de Leeuw, Hox and Huisman 2003).

Similar to item nonresponse, no consent to a within-survey request is source of missing data (Sakshaug et al. 2012). Some surveys request consent to link the survey data with external data sources, such as administrative data. If respondents do not agree with such linkage, this type of nonresponse is known as *non-consent*. While a respondent not responding to a single question generates only few missing data, a respondent who has started an interview but decides to stop mid-interview may generate a larger amount of missing data. Such interviews are also referred to as *breakoff* in survey methodology (Peytchev 2009).

Unit nonresponse and *item nonresponse* are typically treated as "two separate problems with different impacts on data quality, different statistical treatments and adjustments, and different underlying causes" (Yan and Curtin 2010, 535). However, some studies have found evidence of an interconnection between *unit nonresponse* and *item nonresponse* (e.g., Campanelli, Sturgis and Purdon 1997; Couper 1997; Loosveldt, Pickery and Billiet 2002).

The *response continuum model* describes the interconnection between unit nonresponse and item nonresponse. In this model, respondents are placed on a continuum based on their propensity to participate in a survey and to answer survey questions (Yan and Curtin 2010). Respondents with zero propensity to participate in a survey are placed on the left of the continuum and those with a high propensity to participate in a survey and answer all questions are on the right of the continuum (Figure 1.3). From left to right on the continuum, the respondent's propensity to participate in a survey increases as well as their propensity to answer survey questions (Yan and Curtin 2010).

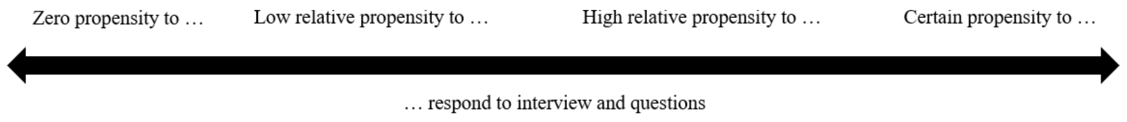

Figure 1.3 The response continuum model. Adapted from Yan and Curtin (2010, 536).

According to this *response continuum model*, the relation between unit nonresponse and item nonresponse can be bi-directional (Yan and Curtin 2010). Researchers who studied item nonresponse and wave nonresponse found that panel members who were more likely to provide missing data in an earlier wave are less likely to participate in the next wave (Loosveldt, Pickery and Billiet 2002). This relation has been also observed with the focus on income item nonresponse and nonresponse to the entire survey in longitudinal surveys (e.g., Schräpler 2004; Frick and Grabka 2005; Taylor 2006; Müller and Castiglioni 2015). Other researchers who studied the link between unit nonresponse and item nonresponse found that sampled subjects who were less likely to participate in the survey were more likely to provide missing data (Campanelli, Sturgis and Moon 1996; Couper 1997). This reciprocity of the response continuum is important to be considered in longitudinal studies. If this relationship of unit nonresponse and item nonresponse and their causes were well understood, missing data caused by nonresponse could be efficiently reduced.

Missing data poses a challenge particularly in longitudinal surveys, which are designed to collect data from the same subjects over time, because missing data makes it impossible to measure changes between waves (Lynn 2009). Observing changes are important to answer many research questions since it is certain events that change the people’s life thereby influence many other decisions. Missing data on changes is especially challenging when changes in people's life are correlated with missing data and panel respondents who were not observed differ from those who were observed (Lynn et al. 2005). For example, changes in people's life often coincidence with relocation. Older people are likely to experience health shock or bereavement. These life events often result in relocating to a nursing home or moving in with relatives (Stoeckel and Porell 2010). If the move of a sampled subject is not properly registered and followed-up the probability to relocate and recontact the subject decreases once the sampled subject has moved (Watson 2020). And even if the follow-up process of such subjects is adequate, they may be less cooperative due to the experienced shock (Riley et al. 1972). Thus, life events increase the probability of unobserved changes. Consequently, certain life events and their impact on changes in the people's life may be underestimated, when those who experienced life events are more likely to drop out of the study than those who have not.

In conclusion, all types of nonresponse may lead to bias of estimates. If missing data was completely randomly occurring, conclusions drawn from the survey statistics would be unbiased. To reduce missing data, in particular missing data that is likely to occur non-randomly, we do not only need to identify its origin and causes, but also to understand the determinants of missing data in surveys in order to prevent them in future studies. To do so, researchers can use the *conceptual framework for survey cooperation* developed by Groves and Couper (1998). This framework can be used to identify the determinants of missing data which provide the basis for modelling missing data mechanisms and with which we can reduce, or even prevent, missing data. This framework can be applied independently of whether the missing data has been generated randomly or non-randomly.

1.4 Determinants of Missing Data

We can identify three main determinants of missing data for surveys: the sampled subject, the social environment, and the survey design. For surveys that collect data by means of interviewers, the interviewer plays a crucial role as well (see Figure 1.4).

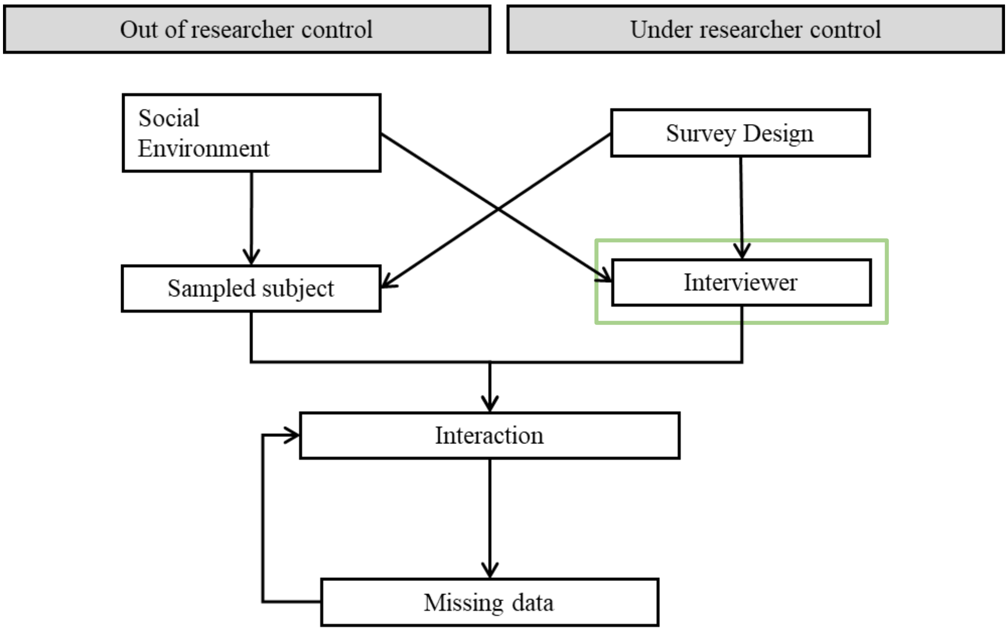

Figure 1.4 Conceptual framework for missing data. Adapted from Groves and Couper (1998, 30).

The interviewer is a key determinant in surveys that collect data face-to-face because the interviewer receives instructions from the researcher and executes numerous tasks on the researcher's behalf. Another determinant of missing data is the survey design, including questionnaire design (question formulation, answer options, instructions, routings, layout, length, etc.), interview mode, and use of incentives (Groves and Couper 1998). These determinants are under the researchers' control. If the survey design and the interviewer meet the sampled subject's needs and align with the respondent's social environment, the probability of missing data can be reduced. Note, however, while the survey design can be controlled by the researcher, individual and household characteristics as well as the social environment cannot.

For instance, some people are less likely to trust strangers than others (Li and Fung 2012). Therefore, achieving cooperation with those who do not trust strangers may be more difficult than with those who trust strangers. In order to reduce this potential cause for missing data, researchers usually send the sampled subjects study materials beforehand to help respondents familiarize themselves with the survey and potential interview partners. If researchers inform about their study and about interviewers that will get in contact with the respective respondents, the sampled subjects will expect interviewers that knock at their door or call and ask to schedule an interview. Previous information, such as advance letters, may reduce the probability of missing data because they can increase the probability of successful contact and cooperation (Toepoel and Schonlau 2017).

Another example of a study characteristic that can be influenced by researchers, so that the probability of missing data decreases, is the questionnaire design. For example, we know that information about income and assets are hard to collect because income and assets are complex and sensitive topics (Tourangeau and Yan 2007). In order to collect such complex and sensitive data interviewers could therefore ask questions that allow respondents to range their amount of income and assets instead of providing exact numbers (Pleis, Dahlhamer and Meyer 2007). Moreover, when sampled subjects mention concerns about having their income information disclosed to taxation authorities, we could train interviewers to highlight the strict data protection regulations of the study to increase trust, thereby decreasing the probability of missing data in income.

Another example of meeting the sampled subjects needs and thereby decrease the probability of missing data is to split the interview into two or three parts and allowing assistance of relatives when necessary (Rodgers and Herzog 1992). For instance, older people are more likely to have health problems (Milanović et al. 2013), and less healthy people are more likely to drop out of studies (Fitzgerald, Gottschalk and Moffitt 1998;

Halliday and Kimmitt 2008), because they may feel not strong enough for an interview or are unable to answer all questions. Interviewers can conduct partial or assisted interviews if rest or assistance are necessary and thereby reduce the risk of this type of missing data. Additionally, this flexibility may avoid breakoffs because it may lower the burden for respondents with health problems.

Implementing a survey with as little as missing data as possible is therefore more likely, when researchers take into account the given social environment and characteristics of sampled subjects when designing the survey. The other determinant of missing data that is under researcher control, is the interviewer. Consequently, the interviewer is one of the most significant parts of this thesis.

1.5 The Role of the Interviewer in the Missing Data Generation Process

Interviewers are the key agent in face-to-face surveys because they interact with the sampled subject and execute many tasks on the researchers' behalf (Groves and Couper 1998). The role of the interviewer can be divided in two parts within the data collection process: the first part describes the recruiting process, when interviewers approach the sampled subjects or re-approach panel members of a longitudinal survey, i.e., contact and try to motivate respondents to participate. The second part describes the interview process itself, that includes the question-answer-interaction with the respondents, the coding and editing of answers, and the transfer of the data. Another task that may be fulfilled by interviewers is obtaining consent for additional survey requests from respondents, e.g., for physical tests, and/or for linkage of the survey data with other data sources, e.g., pension insurance data. During their work interviewers can impact data collection and thereby survey outcomes and/or survey statistics (Olson et al. 2020).

Interviewers having an impact on data is called the interviewer effect (Kish 1962). Interviewer effects occur when interviewers induce a dependency between survey outcomes and/or survey statistics. As a result, the data collected by "either a specific individual interviewer or a specific set of interviewers may be different than data collected by another individual or set of interviewers administering the same questionnaire to a sample from the same population of respondents" (Davis et al. 2010, 15). Such interviewer effects increase the variances of estimates and thus, reduce the precision of estimates. Moreover, if this variation induced by the interviewers is systematic, estimates may be biased (Davis et al. 2010).

Furthermore, interviewer effects are considered to vary in their extent and direction. For some survey tasks, such as recruiting respondents or conducting physical tests, researcher assume that interviewers are helpful and increase the probability of participation (Campanelli, Sturgis and Purdon 1997; Groves et al. 2009) whereas for others, such as obtaining answers to sensitive questions, researchers assume that interviewers can be disruptive and reduce the probability of reporting the 'true' value (Davis et al. 2010; Krumpal 2013). Interviewer effects in factual questions are presumed and found to be rather small (Mangione, Fowler and Louis 1992; Schnell and Kreuter 2005) and rather large in asking for consent (Korbmacher and Schroeder 2013; Sakshaug et al. 2017). Several other studies have additionally shown, that interviewers can – positively and negatively – affect survey outcomes and survey statistics, some to a larger extent than others (see Olson et al. 2020; West and Blom 2017).

In the context of missing data and longitudinal studies, interviewer effects occur in form the interviewers having an interviewer-specific influence on the emergence of unit nonresponse, wave nonresponse, and/or item nonresponse (West and Blom 2017). Interviewer effects in wave nonresponse, for instance, manifest in the panel member's willingness to cooperate being dependent on the interviewer who approached the respondent. In other words, panel member X may not cooperate when approached by interviewer A but may cooperate if approached by interviewer B. In the case of item nonresponse, interviewer effects mean that the respondent's choice to answer or not to answer a particular question is dependent on the interviewer who conducted the interview. Hence, respondent X may have answered differently if interviewed by interviewer A than interviewed by interviewer B. But why do some interviewers obtain more missing data than others?

The amount of missing data individual interviewers collect differs because they are human agents and not "'neutral' collectors of data" (Pickery and Loosveldt 2001, 338). Different interviewers have different characteristics, expectations, personality, and attitudes. Their individual behavior and other individual differences may lead to different outcomes when executing the task of data collection (Groves and Couper 1998), i.e., different cooperation, consent, and/or item response rates across interviewers. For instance, some interviewers may feel uncomfortable with asking sensitive questions (Ongena and Haan 2020) and are more willing to accept "don't know" as an answer than other interviewers. As a result, some interviewers may collect more item nonresponse than others.

Researchers have developed conceptual frameworks to explain interviewer effects on multiple survey errors (Dijkstra and van der Zouwen 1987; Japec 2007; Reinecke and Schmidt 1993; van der Zouwen, Dijkstra and Smit 1991). A more recent conceptual model for understanding interviewer effects by West and Blom (2017) explains interviewer effects with background characteristics of the interviewer, such as sociodemographic characteristics, experience, workload, monitoring and trainings can explain interviewer effects. Interviewer attitudes, personality, beliefs and expectations, behavior and skills and respondent's activation of stereotypes/perceived norms potentially mediate the relationship between background characteristics of the interviewer and their influence on survey outcomes, whereas question, respondent, and interviewer features potentially moderate the relationship (Figure 1.5).

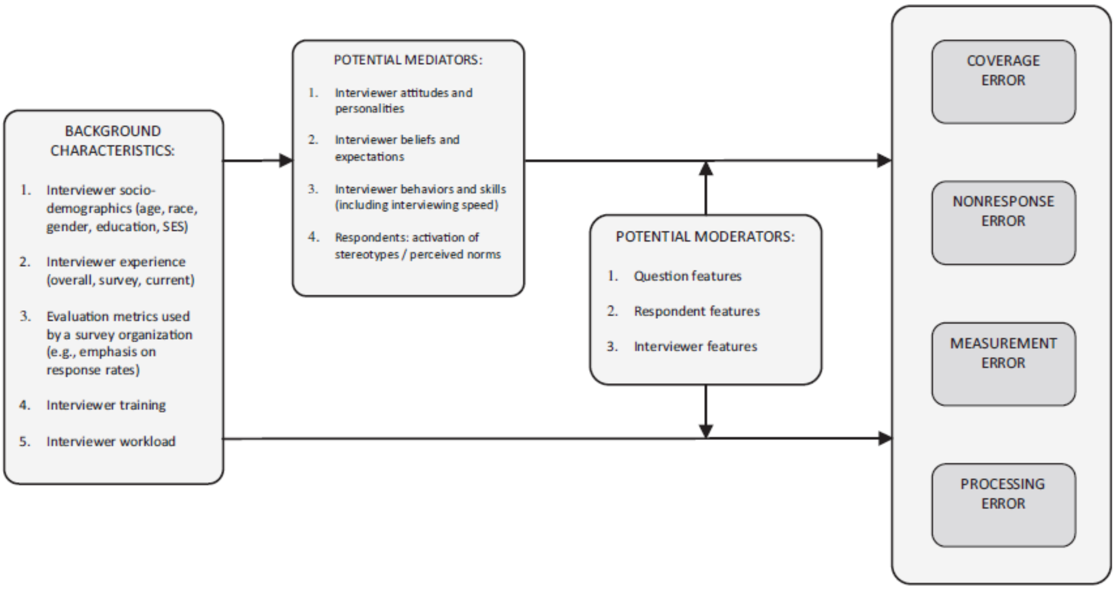

Figure 1.5 Conceptual model to explain interviewer effects on multiple survey errors. Source: West and Blom (2017, 196).

So far, no study on interviewer effects could test this framework in empirical settings because of the limited available data on interviewers. There are, however, many studies that have investigated the association between single, individual interviewer characteristics and survey outcomes and/or survey statistics (see Olson et al. 2020 and West and Blom 2017). The majority of studies either focuses on nonresponse error or measurement error. In the latter case, researchers prioritize differences in substantive responses over item nonresponse. The synthesis of West and Blom (2017) concluded, that interviewer attitudes seem to be a promising predictor of survey outcomes and/or survey statistics because they found, for example, that interviewers with more positive attitudes tend to achieve better survey outcomes.

Explaining interviewer effects is necessary to understand how and why interviewers impact data. This insight provides the basis for the assumptions on missing data mechanisms. Additionally, researchers usually train their interviewers to avoid missing data (Daikeler and Bosnjak 2020; Groves and McGonagle 2001). In interview trainings, interviewers learn how to recruit respondents, how to handle the survey-specific instrument, and other survey-relevant administrative procedures. After having completed their trainings, interviewers can fulfill their task, i.e., they are able to recruit sampled subjects and conduct interviews. Researchers can provide further manuals, instructions and, guidelines, such as contact attempt time sheets or FAQs for interviewers (Daikeler et al. 2017) to minimize the risk of missing data, and other survey errors, caused by interviewers as much as possible.

These trainings can be complemented with interviewer trainings to equip all interviewers with standardized interviewing techniques, such as keeping to the interview script (i.e., reading out questions *verbatum*), probing inadequate answers, or taking a neutral stand towards the respondents (Fowler and Mangione 1990). These standardized interviewer trainings aim at removing the possibility of variation caused by the interviewer, for instance some interviewers collecting more missing data than others (Fowler 2004; Fowler and Mangione 1990; Garbarski, Schaeffer and Dykema 2016). This again is a critical contribution to further reduce the risk of errors in survey statistics. As well-trained interviewers have proven to be a vital part of a successful survey, standardized interviewer trainings are common in large-scale surveys (Mneimneh et al. 2019).

However, the general and standardized interviewer trainings and further interviewer materials are not necessarily universally applicable and cannot capture all potential situations interviewers will encounter (Billiet and Loosveldt 1988). Special trainings where interviewers are trained on additional interview methods can be introduced to complement standardized trainings and prepare interviewers for more particular, difficult situations they may encounter (Maynard and Schaeffer 2002). One example is special training on refusal avoidance (Laurie, Smith and Scott 1999). Refusal conversation trainings aim to equip interviewers with techniques that allow them to persuade sampled subjects that are reluctant. The combination of special interviewer trainings and standardized interviewer trainings can reduce the probability of missing data (Schnell and Trappmann 2006).

In practice, interviewers decide whether or not, and in which way and to what extent, they apply trained methods. Particularly, experienced interviewers may also identify the best strategy to avoid missing data. Applying the best strategy for recruiting is known as "tailoring" (Groves and Couper 1998). This technique is "employed by expert interviewers who respond to cues in their immediate setting – verbal, nonverbal, and visual – to produce utterances and behavior that respond to sample persons' utterances and behavior" (Schaeffer, Dykema and Maynard 2010, 447–448). For instance, if interviewers encounter a sampled subject that is reluctant, they need to identify which method of their trained techniques may be useful to persuade the respective individual to participate in the survey. Moreover, while communicating with the sampled subject, interviewers need to maintain the conversation and interact. For example, if respondents are in a hurry, interviewers may offer callback appointments to overcome the sampled subjects time constraints and to avoid nonresponse. This technique is known as "maintaining interaction" (Groves and Couper 1998). Both techniques require interviewers to use their trained techniques and adapt them where necessary and always according to the respondent's need. At the same time, they must not violate the standardized interview regulations and maintain the comparability of the collected data as much as possible. Therefore, maintaining interaction and tailoring is vital in reducing the risk of missing data. These techniques can be trained and acquired throughout an interviewer's career.

In conclusion, missing data can occur at many steps of the data collection process. Moreover, within the *Total Survey Error (TSE) framework*, nonresponse is only one of numerous error sources that cause missing data and that can impact the survey statistics and introduce bias into estimates if the missing data mechanism is not completely random. Since any survey error is complex, researchers should carefully examine each of them and to piece the puzzle of the TSE together in order to evaluate the quality of survey statistics. My thesis focuses on a small piece of this puzzle, namely missing data in longitudinal surveys that is caused by unit nonresponse and income item nonresponse. The importance of my thesis and the approach it takes is manifold. First, missing data is crucial to study, as it is likely to occur non-randomly. Second, it is particularly important to study missing data in the context of longitudinal surveys, as it would allow researchers to benefit from the full range of opportunities that data from longitudinal surveys can offer. Third, it is worthwhile to explore the link between missing data and interviewers in face-to-face surveys, as it allows us to focus on preventing missing data from occurring, thereby avoiding the need to resort to complex corrections and adjustments for missing data after data collection. Interviewers are the key agents in face-to-face data collection and researchers have some control over how they conduct interviews, e.g., through interviewer trainings. Using the appropriate interviewer training designs, we may be able to prevent missing data, in particular. Having a nuanced understanding of missing data that is likely to bias analysis based on longitudinal surveys and of interviewer-based methods of preventing missing data, survey researchers and social scientist can improve the quality of longitudinal survey statistics.

1.6 Outline of the Thesis

My dissertation consists of three studies that examine missing data caused by nonresponse in a longitudinal face-to-face survey, namely the Survey of Health Ageing and Retirement in Europe (SHARE). Applying the *Total Survey Error (TSE) framework*, I focus on nonresponse error and measurement error, both of which generate missing data and can cause bias in survey statistics. Using the *conceptual framework for survey cooperation*, I examine the role of the interviewer in the missing data generation process. Moreover, in accordance with the *response continuum model*, I connect interviewer effects with unit nonresponse and item nonresponse. The goal of the three studies is to improve longitudinal survey data by understanding missing data that is likely to introduce bias.

The first study (chapter 2) focuses on attrition and answers the research question "How many initially recruited individuals for a longitudinal survey drop out over 12 years of data collection and do those who drop out differ systematically from those who do not?". Attrition is one type of nonresponse error and according to the TSE framework, it can impact survey statistics. If those who attrite are systematically different from those who continue participating in the survey, this attrition can bias estimates based on the data. The reasons for attrition are manifold, but since this study is based on the data of the first interviewed SHARE panel members and since SHARE is a longitudinal survey of an older population, there is a greater risk of attrition due to death, a factor which is additionally considered in this study. Moreover, investigating nonresponse in nine countries, the first study in my thesis provides a multi-country perspective.

The second article (chapter 3) focuses on item nonresponse and answers the research question "To what extent do interviewers contribute to the occurrence of missing data in income and can we explain this link between income item nonresponse and interviewers?". Item nonresponse can be conceptualized as nonresponse error or measurement error within the TSE framework. Regardless of how it is conceptualized, item nonresponse is likely to be high for income survey data, as those who have a very high income and those who have a very low income are likely to not report their income. The high item nonresponse means that income estimates are very likely to be biased. Since the *conceptual framework for survey cooperation* has shown that interviewers are key agents in face-to-face surveys, interviewers are of particular interest in this article. Furthermore, the role of interviewers in SHARE is special, since they interact with a broad range of the older population, ranging from those who are 55–65 years old and are still employed, to retirees, to the elderly and oldest old with poor health. This study uses SHARE data from five countries that recruited new samples in 2015 as well as data on interviewers' sociodemographic characteristics, attitudes, and expectations that were collected through an additional online survey. The interviewer survey provides the opportunity to link data from interviewers and respondents in order to study interviewer effects. This study covers five countries that recruited new panel members and successfully conducted the interviewer survey.

The third study (chapter 4) sheds light on the link between the interviewer and wave nonresponse and income item nonresponse at the same time. In this study, the research question is "Does missing data caused by unit nonresponse and income item nonresponse have common causes that can be located with interviewers collecting the data?". In addition to focusing on the interviewer as a key agent in face-to-face surveys, the study views the data collection process viewed within the *response continuum model*. This model states that the respondent's propensity to participate in a survey is correlated with the propensity to answer survey questions. Like the second study of my thesis, this study incorporates data from the interviewer survey. However, in contrast to the second study, which focused on item nonresponse, this one focuses on wave nonresponse and item nonresponse. By investigating item nonresponse and unit nonresponse in four countries, this last study of my thesis offers a multi-country perspective as well.

References

Allison, Paul D. 2009. "Missing data." In *The SAGE Handbook of Quantitative Methods in Psychology*, edited by Roger E. Millsap and Alberto Maydeu-Olivares, 72–89. Thousand Oaks, CA: SAGE.

Bergmann, Michael, Giuseppe De Luca, and Annette Scherpenzeel. 2017. "Sampling Design and Weighting Strategies in SHARE Wave 6." In *SHARE Wave 6: Panel innovations and collecting Dried Blood Spots*, edited by Frederic Malter and Axel Börsch-Supan, 77–90. Munich, Germany: Munich Center for the Economics of Aging, Max Planck Institute for Social Law and Social Policy.

Bethlehem, Jelke, Fannie Cobben, and Barry Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. New York: John Wiley.

Bianchi, Annamaria, and Silvia Biffignandi. 2019. "Social Indicators to Explain Response in Longitudinal Studies." *Social Indicators Research* 141 (3): 931–957. https://doi.org/10.1007/s11205-018-1874-7.

Billiet, Jacques, and Geert Loosveldt. 1988. "Improvement of The Quality of Responses to Factual Survey Questions by Interviewer Training." *Public Opinion Quarterly* 52 (2): 190–211. https://doi.org/10.1086/269094.

Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. 2019. "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch." *Journal of Political Economy* 127 (5): 2143–2185. https://doi.org/10.1086/701807.

Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie M. Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology* 42 (4): 992–1001. https://doi.org/10.1093/ije/dyt088.

Bull, Eleanor R., Stephan U. Dombrowski, Nicola McCleary, and Marie Johnston. 2014. "Are Interventions for Low-income Groups Effective in Changing Healthy Eating, Physical Activity and Smoking Behaviours? A Systematic Review and Meta-analysis." *BMJ Open* 4 (11): 1–9. http://dx.doi.org/10.1136/bmjopen-2014-006046.

Campanelli, Pamela, Patrick Sturgis, and Nick Moon. 1996. "Exploring the Impact of Survey Introduction." *Proceedings of the Section on Survey Research Methods 2*:1054–1059. Alexandria, VA: American Statistical Association.

Campanelli, Pamela, Patrick Sturgis, and Susan Purdon. 1997. *Can You Hear Me Knocking? An Investigation into the Impact of Interviewers on Survey Response Rates.* London: Survey Methods Centre, Social and Community Planning Research, University of London.

Carter, Kristie N., Fiona I. Gunasekara, Tony Blakely, and Ken Richardson. 2013. "Health Shocks Adversely Impact Participation in the Labour Force in a Workinig Age Population: A Longitudinal Analysis." *Australian and New Zealand Journal of Public Health* 37(3): 257–263. https://doi: 10.1111/1753-6405.12068.

Ciol, Marcia A., Jeanne M. Hoffman, Brian J. Dudgeon, Anne Shumway-Cook, Kathryn M. Yorkston, and Leighton Chan. 2006. "Understanding the Use of Weights in the Analysis of Data From Multistage Surveys." *Archives of Physical Medicine and Rehabilitation* 87 (2): 299–303. https://doi.org/10.1016/j.apmr.2005.09.021.

Couper, Mick P. 1997. "Survey Introductions and Data Quality." *Public Opinion Quarterly* 61 (2): 317–338. https://doi.org/10.1086/297797.

Cunningham, Peter J., Tiffany L. Green, and Robert T. Braun. 2018. "Income Disparities in the Prevalence, Severity, and Costs of Co-occurring Chronic and Behavioral Health Conditions." *Medical Care* 56 (2): 139–145. https://doi.org/10.1097/MLR.0000000000000864.

Daikeler, Jessica, and Michael Bosnjak. 2020. "How to Conduct Effective Interviewer Training: A Meta-Analysis and Systematic Review." In *Interviewer Effects from a Total Survey Error Perspective*, edited by Kristen Olson, Jolene D. Smyth, Jennifer Dykema, Allyson L. Holbrook, Frauke Kreuter, and Brady T. West, 46–56. New York: Chapman and Hall/CRC.

Daikeler, Jessica, Henning Silber, Michael Bosnjak, Anouk Zabal, and Silke Martin. 2017. *A General Interviewer Training Curriculum for Computer-Assisted Personal Interviews (GIT-CAPI)*. Mannheim: GESIS – Leibniz-Institute for the Social Sciences.

Das, Marcel, Vera Toepoel, and Arthur van Soest. 2011. "Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys." *Sociological Methods & Research* 40 (1): 32–56. https://doi.org/10.1177/0049124110390765.

Davies, Richard B. 1994. "From cross-sectional to longitudinal analysis." In *Analyzing Social and Political Change: A Casebook of Methods*, edited by Angela Dale and Richard B. Davies, 20–40. Thousand Oaks, CA: SAGE.

Davis, Rachel E., Mick P. Couper, Nancy K. Janz, Cleopatra H. Caldwell, and Ken Resnicow. 2010. "Interviewer Effects in Public Health Surveys." *Health Education Research* 25 (1): 14–26. https://doi.org/10.1093/her/cyp046.

de Leeuw, Edith D. 2001. "Reducing Missing Data in Surveys: An Overview of Methods." *Quality and Quantity* 35 (2): 147–160. https://doi.org/10.1023/A:1010395805406.

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

de Leeuw, Edith D., Joop J. Hox, and Annemieke Luiten. 2018. "International Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data." *Survey Insights: Methods from the Field.* https://doi.org/10.13094/SMIF-2018-00008. Available at: https://surveyinsights.org/?p=10452.

De Luca, Giuseppe. 2018a. *Flexible Stata Code for Hot-Deck Imputations of Non-Monetary Variables in SHARE.* Available at: www.seriss.eu/resources/deliverables (accessed August 2020).

De Luca, Giuseppe. 2018b. *Flexible Stata Code for Multiple FCS Imputations of Monetary Variables in SHARE.* Available at: www.seriss.eu/resources/deliverables (accessed August 2020).

DeVaney, Sharon A., and Sophia Chiremba. 2005. "Comparing the Retirement Savings of the Baby Boomers and Other Cohorts." *Monthly Labor Review* March, U.S. Department of Labor, Bureau of Labor Statistics Washington, DC.

Dijkstra, Wil, and Johannes van der Zouwen. 1987. "Styles of Interviewing and the Social Context of the Survey-Interview." In *Social Information Processing and Survey Methodology. Recent Resaerch in Psychology*, edited by Hans-Jürgen Hippler, Norbert Schwarz, and Seymour Sudman, 200–211. New York: Springer.

Dillman, Don A., John L. Eltinge, Robert M. Groves, and Roderick J. A. Little. 2002. "Survey Nonresponse in Design, Data Collection, and Analysis." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 3–26. New York: John Wiley.

Edwards, Susan L., Berzofsky, Marcus E., and Paul P. Biemer. 2017. "Effect of Missing Data on Classification Error in Panel Surveys." *Journal of Official Statistics* 33 (2): 551–570.

Etgeton, Stefan, Björn Fischer, and Han Ye. 2019. "The Effect of Increasing the Early Retirement Age on Savings Behavior Before Retirement." *IZA Discussion Paper No. 12744*. Bonn, Germany: Institute for the Study of Labor (IZA).

Eurostat. 2020. Available at: https://ec.europa.eu/eurostat/en/web/products-datasets/-/ILC_DI03 (accessed October 2020).

Fienberg, Stephen E., and Judith M. Tanur. 1986. "The Design and Analysis of Longitudinal Surveys: Controversies and Issues of Cost and Continuity." In *Survey Research Designs: Towards a Better Understanding of Their Costs and Benefits: Prepared under the Auspices of the Working Group on the Comparative Evaluation of Longitudinal Surveys Social Science Research Council*, edited by Robert W. Pearson and Robert F. Boruch, 60–93. New York: Springer.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *The Journal of Human Resources* 33 (2): 251–299. https://doi.org/10.2307/146433.

Fowler, Floyd J. Jr. 2004. "Reducing Interviewer-Related Error Through Interviewer Training, Supervision, and Other Means." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymor Sudman, 259–278. New York: John Wiley.

Fowler, Floyd J. Jr., and Thomas W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Thousand Oaks, CA: SAGE.

Frick, Joachim R., and Markus M. Grabka. 2005. "Item Nonresponse on Income Questions in Panel Surveys: Incidence, Imputation and the Impact on Inequality and Mobility. " *Allgemeines Statistisches Archiv* 89 (1): 49–61. https://doi.org/10.1007/s101820500191.

Frick, Joachim R., and Markus M. Grabka. 2014. "Missing Income Data in the German SOEP: Incidence, Imputation and its Impact on the Income Distribution." *SOEP Survey Papers No.225*. Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Garbarski, Dana, Nora C. Schaeffer, and Jennifer Dykema. 2016. "Interviewing Practices, Conversational Practices, and Rapport:Responsiveness and Engagement in the Standardized Survey Interview." *Sociological Methodology* 46 (1): 1–38. https://doi.org/10.1177/0081175016637890.

Giesecke, Matthias. 2018. "The Effect of Benefit Reductions on the Retirement Age: The Heterogeneous Response of Manual and Non-Manual Workers." *Review of Income and Wealth* 64 (1): 213–238. https://doi.org/10.1111/roiw.12257.

Giusti, Caterina, and Roderick J. A. Little. 2011. "An Analysis of Nonignorable Nonresponse to Income in a Survey with a Rotating Panel Design." *Journal of Official Statistics* 27 (2): 211–229.

Gough, Orla, Roberta Adami, and James Waters. 2008. "The Effects of Age and Income on Retirement Decisions: A Comparative Analysis between Italy and the UK." *Pensions: An International Journal* 13 (3): 167–175. https://doi.org/10.1057/pm.2008.12.

Graham, John W., and Scott M. Hofer. 2000. "Multiple Imputation in Multivariate Research." in *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples,* edited by Todd D. Little, Kai U. Schnabel, and Jürgen Baumert, 201–218, 269-281. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview*. New York: John Wiley.

Groves, Robert M., Don A. Dillman, John L. Eltinge, and Roderick J. A. Little. 2002. *Survey Nonresponse*. New York: John Wiley.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. New York: John Wiley.

Groves, Robert M., and Katherine A. McGonagle. 2001. "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation." *Journal of Official Statistics* 17 (2): 249–265.

Groves, Robert M., and George J. Schoeffel. 2018. "Use of Administrative Records in Evidence-Based Policymaking." *The Annals of the American Academy of Political and Social Science* 678 (1): 71–80. https://doi.org/10.1177/0002716218766508.

Halliday, Timothy, and Michael Kimmitt. 2008. "Selective Migration and Health." *IZA Discussion Paper No. 3458*. Bonn, Germany: Institute for the Study of Labor (IZA).

Hand, David J. 2020. *Dark Data: Why What You Don't Know Matters.* Oxford: Princeton University Press.

Haziza, David, and Jean-François Beaumont. 2017. "Construction of Weights in Surveys: A Review." *Statistical Science* 32 (2): 206–226. https://doi.org/10.1214/16-STS608.

International Conference on Methodology of Longitudinal Surveys (MOLS). 2006. Report available by Peter Lynn (2008), *Bulletin de Méthodologie Sociologique*, 97:48–65.

Japec, Lilli 2007. "Interviewer Error and Interviewer Burden." In *Advances in Telephone Survey Methodology,* edited by James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 185–211. New York: Wiley.

Jones, Andrew M., Nigel Rice, and Jennifer Roberts. 2010. "Sick of Work or Too Sick to Work? Evidence on Self-reported Health Shocks and Early Retirement from the BHPS." *Economic Modelling* 27 (4): 866–880. https://doi.org/10.1016/j.econmod.2009.10.001

Kalton, Graham, and Ismael Flores-Cervantes. 2003. "Weighting Methods." *Journal of Official Statistics* 19 (2): 81–97.

Kesmodel, Ulrik S. 2018. "Cross-sectional Studies – What Are They Good for?" *Acta Obstetricia et Gynecologica Scandinavica* 97 (4): 388–393. https://doi.org/10.1111/aogs.13331.

Kish, Leslie. 1962. "Studies of Interviewer Variance for Attitudinal Variables." *Journal of the American Statistical Association* 57 (297): 92–115. https://doi.org/10.1080/01621459.1962.10482153.

Klevmarken, N. Anders, Bengt Swensson, and Patrick Hesselius. 2005. "The SHARE Sampling Procedures and Calibrated Design Weights." In *The Survey of Health, Aging, and Retirement in Europe – Methodology*, edited by Axel Börsch-Supan and Hendrik Jürges, 28–69. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging.

Knoll, Melissa A. Z. 2011. "Behavioral and Psychological Aspects of the Retirement Decision." *Social Security Bulletin* 71 (4): 15–32.

Korbmacher, Julie M., and Mathis Schroeder. 2013. "Consent when Linking Survey Data with Administrative Records: The Role of the Interviewer." *Survey Research Methods* 7 (2): 115–131. https://doi.org/10.18148/srm/2013.v7i2.5067.

Krumpal, Ivar. 2013. "Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review." *Quality & Quantity* 47 (4): 2025–2047. https://doi.org/10.1007/s11135-011-9640-9.

Laaksonen, Mikko, Ritva Prättälä, Ville Helasoja, Antti Uutela, and Eero Lahelma. 2003. "Income and Health Behaviours. Evidence from Monitoring Surveys among Finnish Adults." *Journal of Epidemiology and Community Health* 57 (9): 711–717. https://doi.org/10.1136/jech.57.9.711.

Laird, Nan M. 1988. "Missing data in longitudinal studies." *Statistics in Medicine* 7 (1-2): 305–315.

Lang, Kyle M., and Todd D. Little. 2018. "Principled Missing Data Treatments" *Prevention Research* 19:284–294. https://doi.org/10.1007/s11121-016-0644-5.

Laurie, Heather, Rachel Smith, and Lynne Scott. 1999. "Strategies for Reducing Nonresponse in a Longitudinal Panel Survey." *Journal of Official Statistics* 15 (2): 269–282.

Lepkowski, James M., and Mick P. Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little, 259–272. New York: John Wiley.

Li, Tianyuan, and Helene H. Fung. 2012. "Age Differences in Trust: An Investigation Across 38 Countries." *The Journals of Gerontology: Series B* 68 (3): 347–355. https://doi.org/10.1093/geronb/gbs072.

Little, Roderick J. A., and Donald B. Rubin. 2020. *Statistical Analysis with Missing Data*. New York: John Wiley.

Little, Roderick J., and Sonya Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31 (2): 161–168.

Loosveldt, Geert, Jan Pickery, and Jaak Billiet. 2002. "Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey." *Journal of Official Statistics* 18 (4): 545–557.

Lynn, Peter. 2009. "Methods for Longitudinal Surveys." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 1–19. Chichester, UK: John Wiley.

Lynn, Peter, Nicholas Buck, Jonathan Burton, Annette Jäckle, and Heather Laurie. 2005. "A Review of Methodological Research Pertinent to Longitudinal Survey Design and Data Collection. " *Working Papers of the Institute for Social and Economic Research, paper 2005-29.* Colchester, UK: University of Essex.

Majcen, Špela. 2017. "Evidence Based Policy Making in the European Union: The Role of the Scientific Community." *Environmental Science and Pollution Research* 24 (9): 7869–7871. https://doi.org/10.1007/s11356-016-6247-7.

Mangione, Thomas W., Floyd J. Fowler Jr., and Thomas A. Louis. 1992. "Question Characteristics and Interviewer Effects." *Journal of Official Statistics* 8 (3): 293–293.

Maynard, Douglas W., and Nora C. Schaeffer. 2002. "Refusal Conversation and Tailoring." In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, edited by Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora C. Schaeffer, and Johannes van der Zouwen, 219–243, New York: Wiley.

Milanović, Zoran, Saša Pantelić, Nebojša Trajković, Goran Sporiš, Radmila Kostić, and Nic James. 2013. "Age-related Decrease in Physical Activity and Functional Fitness among Elderly Men and Women." *Clinical Interventions in Aging* 8:549–556. https://doi.org/10.2147/CIA.S44112.

Mneimneh, Zeina, Lars E. Lyberg, Sharan Sharma, Mahesh Vyas, Dhananjay B. Sathe, Frederic Malter, and Yasmin Altwaijri. 2019. "Case Studies on Monitoring Interviewer Behaviour in International and Mutinational Surveys." In *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A. L. Stoop, and Brita Dorer, 731–770, Hoboken, NJ: John Wiley.

Müller, Bettina, and Laura Castiglioni. 2015. "Attrition im Beziehungs- und Familienpanel pairfam." In *Nonresponse Bias Qualitätssicherung sozialwissenschaftlicher Umfragen*, edited by Jürgen Schupp und Christof Wolf, 383–408, Wiesbaden, Germany: Springer.

Olson, Kristen, Jolene D. Smyth, Jennifer Dykema, Allyson L. Holbrook, Frauke Kreuter, and Brady T. West. 2020. *Interviewer Effects from a Total Survey Error Perspective*. New York: Chapman and Hall/CRC.

Ongena, Yfke, and Marieke Haan. 2020. "Differences in Interaction Quantity and Conversational Flow in CAPI and CATI Interviews." In *Interviewer Effects from a Total Survey Error Perspective*, edited by Kristen Olson, Jolene D. Smyth, Jennifer Dykema, Allyson L. Holbrook, Frauke Kreuter, and Brady T. West, 164–175. New York: Chapman and Hall/CRC.

Palmore, Erdman. 1978. "When Can Age, Period, and Cohort be Separated?" *Social Forces* 57 (1): 282–295. https://doi.org/10.2307/2577639.

Pampel, Fred C., Patrick M. Krueger, and Justin T. Denney. 2010. "Socioeconomic Disparities in Health Behaviors." *Annual Review of Sociology* 36:349–370. https://doi.org/10.1146/annurev.soc.012809.102529.

Peytchev, Andy. 2009. "Survey Breakoff." *Public Opinion Quarterly* 73 (1): 74–97. https://doi.org/10.1093/poq/nfp014.

Pickery, Jan, and Geert Loosveldt. 2001. "An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse." *Journal of Official Statistics* 17 (3): 337–350.

Pleis, John R., James M. Dahlhamer, and Peter S. Meyer. 2007. "Unfolding the Answers? Income Nonresponse and Income Brackets in the National Health Interview Survey." *Proceedings of the 2006 Joint Statistical Meetings:*3540–3547. Alexandria, VA: American Statistical Association.

Reinecke, Jost, and Peter Schmidt. 1993. "Explaining Interviewer Effects and Respondent Behavior: Theoretical Models and Empirical Analysis." *Quality and Quantity* 27 (3): 219–247. https://doi.org/10.1007/BF01097151.

Riley, Matilda White, Marilyn Johnson, Anne Foner, John A. Clausen, Richard Cohn, Beth Hess, Robert K. Merton, et al. 1972. *Aging and Society, Volume 3: A Sociology of Age Stratification*. New York: Russell Sage Foundation.

Rodgers, Willard L., and A. Regula Herzog. 1992. "Collecting Data about the Oldest Old: Problems and Procedures." In *The Oldest Old,* edited by Richard M. Suzman, David P. Willis, and Kenneth G. Manton, 135–156. New York: Oxford University Press.

Sakshaug, Joseph W., Mick P. Couper, Mary B. Ofstedal, and David R. Weir. 2012. "Linking Survey and Administrative Records: Mechanisms of Consent." *Sociological Methods & Research* 41 (4): 535–569. https://doi.org/10.1177/0049124112460381.

Sakshaug, Joseph W., Sebastian Hülle, Alexandra Schmucker, and Stefan Liebig. 2017. "Exploring the Effects of Interviewer- and Self-Administered Survey Modes on Record Linkage Consent Rates and Bias." *Survey Research Methods* 11 (2): 171–188. https://doi.org/10.18148/srm/2017.v11i2.7158.

Särndal, Carl-Erik, and Sixten Lundström. 2005. *Estimation in Surveys with Nonresponse*. Chichester, UK: John Wiley.

Schaeffer, Nora C., Jennifer Dykema, and Douglas W. Maynard. 2010. "Interviewers and Interviewing." In *Handbook of Survey Research,* edited by Peter V. Marsden and James D. Wright, 437–470. Bingley, UK: Emerald.

Schafer, Joseph L., and John W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7 (2): 147–177. https://doi.org/10.1037/1082-989X.7.2.147.

Schnell, Rainer, and Mark Trappmann. 2006. "The Effect of the Refusal Avoidance Training Experiment on Final Disposition Codes in the German ESS-2." *Working Paper No. 3*. Konstanz: Center for Quantitative Methods and Survey Research University of Konstanz, Germany.

Schnell, Rainer, and Frauke Kreuter. 2005. "Separating Interviewer and Sampling-Point Effects." *Journal of Official Statistics* 21 (3): 389–410.

Schräpler, Jörg-Peter. 2004. "Respondent Behavior in Panel Studies: A Case Study for Income Nonresponse by Means of the German Socio-Economic Panel (SOEP)." *Sociological Methods & Research* 33 (1): 118–156. https://doi.org/10.1177/0049124103262689.

Stoeckel, Kimberly J., and Frank Porell. 2010. "Do Older Adults Anticipate Relocating? The Relationship Between Housing Relocation Expectations and Falls." *Journal of Applied Gerontology* 29 (2): 31–50. https://doi.org/10.1177/0733464809335595.

Taylor, Rebecca. 2006. "Using Item Non-response to Predict Future Unit-nonresponse: ELSA as a Case Study." Paper presented at the International Conference on Methodology of Longitudinal Surveys, Colchester, UK, July.

Toepoel, Vera, and Matthias Schonlau. 2017. "Dealing with Nonresponse: Strategies to Increase Participation and Methods for Postsurvey Adjustments." *Mathematical Population Studies* 24 (2): 79–83. https://doi.org/10.1080/08898480.2017.1299988.

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

Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133 (5): 859–883. https://doi.org/10.1037/0033-2909.133.5.859.

Uhrig, S. C. Noah. 2008. "The Nature and Causes of Attrition in the British Household Panel Survey." ISER Working Paper Series No. 2008-05. Colchester, UK: University of Essex, Institute for Social and Economic Research. Available at: https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2008-05 (accessed August 2020).

van den Berg, Gerard J., Maarten Lindeboom, and Geert Ridder. 1994. "Attrition in Longitudinal Panel Data and the Empirical Analysis of Dynamic Labour Market Behaviour." *Journal of Applied Econometrics* 9 (4): 421–435. https://doi.org/10.1002/jae.3950090408.

van der Zouwen, Johannes, Wil Dijkstra, and Johannes H. Smit. 1991. "Studying Respondent-Interviewer Interaction: The Relationship between Interviewing Style, Interviewer Behavior, and Response Behavior." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymor Sudman, 419–437, New York: John Wiley.

van der Zouwen, Johannes, and Theo van Tilburg. 2001. "Reactivity in Panel Studies and its Consequences for Testing Causal Hypotheses." *Sociological Methods & Research* 30 (1): 35–56. https://doi.org/10.1177/0049124101030001003.

Verbeek, Marno. 2008. "Pseudo-Panels and Repeated Cross-Sections." In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, edited by László Mátyás and Patrick Sevestre, 369–383. Berlin, Heidelberg: Springer Berlin Heidelberg.

Watson, Nicole. 2020. "Measuring Geographic Mobility: Comparison of Estimates from Longitudinal and Cross-Sectional Data." *Survey Research Methods* 14 (1): 1–18. https://doi.org/10.18148/srm/2020.v14i1.7422.

Watson, Nicole, and Mark Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 157–181. Chichester, UK: John Wiley.

West, Brady T., and Annelies G. Blom. 2017. "Explaining Interviewer Effects: A Research Synthesis." *Journal of Survey Statistics and Methodology* 5 (2): 175–211. https://doi.org/10.1093/jssam/smw024.

Yan, Ting, and Richard Curtin. 2010. "The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective." *International Journal of Public Opinion Research* 22 (4): 535–551. https://doi.org/10.1093/ijpor/edq037.

Yan, Ting, Richard Curtin, and Matthew Jans. 2010. "Trends in Income Nonresponse Over Two Decades." *Journal of Official Statistics* 26 (1): 145–164.

Zucchelli, Eugenio, Anthony Harris, Andrew M. Jones, and Nigel Rice. 2010. "The Effects of Health Shocks on Labour Market Exits: Evidence from The HILDA Survey." *Australian Journal of Labour Economics* 13 (2): 191–218.

2. Evolution of the Initially Recruited SHARE Panel Sample Over the First Six Waves[1]

[1] This chapter has been published as Friedel, Sabine, and Tim Birkenbach. (2020). "Evolution of the Initially Recruited SHARE Panel Sample Over the First Six Waves." *Journal of Official Statistics* 36 (3): 507–527. https://doi.org/10.2478/jos-2020-0027.

Abstract

Attrition is a frequently observed phenomenon in panel studies. The loss of panel members over time can hamper the analysis of panel survey data. Based on data from the Survey of Health, Ageing and Retirement in Europe (SHARE), this study investigates changes in the composition of the initially recruited first-wave sample in a multi-national face-to-face panel survey of an older population over waves. By inspecting retention rates and R-indicators, we found that, despite declining retention rates, the composition of the initially recruited panel sample in Wave 1 remained stable after the second wave. Thus, after the second wave there is no further large decline in representativeness with regard to the first wave sample. Changes in the composition of the sample after the second wave over time were due mainly to mortality-related attrition. Non-mortality-related attrition had a slight effect on the changes in sample composition with regard to birth in survey country, area of residence, education, and social activities. Our study encourages researchers to investigate further the impact of mortality- and non-mortality-related attrition in multi-national surveys of older populations.

Key words: R-indicator; wave nonresponse; mortality- and non-morality-related attrition; panel sample composition

2.1 Introduction

Panel surveys of older populations in Europe have become the focus of widespread interest in recent decades. Falling fertility rates (Myrskylä, Goldstein, and Cheng 2013) and greater life expectancy (Leon 2011) bring many challenges for Western European societies. To investigate these dynamic processes, researchers need data that allow them to provide evidence of changes over time (Olsen 2018). In contrast to cross-sectional surveys, panel surveys fulfil this requirement because they repeatedly collect data from the same respondents over time (Lynn 2009).

However, a major detracting feature of panel surveys is the risk of attrition – that is, the loss of panel members from the initially recruited sample over time (Binder 1998). Panel attrition is a frequent phenomenon that has been observed during the last decades (Fitzgerald, Gottschalk, and Moffitt 1998; Watson 2003; Buck et al. 2006). Attrition may occur because panel members are no longer able or willing to participate or because they can no longer be located or contacted (Lynn and Lugtig 2017). The largest amount of drop out occurs in the second wave (Watson and Wooden 2009; Schoeni et al. 2013). When attrition occurs, changes over time cannot be observed from the beginning to the end of the panel because one measure is missing in two consecutive waves (Lynn and Lugtig 2017). This absence of data can lead to restrictions when researchers want to analyze changes in the data. Thus, we need to inform researchers about attrition in the data they use.

Particularly in panel surveys of older populations, researchers are faced with a greater risk of attrition due to death. In an investigation of characteristics associated with attrition in the English Longitudinal Study of Ageing (ELSA) and the U.S. Health and Retirement Study (HRS), Banks, Alastair, and Smith (2011) found that the mortality rate between two waves among panel members aged 70–80 years was 15 percent, and that among 55–64 year-old panel members it was four percent. In contrast, for the Panel Study of Income Dynamics (PSID), which is a household panel survey, Watson (2003) reported a mortality rate of only 0.5 percent between two waves. Thus, the risk of mortality-related attrition is much higher in panel surveys of older populations compared to those that collect data on younger populations.

Deaths in panel surveys of older populations are not problematic per se. Older populations are not fixed, and all older populations are affected by deaths (Smith, Lynn, and Elliot 2009). Deaths occur both in the population and in the sample, and thus deaths of panel members change the composition of the data sample and of the population about which researchers want to draw conclusions. In both settings – the population and the sample – individuals who have a lower risk of dying, for example because they have a high socioeconomic and health status, are more likely to survive to old age than individuals with a low socioeconomic and health status (Banks, Alastair, and Smith 2011). Thus, we assume that mortality in panel surveys of older populations is selective. However, deaths reflect changes in the composition of the population to which the data refer, and, as Smith, Lynn, and Elliot (2009, 29) noted, “as long as these [deaths] can be identified and distinguished from nonresponse, they are easily incorporated in analyses by using a code for dead units.”

In contrast to mortality-related attrition, respondents who drop out for other reasons are still present in the population, and their non-participation changes only the composition of the sample. Changes in these individuals’ outcomes of interest can no longer be observed in the survey data, although they are occurring in the population. However, this type of attrition is not problematic per se, either, unless it is selective, and thus can affect the validity and interpretation of estimates (Watson and Wooden 2019).

The present study focuses on the Survey of Health, Ageing and Retirement in Europe (SHARE) (Börsch-Supan et al. 2013), a biennial panel study based on people in Europe aged 50 years and older. With its harmonized collection of data in many European countries, SHARE is unique and offers many opportunities to analyze dynamic processes in the European societies. Although previous research has shown that attrition occurs in the SHARE panel (Bergmann et al. 2019), little research has investigated in more detail the changes in the composition of the initially recruited panel sample over time (e.g., Bristle et al. 2019). Moreover, little is known about the relation between attrition and the changes in the panel composition over waves when mortality is particularly considered. Both aspects can inform researchers about the impact of attrition on the evolution of the SHARE panel.

To obtain a clear picture of how the composition of the SHARE panel has evolved over waves, we define two samples of interest:

- A: the initially recruited SHARE sample (i.e., the sample first interviewed in Wave 1)
- B: the initially recruited SHARE sample, excluding respondents who were reported to have died.

Whereas Sample A is fixed over waves and includes all respondents who dropped out, Sample B is dynamic over waves and excludes for each wave separately respondents who were reported to have died before the corresponding wave started. For instance, Sample B in Wave 2 is based on the initially recruited SHARE sample, excluding respondents who were reported to have died before the second wave started, or Sample B in Wave 3 is based on the initially recruited SHARE sample, excluding respondents who were reported to have died before the third wave started. Thus, Sample A investigates total attrition (non-mortality-related and mortality-related), whereas Sample B investigates non-mortality-related attrition only.

With these two definitions of the samples of interest, we aim to answer the following research questions:

1. How has the initially recruited first-wave sample (A and B) evolved over the survey waves?
2. Has the evolution of the initially recruited first-wave sample (A and B) over waves varied across countries?
3. What variables/characteristics have played the most important role in the evolution over waves of the sample that excludes reported deaths (Sample B)?

The remainder of this article is organized as follows: In the next section, we describe our SHARE dataset and the variables considered in our analyses. We then answer Research Questions 1 and 2 by applying two aggregate-level measures (retention rates, R-indicators). In section 4, we apply two variable-level measures (subgroup retention rates, logistic regressions) to answer Research Question 3. Thus, the methods and results for the first two research questions and the methods and results for the third question are presented separately. The article concludes with a summary of the findings and discussion for all three research questions.

2.2 Data and Variables

2.2.1 Data

We used data from the Survey of Health, Ageing and Retirement in Europe (SHARE) (Börsch-Supan 2017)[2]. SHARE is a biennial multidisciplinary, cross-national panel survey that collects microdata on the health, socioeconomic status, and social and family networks of individuals aged 50 years and older and of their partners, regardless of their

[2] This article uses data from SHARE Wave 1 (DOIs: 10.6103/SHARE.w1.600). The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

age. The target persons and their partners are interviewed face-to-face using computer-assisted personal interviewing (CAPI) (Börsch-Supan et al. 2013). The first wave of SHARE was conducted in 2004 in 11 European countries and in Israel. Samples from each country are based on a probability sample that is representative of the non-institutionalized population aged 50 years and older (De Luca, Rosetti, and Malter 2013). The initial individual response rates (RR1, American Association for Public Opinion Research, AAPOR, 2016) ranged between 27.9 and 58.8 percent (Bergmann et al. 2019).

For our analyses, we used the first-wave data about respondents' individual and household characteristics and supplemented these data with information about whether or not the respondents had participated in later waves. We restricted our sample to countries that participated in all six observed waves. This selection criterion reduced the sample to nine countries (Austria, Belgium, Denmark, France, Germany, Italy, Spain, Sweden, and Switzerland). Moreover, we restricted our sample to respondents aged 50 years or older. Together, these restrictions decreased the sample to 21,227 panel respondents (table 2.1, Respondents aged 50+). About five percent of the respondents could not be considered because they did not know or refused to report the answer to questions that were used to measure variables included in the analyses. As a consequence, the first analysis sample of Sample A consisted of 20,236 respondents. The sample size by country ranged from 898 in Switzerland to 3,521 in Belgium (see table 2.1, Analysis Sample A).

To study further non-mortality-related attrition, we excluded respondents who were reported to have died before a given wave. This exclusion resulted in a dynamic Analysis Sample B (see table 2.1, Analysis Sample B, Wave 1– Wave 6). However, the quality of information we used to identify deaths differs between countries. This is due mainly to the fact that most European countries lack a national mortality register or similar records. Therefore, SHARE cannot reliably ascertain the vital status of nonrespondents who drop out because they cannot be located or contacted or because they refuse to be re-interviewed (Bergmann et al. 2019). Thus, the dynamic Analysis Sample B may include unreported deaths.

Table 2.1. Sample selection of initially recruited first-wave SHARE respondents

Country	Respondents aged 50+	Analysis Sample A	Analysis Sample B					
	Wave 1	Wave 1-6	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Austria	1,516	1,487	1,487	1,442	1,361	1,287	1,213	1,174
Belgium	3,631	3,521	3,521	3,474	3,356	3,237	3,120	3,017
Denmark	1,597	1,527	1,527	1,480	1,390	1,310	1,220	1,134
France	2,955	2,706	2,706	2,650	2,519	2,428	2,298	2,221
Germany	2,909	2,768	2,768	2,718	2,648	2,545	2,508	2,486
Italy	2,495	2,406	2,406	2,353	2,268	2,189	2,081	1,984
Spain	2,232	2,075	2,075	1,984	1,884	1,769	1,655	1,547
Sweden	2,961	2,848	2,848	2,778	2,640	2,486	2,349	2,268
Switzerland	931	898	898	882	860	839	816	788
Total	21,227	20,236	20,236	19,761	18,926	18,090	17,260	16,619

2.2.2 Variables

Investigating the evolution of the SHARE panel offered the possibility of including a rich set of variables in the models. To examine the evolution of the panel, we selected 23 first-wave key variables from the areas of demographics, social embeddedness, health, and economics, and three survey-specific variables of the questionnaire design (see table 2.2).

When selecting variables to investigate changes in the composition of the initially recruited sample over waves, care was taken to ensure that they represented the main publication domains, related to key survey items, and/or related to survey-specific motives for nonresponse (Schouten, Shlomo, and Skinner 2011).

Table 2.2. Operationalization of information used to examine the evolution of the SHARE panel

Variable	Operationalization
Demographics	
Gender	0: male; 1: female
Age	1: 50–59 years; 2: 60–69 years; 3: 70–79 years; 4: 80+ years
Born in survey country	1: yes; 0: no
Education level	1: low; 2: medium and other; 3: high
Household (HH) size	1: 1-person HH; 2: 2-person HH; 3: 3+-person HH
Partner in HH	0: no; 1: yes
Area of residence	1: city/large town; 2: small town; 3: rural village
Social embeddedness variables	
Residential proximity of child(ren)	1: no children; 2: child living in household; 3: child living ≤1 km away; 4: child living >1 km away
Social activities	0: no activities; 1: at least one activity
Received help from others	0: no; 1: yes
Gave help to others	0: no; 1: yes
Health variables	
Health status	0: good or better; 1: fair or poor
Chronic diseases	0: none; 1: at least one chronic disease
Depression (Euro-D)	0: no or insufficient symptoms; 1: 4 or more depressive symptoms
Maximum grip strength	1: item nonresponse; 2: 1st quartile; 3: 2nd quartile; 4: 3rd quartile; 5: 4th quartile
Memory recall ability	0: recalled less than half of the words; 1: recalled more than half of the words
Hospital overnight stays in last 12 months	0: no; 1: yes
Currently smoking	0: no; 1: yes
Currently drinking	0: never; 1: less than once a week; 2: 1-6 times a week; 3: daily
Limitation of instrumental activities of daily living (IADL)	0: no IADL limitation; 1: at least one IADL limitation
Economic variables	
Employment status	1: retired; 2 working; 3: not working and other
Make ends meet	0: difficulties; 1: no difficulties
Total household income	1: item nonresponse; 2: 1st quartile; 3: 2nd quartile; 4: 3rd quartile; 5: 4th quartile
Interview process variables	
Financial respondent	0: no; 1: yes
Family respondent	0: no; 1: yes
Household respondent	0: no; 1: yes

We included sociodemographic and socioeconomic variables (gender, age, education, citizenship, number of children, and income) in our models. Researchers have used these individual characteristics in almost all models for their substantive analyses based on SHARE data (SHARE-ERIC 2018). Additionally, some of these variables have been found to predict attrition in SHARE (Bristle et al. 2019). As Bristle et al. (2019) showed that item nonresponse to financial questions in SHARE negatively affected cooperation in the next wave, we supplemented the income quartiles with an additional category indicating that respondents did not answer the household income question.

We also included information on household composition, area of residence, employment status, and making ends meet, because this information has been widely used in economic research (SHARE-ERIC 2018) and has been found to predict cooperation in SHARE (Bristle et al. 2019). We included several key health variables that have been extensively used in the literature because researchers have also used SHARE data to study health (SHARE-ERIC 2018). Moreover, research has shown that persons with poor health tend to cooperate less than healthy persons (Bristle et al. 2019). Our selection of health variables included self-assessed health, chronic diseases, depression symptoms (Euro-D), limitations of instrumental activities of daily living (IADL), smoking and drinking behavior, and two objective health measurements/tests (grip strength and recall memory). As SHARE data are also used by researchers in the field of family and social networks, well-being, and charity, we included information on the spatial proximity of children, giving help to others, and receiving help from others. Additionally, as the literature shows that being socially active can predict cooperation in longitudinal studies (Bianchi and Biffignandi 2019), information on the number of social activities was also included.

Furthermore, research has shown that respondent burden in the previous SHARE wave influenced cooperation in the next wave (Bristle et al. 2019). In SHARE, selected household members serve as so-called family, financial, or household respondents and answer specific questions on behalf of the whole household. Being selected for one of these roles means that the duration of the interview is usually longer than average and that the respondent provides more information. To capture this respondent burden, we selected three interview process variables (financial, family, and household respondent).

2.3 Evolution of the SHARE Panel Sample Over Waves and Across Countries

2.3.1 Analytical approach

Addressing Research Questions 1 and 2, we examined changes in the composition of the initially recruited SHARE sample over waves and across countries by calculating retention rates and estimating R-indicators for Analysis Samples A and B (the latter excludes reported deaths before the start of the corresponding wave and potentially includes unreported deaths). To investigate changes in the sample composition over waves, we coded participation for each wave. We denoted by y_i the outcome for respondent *i* as follows:

$$y_i = \begin{cases} 0 \ no\ participation \\ 1 \quad\ participation \end{cases}, \tag{1}$$

where participation y_i equals 1 if respondent *i* participated in the survey and 0 otherwise.

The retention rates in the present study measured the proportion of respondents who participated in each wave, conditional upon having participated in the first wave. The R-indicator (where "R" stands for representativeness) was originally designed to measure the degree to which the respondents in a sample resemble the total target population or gross sample (Schouten, Cobben, and Bethlehem 2009). By contrast, the R-indicators in our study measured the degree to which the respondents in Analysis Sample A resemble the initially recruited first-wave respondents over waves, and the degree to which respondents in the dynamic Analysis Sample B resembles the initially recruited first-wave respondents over waves but excluding respondents who were reported to have died before a given wave.

Researchers have used R-indicators to assess the extent to which a net sample is representative of the target population or a gross sample. For instance, data of recruited samples have been compared with census, administrative, or population register data (e.g., Moore, Durrant, and Smith 2016; Schouten et al. 2012; Luiten and Schouten 2013;

Roberts, Vandenplas, and Stähli 2014). R-indicators can also be used as indicators for representativeness in panel studies (Schouten et al. 2012). Bianchi and Biffignandi (2017) used R-indicators to compare the panel sample of the UK household longitudinal study Understanding Society over waves with administrative data to assess population representativeness. In sum, they showed that R-indicators were a valuable measure of representativeness.

R-indicators are estimated as follows (Schouten, Cobben, and Bethlehem 2009):

$$\hat{R}_{\hat{\rho}} = 1 - 2\,\hat{S}_{\hat{\rho}}, \tag{2}$$

where $\hat{S}_{\hat{\rho}}$ is the estimated standard deviation of the individual response propensities. Therefore, the R-indicator is a measure of variation in response propensities. The estimated R-indicator $\hat{R}_{\hat{\rho}}$ ranges between 1 and 0, where the value 1 denotes strong representativeness and the value 0 denotes the maximum deviation from strong representativeness.

Our approach differed from that of Schouten, Cobben, and Bethlehem (2009) with respect to the meaning of the term "representativeness." Schouten, Cobben, and Bethlehem (2009) designed R-indicators to assess the extent to which a net sample was representative of the total target population or a gross sample, whereas we used R-indicators to compare the composition of the initially recruited sample in Wave 1 of SHARE with the composition of the sample in subsequent waves, including any recruitment bias that might have existed in the original sample. The main advantage of our approach was that a rich set of individual-level data could be used rather than the sparse data that are available at population level. For our analyses of the evolution of the panel sample, all information already provided by the participants in the first wave could be used. This approach allowed for the detection of systematic dropout from the panel with respect to many important and substantive survey variables, and not only with respect to a few demographic variables available at the population level.

Thus, we adapted Schouten and colleagues' concept (2009) to examine changes in the composition of the initially recruited SHARE sample over waves. We defined a panel response subset of variables X as "fully representative" if the average propensity to participate again over these categories of X was constant for all possible values of X (Equation 2). For Analysis Sample A, samples in later waves were "fully representative" if their propensities to participate again were equal over the categories of X. As a consequence, the distributions of the selected respondent and household characteristics X remained identical as in the first observed wave. For the dynamic Analysis Sample B, samples in later waves were "fully representative" if their propensities to participate again were equal over the categories of X when reported deaths before a given wave were excluded. As a consequence, the distributions of the selected respondent and household characteristics X remained identical as in the first observed wave excluding reported deaths before a given wave. The estimated R-indicator $\hat{R}_{\hat{\rho}}$ (Equation 2) in our study also ranged between 1 and 0. However, 1 means no change in the composition of the original sample and 0 means total change. Confidence intervals for each R-indicator in each wave were estimated at the five percent level.

The probability that the R-indicators would reach high values differed for our two analysis samples. We expected that the exclusion of reported deaths in the Analysis Samples B would lead to higher R-indicator values for the dynamic Analysis Sample B compared to the fixed Analysis Sample A because we assumed that respondents who dropped out because they died belonged to a selective group of respondents. In contrast, if we had perfect response or if we had equal response propensities over waves, the value of the R-indicator of both analysis samples (A and B) would remain at 1.

To estimate the R-indicators, we used a specially adapted tool provided by the Representative Indicators for Survey Quality Project (RISQ 2015). In more detail, to compute R-indicators, we used a version of Version 2.1 of RISQ that was adapted for our purposes by the RISQ team. As RISQ recommends that representativeness be analyzed by using categorical information rather than continuous information, we applied a categorical approach to describe and explore the evolution of the SHARE panel. We fitted several R-indicator models with the 26 selected variables based on participation outcome as the dependent variable. First, we estimated overall R-indicators for all countries (Analysis Sample A). Second, we estimated overall R-indicators that excluded reported deaths before a given wave for all countries (Analysis Sample B) to focus on non-mortality-related attrition. Third, we estimated the R-indicator based on Analysis Sample A and the R-indicator based on the dynamic Analysis Sample B for each country separately.

2.3.2 Results

To answer the first research question as to how the composition of the initially recruited first-wave sample evolved over waves, we calculated retention rates and estimated R-indicators for each wave, averaged across all countries.

The overall retention rate of Analysis Sample A declined almost linearly over the waves from 69 percent to 42 percent (see fig. 2.1), with a kink at the first follow-up interview. Around 30 percent of the initially recruited first-wave respondents (Analysis Sample A) did not participate in the second wave. Also in the case of the R-indicator (Analysis Sample A), the largest decrease in the value was observed from the first to the second wave (−.16). However, in contrast to the retention rate, the R-indicator (Analysis Sample A) decreased weakly over time afterwards. After six waves, Analysis Sample A reached an R-indicator value of .72. Thus, after the second wave, no further large decline in representativeness of the initially recruited first-wave sample and only few changes in the sample composition were observed.

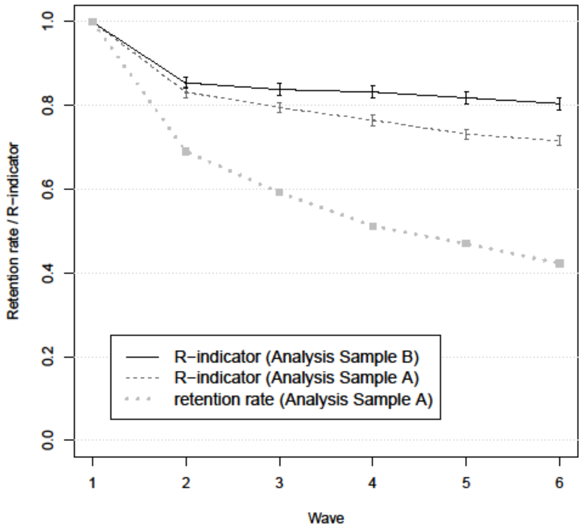

Figure 2.1 Evolution of the initially recruited SHARE sample over waves

Comparing the R-indicator for Analysis Sample A with that for Analysis Sample B, where we excluded reported deaths, we saw that the R-indicators of Analysis Sample B followed the same trend over waves as of Analysis Sample A – a substantial decrease in value after the first wave, and relatively stable values after the second wave. Moreover, we noted that the R-indicators for Analysis Sample B differed significantly from that of Analysis Sample A (see fig. 2.1). After six waves, the R-indicator for – and thus the representativeness of – Analysis Sample A was .72, whereas the R-indicator for the dynamic Analysis Sample B was .80. Thus, a decline in retention rate is not automatically linked to strong changes in the sample composition. In particular, when we eliminated the selective mortality-related attrition in Analysis Sample B, the representativeness of the sample was reasonably strong.

To answer Research Question 2 as to whether the evolution of the initially recruited sample over waves differed across countries, we calculated retention rates and estimated R-indicators for each country separately. Overall, the same pattern of declining retention rates and stabilizing R-indicators after the second wave was observed (see fig. 2.2). Retention rates in Analysis Sample A ranged from 55 to 75 percent across countries in Wave 2 and from 24 to 50 percent in Wave 6. By contrast, the values of the R-indicators in Wave 2 ranged across countries from .76 to .85 for Analysis Sample A and from .77 to .86 for Analysis Sample B. At the last observed wave (Wave 6), R-indicators ranged across countries from .61 to .74 for Analysis Sample A and from .69 to .85 for Analysis Sample B. Despite the fact that the gap between retention rates and R-indicators varied across countries, the observed pattern of change in the composition of the initially recruited first-wave sample (A and B) measured by R-indicators tended to be similar for all countries.

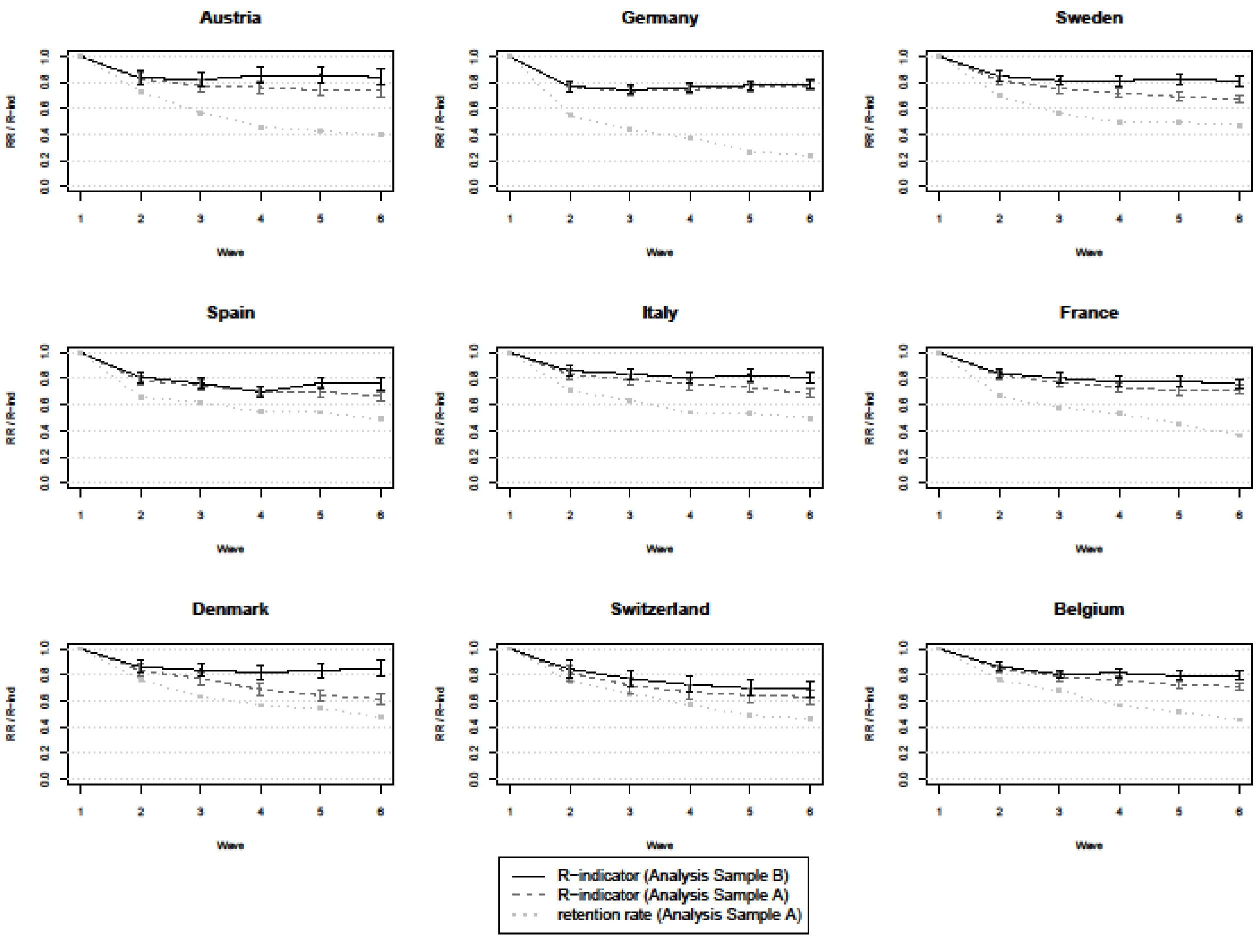

Figure 2.2. Evolution of the initially recruited SHARE sample over waves, by country

2.4 Variable-Level Analysis of Non-mortality-related Attrition in SHARE

2.4.1 Analytical approach

Research Question 3 aims at understanding non-mortality-related attrition and asked what variables/characteristics played the most important role in the evolution of Analysis Sample B (which excludes reported deaths before a given wave) over waves across all countries. To answer this question, we calculated subgroup retention rates and estimated logistic regression models across all countries.

We defined several attrition scenarios for Research Question 3:

- Scenario 1 (W2): attrition in Wave 2
- Scenario 2 (W3|W2): attrition in Wave 3, conditional upon participation in Wave 2
- Scenario 4 (W6): attrition in Wave 6
- Scenario 5 (W6|W3): attrition in Wave 6, conditional upon participation in Wave 4

These scenarios will inform researchers about the changes in the composition of the initially recruited first-wave SHARE sample in later waves. For further exploration, we also defined and analyzed a number of other scenarios (see Appendix, table A2.1).

We compared subgroup retention rates for the defined scenarios with the first-wave subgroup proportions, excluding reported deaths before the given wave (dynamic Analysis Sample B). Only deviations of one percentage point or more are reported in the corresponding figures (see Appendix, figs. A2.1– A2.3), and only deviations of two percentage points or more are discussed in what follows.

In addition to the univariate subgroup retention rates, we explored non-mortality-related attrition within a multivariate framework because multivariate analyses allow several respondent and household characteristics to be taken into account at once. We estimated logit equations to examine which selected key variables have played the most important role in the evolution of the panel for the various selected scenarios. In contrast to the subgroup retention rates, the coding of y_i was reversed intentionally for the multivariate logits. It allows for an interpretation of the results related to attrition rather than participation. Thus, the attrition propensity ρ_i for a panel respondent i is defined as follows:

$$\rho_i(X) = P(y_i = 1 \mid X = x_i). \tag{3}$$

For a respondent $i = 1, \dots, N$, y_i refers to the binary nonresponse outcome, which equals 1 if panel respondent *i* dropped out and 0 otherwise. The outcome y_i can be different for each of the six waves; x_i is a vector of the 26 selected SHARE key variables for panel respondent *i* (see table 2.2).

As standard coefficients in logistic models indicate only the effect direction and provide no information about effect size, we estimated average marginal effects (AME) to evaluate the logistic regression coefficients more appropriately. AMEs represent the average change in probability when the variable predictor increases by one unit (Mood 2010). Moreover, by examining the *z*-scores of the logistic regression models we could quantify the impact of the individual and household characteristics on non-mortality-related attrition (Analysis Sample B). This examination deepened the understanding which variables actually led to a decline of the R-indicators in Section 2.3.2.

2.4.2 Results

To answer Research Question 3 as to what variables/characteristics played the most important role in the changes in the composition of the initially recruited first-wave sample (Analysis Sample B, which excludes reported deaths before a given wave) over waves, we calculated subgroup retention rates on participation and ran logistic regression models on attrition for the selected scenarios.

Wave 2. In the subgroup retention rates in Wave 2, where respondents who were reported to have died before Wave 2 were excluded, we observed a deviation of two or more percentage points from the initially recruited Analysis Sample B in Wave 1 only for social activity (see Appendix, fig. A2.1). The share of respondents who were socially active in Wave 1 increased by 2.4 percentage points in Wave 2, whereas the share of those who were not socially active increased by the same amount of percentage points.

The multivariate analyses of the Analysis Sample B in Wave 2, that excludes reported deaths before Wave 2, in table 2.3 showed that, after controlling for other respondent and household characteristics, the association of being socially active with not participating in the second wave was statistically significant (p<.001; z-score = −5.11). The probability of attrition in Wave 2 decreased by four percentage points if respondents were socially active in Wave 1. However, the significant association of social activity with not participating in Wave 2 was not the strongest association observed. Rather, the strongest association of attrition in Wave 2 was observed with residing in a rural village (p<.001; z-score = −9.12). The probability of dropping out in Wave 2 was seven percentage points lower for respondents residing in rural villages than for those living in cities or large towns.

Table 2.3. Estimated average marginal effects (AME) from logistic regressions of attrition by individual and household characteristics

	W2	W3|W2	W6	W6|W3
Gender: male (ref.)				
- female	−.01	−.01	−.04**	−.01
	(−1.15)	(−.46)	(−2.96)	(−.99)
Age: 50–59 years (ref.)				
- 60–69 years	−.02*	−.04***	−.03*	−.00
	(−2.23)	(−3.48)	(−2.38)	(−.38)
- 70–79 years	.00	−.03*	−.01	.01
	(.06)	(−2.11)	(−.72)	(.58)
- 80+ years	−.00	.01	.11***	.10***
	(−.16)	(.81)	(4.78)	(3.42)
Born in survey country: no (ref.)				
- yes	−.05***	−.05***	−.09***	−.07***
	(−3.86)	(−3.38)	(−6.27)	(−4.05)
Education level: low (ref.)				
- medium	−.00	−.02	−.01	−.01
	(−.51)	(−1.81)	(−1.00)	(−.46)
- high	−.04***	−.05***	−.07***	−.05***
	(−4.14)	(−4.98)	(−6.43)	(−4.20)
HH size: 1–person (ref.)				
- 2–person HH	.00	.03	.02	.03
	(.13)	(1.47)	(1.26)	(1.21)
- 3+ person HH	.01	.02	.02	.04
	(.64)	(.92)	(.76)	(1.37)
Partner in HH: no (ref.)				
- yes	.01	.01	.02	−.01
	(.73)	(.48)	(.89)	(−.35)
Area of residence: city/large town (ref.)				
- small town	−.03***	−.03**	−.05***	−.05***
	(−4.26)	(−3.08)	(−5.56)	(−4.44)
- rural village	−.07***	−.03**	−.07***	−.03*
	(−9.12)	(−3.00)	(−6.90)	(−2.54)
Residential proximity of child(ren): no children (ref.)				
- child in HH	−.08***	−.03	−.07***	−.04*
	(−5.61)	(−1.69)	(−4.23)	(−2.33)
- child ≤ 1 km away	−.06***	−.01	−.08***	−.04*
	(−4.39)	(−.72)	(−5.04)	(−2.03)
- child > 1 km away	−.04***	−.01	−.05***	−.02
	(−3.95)	(−1.10)	(−3.86)	(−1.17)
Social activities: no activities (ref.)				
- at least one activity	−.04***	−.03***	−.06***	−.04***
	(−5.11)	(−3.92)	(−7.06)	(−4.12)
Received help from others: no (ref.)				
- yes	.02*	−.02*	−.02	-.02
	(2.08)	(−2.06)	(−1.86)	(-1.66)
Gave help to others: no (ref.)				
- yes	−.02*	−.01	−.01	-.00
	(−2.44)	(−1.86)	(−.74)	(-.06)
Health status: good/better (ref.)				
- poor or fair	.01	.01	.02	.02
	(1.62)	(1.08)	(1.84)	(1.90)
Chronic diseases: none (ref.)				
- 1+ chronic diseases	−.02*	−.02*	−.01	−.02*
	(−2.25)	(−2.22)	(−1.32)	(−2.25)

Table 2.3. (cont.)

	W2	W3|W2	W6	W6|W3
Depression (Euro-D): insufficient symptoms (ref.)				
- 4+ symptoms	−.02**	−.00	−.02*	−.01
	(−2.96)	(−.39)	(−2.33)	(−.57)
Maximum grip strength: item nonresponse (ref.)				
- 1st quartile (very weak)	−.08***	−.01	−.04	−.02
	(−4.84)	(−.71)	(−2.65)	(−.68)
- 2nd quartile	−.07***	−.01	−.06**	.01
	(−4.60)	(−.29)	(−1.85)	(.50)
- 3rd quartile	−.08***	−.00	−.07**	−.01
	(−4.73)	(−.05)	(−3.20)	(−.33)
- 4th quartile (very strong)	−.08***	.00	−.04	−.00
	(−4.52)	(.10)	(−3.08)	(−.18)
Memory recall ability:				
- less than half of the words (ref.)				
- more than half of the words	−.03***	−.01	–.02**	−.02
	(−4.27)	(−1.66)	(−2.62)	(−1.84)
Hospital overnight stay in last 12 months: no (ref.)				
- yes	−.00	−.00	−.00	−.01
	(−.39)	(−.01)	(−.40)	(−.80)
Currently smoking: no (ref.)				
- yes	.03***	.01	.04***	.01
	(3.95)	(1.48)	(4.43)	(.77)
Currently drinking: never (ref.)				
- less than once a week	−.03**	−.02	−.05***	−.03*
	(−2.70)	(−1.65)	(−3.81)	(−2.41)
- 1–6 times a week	−.04***	−.02	−.05***	−.05***
	(−4.07)	(−1.65)	(−4.68)	(−3.64)
- almost every day	−.04***	−.01	−.04***	−.02
	(−4.33)	(−.89)	(−3.77)	(−1.62)
IADL: no IADL limitations (ref.)				
- 1+ IADL limitations	.03**	.03*	.04***	.04*
	(3.03)	(2.16)	(3.41)	(2.33)
Employment status: retired (ref.)				
- working	.01	−.00	−.01	-.01
	(.97)	(−.03)	(−.95)	(-.64)
- not working and other	−.00	−.00	.00	.01
	(−.34)	(−.47)	(.15)	(.39)
Making ends meet: difficulties (ref.)				
- no difficulties	−.01	−.00	.04***	−.02
	(−1.65)	(−.13)	(4.43)	(−1.67)
Total household income: item nonresponse (ref.)				
- 1st quartile	−.07***	−.04**	−.05***	−.04*
	(−5.54)	(−3.08)	(−3.39)	(−2.11)
- 2nd quartile	−.07***	−.05***	−.06***	−.04*
	(−5.93)	(−3.37)	(−4.51)	(−2.49)
- 3rd quartile	−.06***	−.06***	−.06***	−.02
	(−4.92)	(−4.21)	(−4.64)	(−1.51)
- 4th quartile	−.05***	−.04**	−.04**	−.00
	(−3.75)	(−2.83)	(−2.79)	(−.29)

Table 2.3. (cont.)

	W2	W3|W2	W6	W6|W3
Family respondent: no (ref.)				
- yes	−.01	−.00	−.01	.01
	(−.92)	(−.19)	(.62)	(.80)
Financial respondent: no (ref.)				
- yes	.02	−.00	–.01	−.02
	(1.25)	(−.04)	(−.95)	(−1.27)
Household respondent: no (ref.)				
- yes	−.03	−.00	–.01	−.00
	(−1.95)	(−.03)	(−.57)	(−.08)
N	19,761	13,466	16,619	10,412

Note: W2 = attrition in Wave 2. W3|W2 = attrition in Wave 3, conditional upon participation in Wave 2. W6 = attrition in Wave 6. W6|W3 = attrition in Wave 6, conditional upon participation in Wave 3. Z statistics in parentheses. HH=Household. IADL = instrumental activities of daily living; all models additionally include country dummies. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Other strong associations with attrition in Wave 2 were found for respondents who had participated in the grip strength test and who had reported their total household income in Wave 1, regardless of the value in measure (p<0.001; z-scores between −3.75 and −5.93). They were less likely to drop out in Wave 2 than respondents who had not provided these measures. The decrease in probability to drop out ranged from five to eight percentage points (table 2.3).

The multivariate analyses additionally showed that other numerous individual and household characteristics of Analysis Sample B in Wave 2 were significantly associated with attrition in the second wave (see table 2.3). The probability to drop out increased significantly with: having received help from others, smoking, and having at least reported one limitation in IADL in the first wave. In addition to these positive significant associations with attrition in the second wave, we observed several negative significant associations with attrition in the second wave. Respondents who were between 60 and 69 years old in Wave 1 were less likely to drop out in Wave 2 than respondents who were between 50 and 59 years old in Wave 1. A respondent born in the survey country was less likely to attrite in Wave 2 than a respondent born outside the survey country. Highly educated respondents were less likely to attrite than low educated respondents, and respondents who resided in a small town in Wave 1 had a lower probability to drop out than respondents residing in a city or large town in Wave 1. Having children, among all groups of residential proximity of the child in Wave 1, decreased the probability to drop out in Wave 2 compared to having no children. Moreover, the probability to drop out in Wave 2 decreased significantly at the 5 percent level with: giving help to others, having at least reported to have one chronic disease, having reported at least four depression symptoms, having a larger memory recall ability, and drinking, regardless of the frequency of alcohol consumption in the first wave.

Wave 3. The subgroup retention rates of Analysis Sample B in Wave 3, conditional upon participation in Wave 2, showed no deviations larger than two percentage points from the initially recruited respondents in the first wave when we excluded respondents that that were reported to have died before the third wave. Only one deviation larger than one percentage point was observed from respondents who resided in the city or large town. Their share was 1.1 percentage points lower compared to their share in Wave 1 (result not shown).

Multivariate analyses showed that strong predictors of attrition in Wave 3, conditional upon participation in Wave 2, were: high educational level (p<.001; z-score = −4.98) compared to a low educational level, social activity in Wave 1 (p<.001; z-score = −3.92), age between 60 and 69 years in Wave 1 (p<.001; z-score = −3.48) compared to age between 50 and 59 years in Wave 1, birth in survey country level (p<.001; z-score = −3.38), and reporting the total household income among all income groups (p<.01; z-scores between −2.83 and −4.21) compared to item nonresponse in the total household income in Wave 1 (table 2.3).

Other significant negative associations with attrition were observed for: respondents who were between 70 and 79 years old, resided in a small town or rural village, received help from others, and reported at least one chronic disease in the first wave compared to corresponding reference category. Other positive significant associations with attrition in the third wave were observed with having reported at least one IADL limitation in Wave 1 (see table 2.3).

Some significant effects of individual and household charactersitics on attrition we found in the model for the second wave, that excluded reported deaths before Wave 2, could not be found in the conditional model for the third wave, where we excluded reported deaths before Wave 3 (see table 2.3).

Wave 6. The proportion of respondents who were born in the survey country, and of respondents who self-assessed their health in Wave 1 as good or better, and of respondents who were socially active in Wave 1 was between 2.40 and 3.64 percentage points larger for the panel members who participated in Wave 6 compared to the respective Wave 1 proportions. Moreover, the proportion of respondents who had a medium educational level was 3.06 percentage points smaller compared to the respective Wave 1 proportion (see Appendix, fig. A2.2). In the conditional Wave 6 scenario (attrition in Wave 6, conditional upon participation in Wave 3) no larger deviation than two percentage points were observed (see Appendix, fig. A2.3).

Examining multivariate attrition in Wave 6, we observed for the unconditional scenario that many individual and household characteristics significantly predicted the drop out in the sixth wave (see table 2.3). Strong positive associations with attrition were found for respondents who smoked ($p<.001$; z-score = 4.43), made ends meet with no difficulties ($p<.001$; z-score = 4.43), and reported at least one IADL limitation in the first wave ($p<.001$; z-score = 3.41) compared to respondents who did not smoke, made ends meet with difficulties, and reported no IADL limitation in the first wave. The probability to drop out increased by four percentage points for each of these characteristics (smoking, making ends meet, and having at least one IADL limitation). Strong negative associations with attrition were found for respondents who were socially active (p<.001; z-score = −7.06), had a high educational level (p<.001; z-score = −6.43) compared to low educational level, were born in survey country (p<.001; z-score = −6.27), resided in a rural village (p<.001; z-score = −6.90) or small town (p<.001; z-score = −5.56) compared to city or large town. The decrease in probability to drop out for these groups ranged between four and ten percentage points. For further negative and positive associations in Wave 6 (with a lower significance level than 99.9 percent or with a smaller absolute value in z-score than 5) please see table 2.3.

For attrition in Wave 6, conditional upon participation in Wave 3, we observed at the significance level of 99.9 percent, that highly educated and socially active respondents in Wave 1, and who were born in the survey country were less likely to drop out in Wave 6 than low-educated and socially inactive respondents and those, who were born outside the survey country (table 2.3). Furthermore, residing in a small town and drinking between one and six drinks peer week, compared to residing in a city or large town and not drinking in Wave 1 decreased the probability of dropping out by five percentage points for the respective characteristics (table 2.3). For further negative and positive associations (with a lower significance level than 99.9%) please see table 2.3.

Comparing the conditional Wave 6 attrition model with the unconditional Wave 6 attrition model, we noted that far fewer individual and household characteristics were significantly associated with attrition in the conditional model. However, the age group 80+ in Wave 1, who were aged 92+ years in Wave 6, had a relatively large positive impact in both Wave 6 attrition models. The probability to drop out increased by 11 percentage points in the unconditional model and by 10 percentage points in the conditional model for those old respondents (table 2.3).

2.5 Summary and Discussion

This study examined the evolution of the initially recruited SHARE first-wave sample. With its specific target population, SHARE has a relatively large proportion of respondents who are at a high risk of attrition because of death. As we assumed that people who die are a selective group of the population and of the panel sample, we investigated the evolution of the SHARE panel with two defined samples. We used Analysis Sample A to study total attrition (non-mortality-related and mortality-related attrition), and Analysis Sample B to study exclusively non-mortality-related attrition. We applied different methods to answer our research questions.

We answered Research Question 1 “How has the initially recruited SHARE first-wave sample (A and B) has evolved over waves” by calculating retention rates and estimating R-indicators. We detected declining retention rates with a major loss of respondents in the second wave. This finding is in line with previous literature (Lepkowski and Couper 2002; Schoeni et al. 2013; Lugtig 2014). Moreover, the retention rates observed in our study are about the same as those for second-wave response in other studies of older populations (Banks, Alastair, and Smith 2011). In addition, we observed that the values of the R-indicators of the initially recruited SHARE sample (Analysis Sample A and B) dropped in the second wave but remained stable afterwards. Thus, we could show that, despite declining retention rates, the composition of the first-wave sample changed but was maintained over waves with respect to many individual or household characteristics after the second wave. Furthermore, the results showed, when we excluded respondents that had been reported as dead before a given wave (Analysis Sample B, Wave 1− Wave 6), that, the observed changes in the sample composition over time were mainly due to deaths (with the exception of Wave 2).

As SHARE collects data in various countries, it has to deal with country-specific differences, although it is harmonized ex ante. Therefore, we further investigated the evolution of the SHARE panel by Research Question 2 "Has the evolution of the initially recruited first-wave sample (A and B) over waves varied across countries?". We observed that the changes in the composition of the initially recruited sample over time differed across countries, although the differences were small. All countries followed the same trend, with a stable R-indicator value after the second wave (Analysis Samples A and B). However, comparing R-indicator values for Analysis Sample B (excluding deaths before a given wave) revealed larger differences across countries. These differences may be due to the quality of the respective death reports.

To answer Research Question 3 as to what characteristics and variables played the most important role in the changes in the composition of the initially recruited first-wave sample (dynamic Analysis Sample B) over waves, we examined various attrition scenarios by calculating subgroup retention rates and estimating multivariate logistic regression models on attrition. The results of the subgroup retention rate analyses were supported by those of the multivariate analyses. In all multivariate models, first-wave respondents who were born in the survey country, were residing in a rural area or small town, had a high level of education, and were socially active were less likely to attrite than first-wave respondents who were not born in the survey country, who were residing in a city or a large town, who had a low level of education, and were socially inactive. We did not observe that health-related variables, such as illness or age, were strong predictors of non-mortality-related attrition. Only very old respondents (aged 80+ in the first wave) had a high risk of attrition in later waves. Overall, birth in survey country, area of residence, education, and social activities played an important role in the non-mortality related attrition and their impact led to a decline of the R-indicators.

Comparing logit models from early waves with those from later waves, we noted that some significant associations declined to statistical insignificance in the multivariate models, especially in the models for attrition conditional upon participation in a specified previous wave. This change in significance is reflected in the stabilizing R-indicator after the second wave.

The present study has a number of limitations. To draw conclusions from panel data about the general population aged 50 years or older, researchers need to consider and investigate initial nonresponse – that is, nonresponse that occurs in the recruitment stage of the panel. As the focus of the present study was on the evolution of the initially recruited first-wave sample over waves, we did not consider initial nonresponse. However, as initial nonresponse is an important factor for understanding the overall nonresponse process in SHARE and might have an impact on the data researchers use for analyzing dynamic processes in the European societies, future research should take it into account.

Another limitation of this study relates to the reporting of deaths. The SHARE countries included in the study differed in the share of reported deaths in the initially recruited sample over the course of the panel. Unlike the U.S. Health and Retirement Survey (HRS) or the English Longitudinal Survey of Ageing (ELSA) in England, SHARE cannot be linked to a mortality register because national mortality registers are lacking in most European countries (Bergmann et al. 2019). A comparison of the share of reported deaths in the initially recruited first-wave sample in SHARE with the mortality rate among persons aged 50+ years between 2004 and 2015 in Eurostat data (Eurostat 2004–2015) showed that only in a minority of the SHARE countries in our study was the share of respondents who died over course of the panel lower than the estimated share of persons in the corresponding population group who died between 2004 and 2015 (Appendix, table A2.2). Thus, we may have underestimated the number of deaths in

SHARE for a few countries due to a lack of information. However, we expected the share of deaths in Eurostat and SHARE to differ to some extent because SHARE excludes the hospitalized population from the sampling frame.

Notwithstanding these limitations, our study shows that, despite declining retention rates, the composition of an initially recruited panel sample can remain stable over later waves. The representativeness of the first wave sample (fixed Analysis Sample A and dynamic Analysis Sample B) did not decline further after the second wave. Moreover, this study informs researchers who wish to analyze dynamic processes over time about the impact of mortality-related and non-mortality-related attrition on the composition of the initially recruited first-wave SHARE sample over time. To further inform researchers wishing to analyze dynamic processes in SHARE over time, future research should examine the impact of mortality- and non-mortality-related attrition on cross-sectional and longitudinal estimates.

References

American Association for Public Opinion Research (AAPOR). 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.). AAPOR. Available at: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf (accessed August 2018).

Banks, James, Muriel Alastair, and James P. Smith. 2011. "Attrition and Health in Ageing Studies: Evidence from ELSA and HRS." *Longitudinal and Life Course Studies* 2 (2): 1–29. https://doi.org/10.14301/llcs.v2i2.115.

Bergmann, Michael, Thorsten Kneip, Giuseppe De Luca, and Annette Scherpenzeel. 2019. *Survey Participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1–7*. SHARE Working Paper Series 41-2019. Munich, Germany: Munich Center for the Economics of Aging (MEA), Max Planck Institute for Social Law and Social Policy. Available at: http://www.share-project.org/uploads/tx_sharepublications/WP_Series_41_2019_Bergmann_et_al.pdf (accessed February 2019).

Bianchi, Annamaria, and Silvia Biffignandi. 2017. "Representativeness in Panel Surveys." *Mathematical Population Studies* 24 (2): 126–143. https://doi.org/10.1080/08898480.2016.1271650.

Bianchi, Annamaria, and Silvia Biffignandi. 2019. "Social Indicators to Explain Response in Longitudinal Studies." *Social Indicators Research* 141 (3): 931–957. https://doi.org/10.1007/s11205-018-1874-7.

Binder, David A. 1998. "Longitudinal Surveys: Why Are These Surveys Different from All Other Surveys?" *Survey Methodology* 24 (2): 101–108.

Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie M. Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology* 42 (4): 992–1001. https://doi.org/10.1093/ije/dyt088.

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

Bristle, Johanna, Martina Celidoni, Chiara Dal Bianco, and Gugliemo Weber. 2019. "The Contributions of Paradata and Features of Respondents, Interviewers and Survey Agencies to Panel Co-Operation in the Survey of Health, Ageing and Retirement in Europe." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182 (1): 3–35. https://doi.org/10.1111/rssa.12391.

Buck, Nicholas, Jonathan Burton, Heather Laurie, Peter Lynn, and S.C. Noah Uhrig. 2006. *Quality Profile: British Household Panel. Survey Version 2.0: Waves 1 to 13: 1991–2003.* Colchester, UK: University of Essex, Institute for Social and Economic Research.

De Luca, Giuseppe, Claudio Rossetti, and Frederic Malter. 2013. "Sample Design and Weighting Strategies in SHARE Wave 5." In *SHARE Wave 5: Innovations & Methodology*, edited by Frederic Malter and Axel Börsch-Supan, 75–84. Munich, Germany: Munich Center for the Economics of Aging, Max Planck Institute for Social Law and Social Policy.

Eurostat. 2004-2015. Available at: https://ec.europa.eu/eurostat/web/population-demography-migration-projections/population-projections-/database (accessed June 2019).

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *The Journal of Human Resources* 33 (2): 251–299. https://doi.org/10.2307/146433.

Leon, David A. 2011. "Trends in European Life Expectancy: A Salutary View." *International Journal of Epidemiology* 40 (2): 271–277. https://doi.org/10.1093/ije/dyr061.

Lepkowski, James M. and Mick P. Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little, 259–272. New York: John Wiley.

Lugtig, Peter. 2014. "Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers." *Sociological Methods & Research* 43 (4): 699–723. https://doi.org/10.1177/0049124113520305.

Luiten, Annemieke, and Barry Schouten. 2013. "Tailored Fieldwork Design to Increase Representative Household Survey Response: An Experiment in the Survey of Consumer Satisfaction." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (1): 169-189 https://doi.org/10.1111/j.1467-985X.2012.01080.x.

Lynn, Peter. 2009. "Methods for Longitudinal Surveys." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 1–19. Chichester, UK: John Wiley.

Lynn, Peter, and Peter Lugtig. 2017. "Total Survey Error for Longitudinal Surveys." In *Total Survey Error in Practice*, edited by Paul .P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West, 279–298. Chichester, UK: John Wiley.

Mood, Carina. 2010. "Logistic Regression: Why We Cannot Do what We Think We Can Do, and What We Can Do about It." *European Sociological Review* 26 (1): 67–82. https://doi.org/10.1093/esr/jcp006.

Moore, Jamie C., Gabriele B. Durrant, and Peter W. F. Smith. 2016. "Data Set Representativeness during Data Collection in Three UK Social Surveys: Generalizability and the Effects of Auxiliary Covariate Choice." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181 (1): 229–248. https://doi.org/10.1111/rssa.12256.

Myrskylä, Mikko, Joshua R . Goldstein, and Yen-hsin A. Cheng. 2013. "New Cohort Fertility Forecasts for the Developed World: Rises, Falls, and Reversals." *Population and Development Review* 39 (1): 31–56. https://doi.org/10.1111/j.1728-4457.2013.00572.x.

Olsen, Randall J. 2018. "Respondent Attrition Versus Data Attrition and Their Reduction." In *The Palgrave Handbook of Survey Research*, edited by David L. Vannette and Jon A. Krosnick, 155–158. Cham, Switzerland: Springer.

RISQ. 2015. *Representative Indicators for Survey Quality-Tools*. Manchester, UK: University of Manchester. Available at: https://www.cmist.manchester.ac.uk/research/projects/representative-indicators-for-survey-quality/tools/ (accessed August 2018).

Roberts, Caroline, Caroline Vandenplas, and Michèle E. Stähli. 2014. "Evaluating the Impact of Response Enhancement Methods on the Risk of Nonresponse Bias and Survey Costs." *Survey Research Methods* 8 (2): 67–80. https://doi.org/10.18148/srm/2014.v8i2.5459.

Schoeni, Robert F., Frank Stafford, Katherine A. McGonagle, and Patricia Andreski. 2013. "Response Rates in National Panel Surveys." *Annals of the American Academy of Political and Social Science* 645 (1): 60–87. https://doi.org/10.1177/0002716212456363.

Schouten, Barry, Jelke Bethlehem, Koen Beullens, Øyvin Kleven, Geert Loosveldt, Annemieke Luiten, Katja Rutar, Natalie Shlomo, and Chris Skinner. 2012. "Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response through R-indicators and Partial R-indicators." *International Statistical Review* 80 (3): 382–399. https://doi.org/10.1111/j.1751-5823.2012.00189.x.

Schouten, Barry, Fannie Cobben, and Jelke Bethlehem. 2009. "Indicators for the Representativeness of Survey Response." *Survey Methodology* 35 (1): 101–113.

Schouten, Barry, Nathalie Shlomo, and Chris Skinner. 2011. "Indicators for Monitoring and Improving Representativeness of Response." *Journal of Official Statistics* 27 (2): 231–253.

SHARE-ERIC. 2018. SHARE Publications: Journal Articles. Available at: http://www.share-project.org/share-publications/journalarticles00.html (accessed July 2020).

Smith, Paul, Peter Lynn, and Dave Elliot. 2009. "Sample Design for Longitudinal Surveys." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 21–33. Chichester, UK: John Wiley.

Watson, Dorothy. 2003. "Sample Attrition between Waves 1 and 5 in the European Community Household Panel." *European Sociological Review* 19 (4): 361–378. https://doi.org/10.1093/esr/19.4.361.

Watson, Nicole, and Mark Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 157–181. Chichester, UK: John Wiley.

Watson, Nicole, and Mark Wooden. 2019. "Chasing Hard-to-Get Cases in Panel Surveys: Is it Worth it?" *methods, data, analyses* 13 (2): 199–222. https://doi.org/10.12758/mda.2018.03.

Appendix

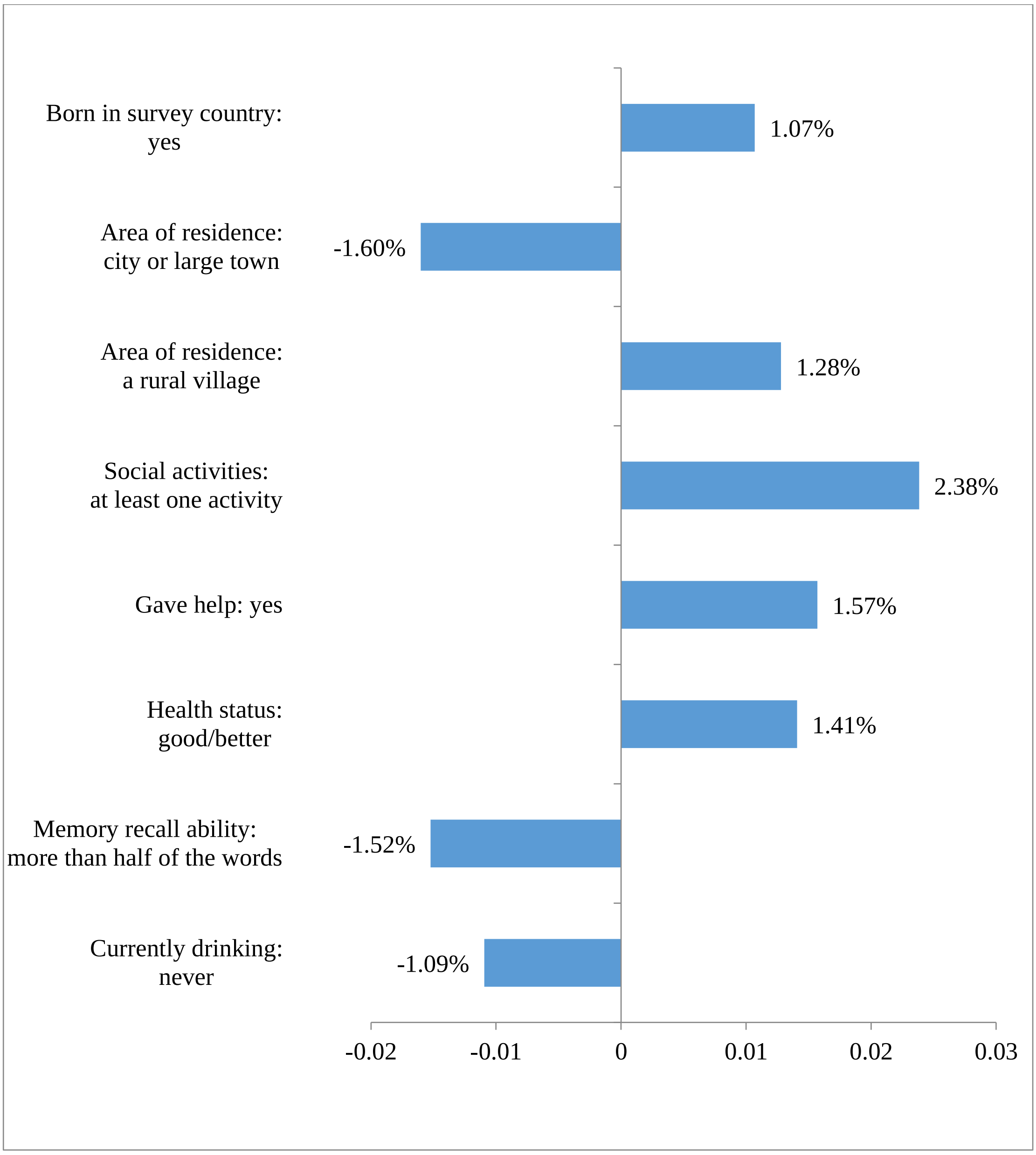

Figure A2.1. Deviation of Wave 2 proportions from Wave 1 proportions, excluding reported deaths before Wave 2
Note: HH = household; Resid. = residential. IADL = instrumental activities of daily living. *n*=13,959.

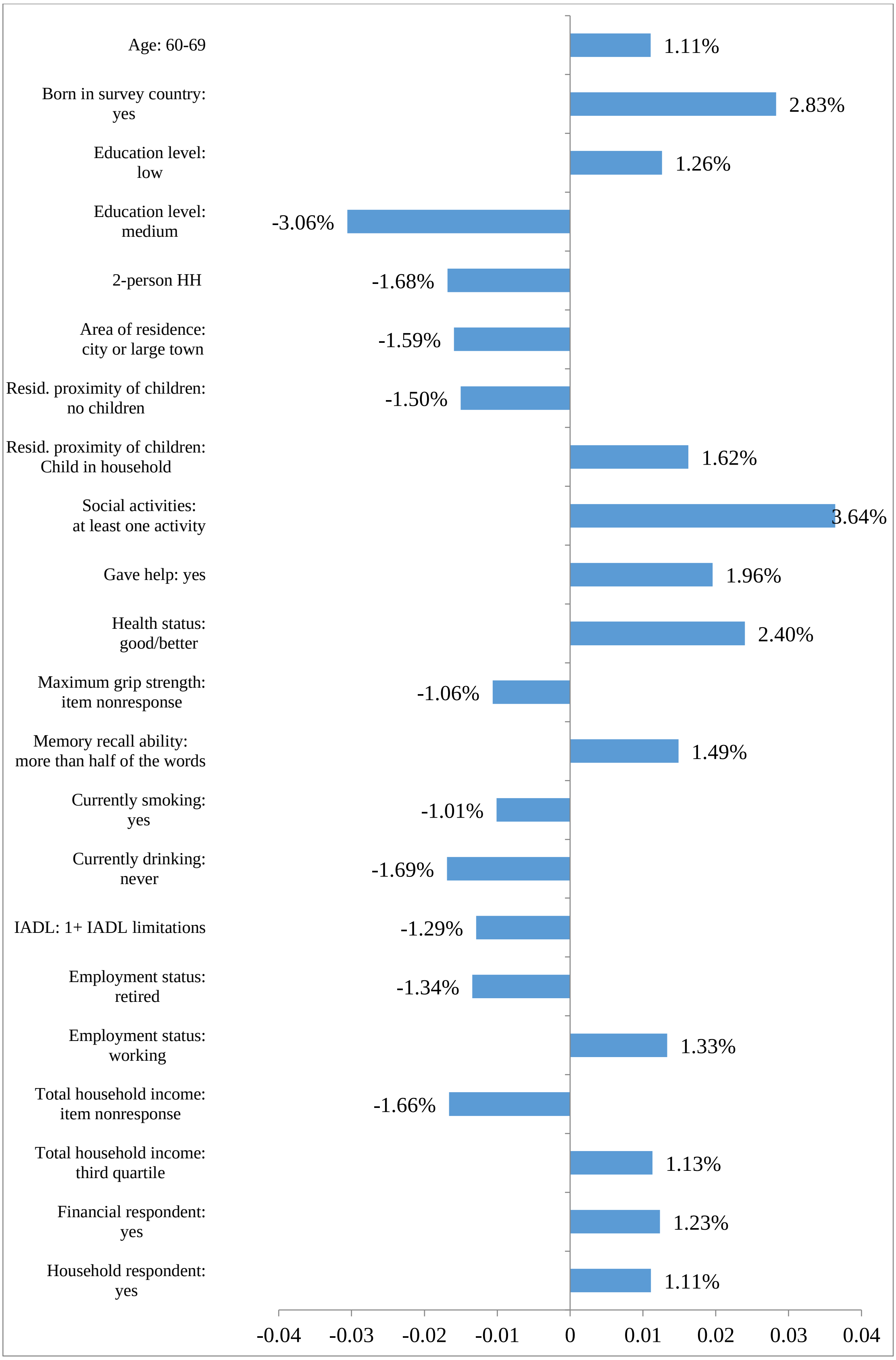

Figure A2.2. Deviation of Wave 6 proportions from Wave 1 proportions, excluding reported deaths before Wave 6
Note: HH = household; resid. = residential. IADL = instrumental activities of daily living; *n*=8,545.

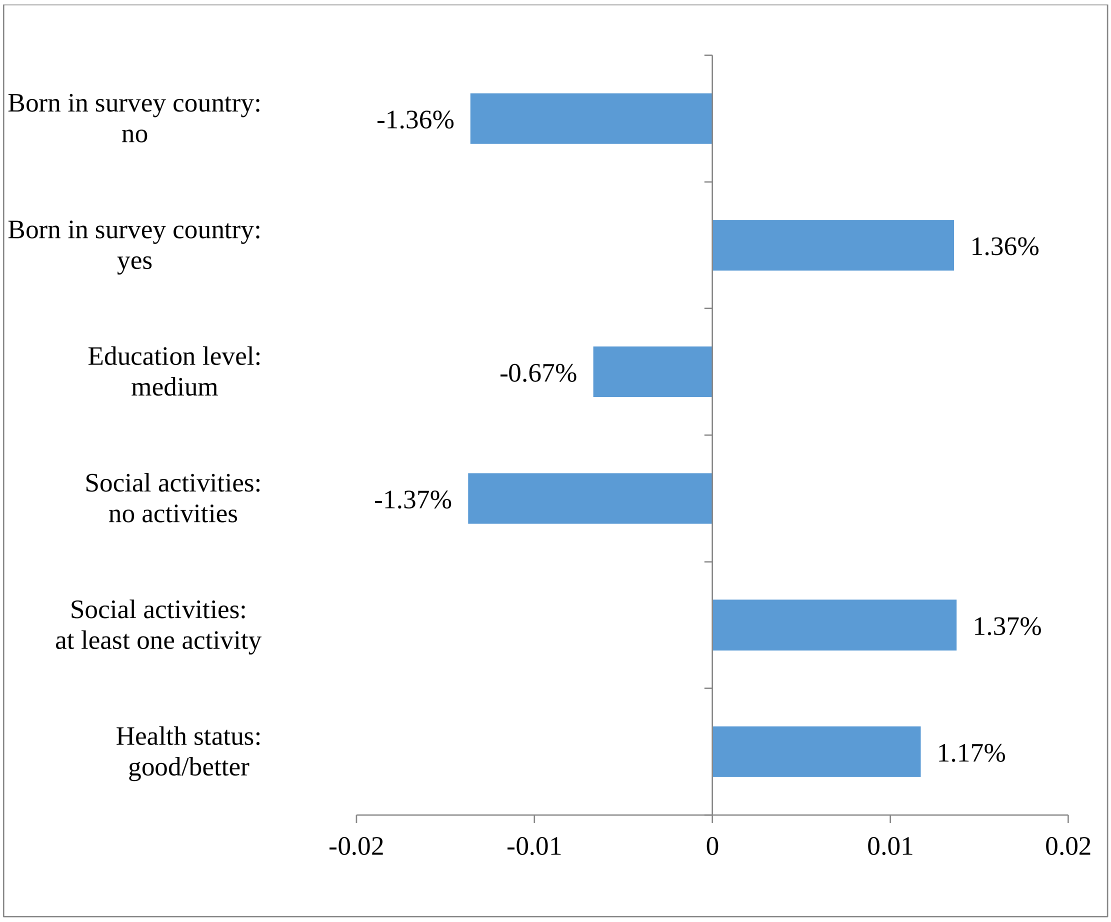

Figure A2.3. Deviation of Wave 6 from Wave 1 proportions, excluding reported deaths before Wave 6 and conditional upon participation in Wave 3
Note: HH = household. resid. = residential. IADL = instrumental activities of daily living; *n*=7,513.

Table A2.1. Estimate average marginal effects (AME) of additional attrition models

	W3	W4	W4|W2	W5	W5|W4
Gender: male (ref.)					
- female	−.01	−.02	−.01	−.02	−.01
	(−1.05)	(−1.57)	(−.56)	(−1.39)	(−.64)
Age: 50–59 years (ref.)					
- 60–69 years	−.04***	−.04***	−.04***	−.03*	−.01
	(−4.13)	(−3.95)	(−3.75)	(−2.44)	(−1.27)
- 70–79 years	−.03**	−.04**	−.04**	−.02	.01
	(−2.63)	(−2.78)	(−2.59)	(−1.72)	(.37)
- 80+ years	−.00	.04*	.06*	.06**	.02
	(−.25)	(2.03)	(2.49)	(2.98)	(.96)
Born in survey country: no (ref.)					
- yes	−.07***	−.07***	−.03*	−.08***	−.03*
	(−5.62)	(−5.16)	(−2.24)	(−5.67)	(−2.27)
Education level: low (ref.)					
- medium	−.01	−.01	−.00	−.02*	−.02*
	(−.93)	(−1.05)	(−.38)	(−2.42)	(−2.48)
- high	−.07***	−.07***	−.03**	−.07***	−.04***
	(−7.10)	(−6.19)	(−2.97)	(−6.80)	(−4.27)
HH size: 1-person (ref.)					
- 2-person HH	.02	.03	.03	.05*	.03
	(1.24)	(1.81)	(1.63)	(2.51)	(1.61)
- 3+person HH	.01	.02	.02	.04	.03
	(.44)	(.81)	(.68)	(1.72)	(1.31)
Partner in HH: no (ref.)					
- yes	.03	.01	.01	.01	.01
	(1.50)	(.72)	(.31)	(.55)	(.56)
Area of residence: city/large town (ref.)					
- small town	−.04***	−.05***	−.04***	−.05***	−.03**
	(−4.48)	(−5.48)	(−3.97)	(−5.75)	(−3.02)
- rural village	−.07***	−.08***	−.05***	−.07***	−.02*
	(−8.23)	(−8.82)	(−4.48)	(−7.44)	(−2.21)
Residential proximity of child(ren): no child (ref.)					
- child lives in HH	−.07***	−.07***	−.04*	−.06***	.00
	(−4.38)	(−4.52)	(−2.32)	(−3.56)	(.08)
- child ≤ 1 km away	−.06***	−.08***	−.04*	−.06***	−.00
	(−4.28)	(−5.32)	(−2.57)	(−3.99)	(−.01)
- child > 1 km away	−.05***	−.05***	−.03*	−.03*	.02
	(−3.72)	(−3.96)	(−2.01)	(−2.18)	(1.82)
Social activities: no (ref.)					
- at least one activity	−.05***	−.05***	−.04***	−.06***	−.03**
	(−6.68)	(−6.37)	(−4.30)	(−6.72)	(−3.22)
Received help from others: no (ref.)					
- yes	−.01	−.01	−.02*	−.02	−.00
	(−1.11)	(−1.46)	(−2.23)	(−1.62)	(−.44)
Gave help to others: no (ref.)					
- yes	−.02**	−.02*	−.01	−.02*	−.01
	(−2.58)	(−2.56)	(−1.68)	(−2.08)	(−.98)
Health status: good/better (ref.)					
- poor or fair	.02*	.02	.01	.01	.00
	(1.96)	(1.60)	(.54)	(1.15)	(.30)
Chronic diseases: none (ref.)					
- 1+ chronic diseases	−.02*	−.01	−.01	−.02*	−.02*
	(−2.45)	(−1.51)	(−.94)	(−2.26)	(−1.99)

Table A2.1. (cont.)

	W3	W4	W4|W2	W5	W5|W4
Depression (Euro-D):					
insufficient symptoms (ref.)					
- 4+ symptoms	−.02*	−.02*	−.00	−.02*	.01
	(−2.52)	(−2.06)	(−.35)	(−1.96)	(1.05)
Maximum grip strength:					
item nonresponse (ref.)					
- 1st quartile					
(very weak)	−.07***	−.07***	−.04*	−.07***	−.01
- 2nd quartile	(−3.76)	(−3.80)	(−2.02)	(−3.74)	(−.58)
	−.06***	−.07***	−.04	−.07***	−.01
- 3rd quartile	(−3.64)	(−3.71)	(−1.88)	(−3.41)	(−.44)
	−.07***	−.08***	−.04	−.08***	−.01
- 4th quartile	(−3.86)	(−4.28)	(−1.87)	(−3.93)	(−.62)
(very strong)	−.06**	−.09***	−.05*	−.08***	−.01
	(−3.09)	(−4.17)	(−2.19)	(−3.56)	(−.45)
Memory recall ability:					
recalled less than half of					
the words (ref.)					
- recalled more than half	−.02**	−.02*	−.01	−.03**	−.02
of the words	(−2.87)	(−2.50)	(−1.32)	(−3.07)	(−1.95)
Hospital overnight stay in last					
12 months: no (ref.)					
- yes	−.00	−.01	−.00	.01	.02
	(−.19)	(−.99)	(−.37)	(.99)	(1.31)
Currently smoking: no (ref.)					
- yes	.04***	.05***	.05***	.04***	−.01
	(3.79)	(5.61)	(4.36)	(4.21)	(−.52)
Currently drinking:					
never (ref.)					
- less than once a	−.03**	−.04***	−.03*	−.04***	−.03*
week	(−3.07)	(−3.31)	(−2.05)	(−3.85)	(−2.22)
- 1–6 times a week	−.03**	−.05***	−.05***	−.05***	−.03**
	(−2.89)	(−4.79)	(−4.19)	(−4.32)	(−3.02)
- almost every day	−.03**	−.04***	−.03*	−.02*	−.00
	(−2.92)	(−3.48)	(−2.27)	(−2.10)	(−.27)
IADL:					
no limitations (ref.)					
- 1+ IADL limitations	.03**	.02	.02	.05***	.03*
	(2.60)	(1.92)	(1.57)	(3.73)	(2.10)
Employment status:					
retired (ref.)					
- working	.00	.00	−.00	−.01	−.02
	(.04)	(.38)	(−.06)	(−.61)	(−1.33)
- not working and other	−.01	−.01	−.00	−.00	.00
	(−.68)	(−.53)	(−.09)	(−.33)	(.04)
Make ends meet:					
difficulties (ref.)					
- no difficulties	.00	.00	.00	.00	−.00
	(.04)	(.11)	(.24)	(.53)	(−.53)
Total household income:					
item nonresponse (ref.)					
- 1st quartile	−.06***	−.06***	−.04**	−.05***	−.02
	(−4.82)	(−4.09)	(−2.76)	(−3.36)	(−1.12)
- 2nd quartile	−.07***	−.05***	−.04*	−.07***	−.05**
	(−5.10)	(−4.02)	(−2.39)	(−4.89)	(−3.24)
- 3rd quartile	−.08***	−.05***	−.04*	−.07***	−.04*
	(−6.36)	(−4.16)	(−2.36)	(−5.25)	(−2.53)
- 4th quartile	−.06***	−.04**	−.03	−.05***	−.02
	(−4.74)	(−2.79)	(−1.77)	(−3.67)	(−1.47)

Table A2.1. (cont.)

	W3	W4	W4|W2	W5	W5|W4
Family respondent: no (ref.)					
- yes	−.00	.01	.01	.00	−.00
	(−.04)	(.64)	(.46)	(.16)	(−.15)
Financial respondent: no (ref.)					
- yes	.01	.01	.00	−.00	−.00
	(.73)	(.82)	(.11)	(−.04)	(−.07)
Country: Austria (ref.)					
- Germany	.17***	.13***	−.00	.23***	.12***
	(1.03)	(7.80)	(−.03)	(13.65)	(5.14)
- Sweden	.04*	−.01	−.02	−.05*	−.01
	(2.20)	(−.65)	(−1.12)	(−2.56)	(−.30)
- Spain	−.09***	−.14***	−.19***	−.20***	−.14***
	(−5.19)	(−7.65)	(−9.30)	(−1.34)	(−7.41)
- Italy	−.05**	−.07***	−.17***	−.11***	−.12***
	(−3.11)	(−3.96)	(−8.76)	(−5.70)	(−6.18)
- France	.01	−.06**	−.11***	−.01	.02
	(.33)	(−3.25)	(−5.51)	(−.73)	(1.02)
- Denmark	−.04*	−.10***	−.12***	−.11***	−.11***
	(−2.29)	(−5.01)	(−5.42)	(−5.59)	(−5.07)
- Switzerland	−.03	−.04*	−.13***	.02	−.08***
	(−1.39)	(−1.96)	(−5.25)	(.94)	(−3.40)
- Belgium	−.08***	−.07***	−.12***	−.05**	−.08***
	(−4.94)	(−4.09)	(−6.32)	(−2.62)	(−4.08)
N	18,926	18,090	12,869	17,260	9,861

Note: W3 = attrition in Wave 3. W4= attrition in Wave 4. W4|W2= attrition in Wave 4, conditional upon participation in Wave 2. W5= attrition in Wave 5. W5|W4= attrition in Wave 5, conditional upon participation in Wave 4. Z statistics in parentheses. HH = Household. IADL = instrumental activities of daily living.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A2.2. Comparison of cumulative Eurostat death statistics (persons aged 50+) with cumulative SHARE death statistics (respondents aged 50+) between 2004 and 2015 (in percent)

	Austria	Belgium	Denmark	France	Germany	Italy	Spain	Sweden	Switzerland
SHARE	.21	.14	.26	.18	.10	.18	.25	.20	.12
Eurostat	.24	.25	.26	.22	.25	.24	.23	.26	.21
Difference	−.03	−.11	−.01	−.04	−.15	−.07	.03	−.06	−.09

Note: Own calculations based on SHARE W1-W6 and Eurostat 2004-2015.

3. What They Expect is What You Get: The Role of Interviewer Expectations in Nonresponse to Income and Asset Questions[3]

[3] This chapter has been published as Friedel, Sabine (2019). "What They Expect Is What You Get: The Role of Interviewer Expectations in Nonresponse to Income and Asset Questions." *Journal of Survey Statistics and Methodology* (online first). https://doi.org/10.1093/jssam/smz022.

Abstract

Personal income and assets are sensitive topics to discuss. This phenomenon is reflected in high rates of nonresponse to financial questions in surveys. In face-to-face surveys, item nonresponse is influenced by interviewers. Although interviewers are trained to conduct standardized interviews, some obtain a higher number of item nonresponses than others. This study examines interviewer effects on nonresponse to questions about household income, bank balances, and interest and dividend income in the Survey of Health, Ageing and Retirement in Europe (SHARE). It investigates, first, the extent to which interviewers affect nonresponse to income and asset questions, and second whether interviewers' prior expectations regarding respondents' likelihood to provide information about their income predict actual nonresponse rates. Results of multilevel modeling show that interviewer influence on nonresponse to the income and asset questions was significant at the five percent level. In addition, interviewer expectations were significantly correlated with "don't know" responses and "refusals." These results indicate that interviewer expectations matter in the context of income and asset questions and that survey practitioners should take this into account when designing interviewer training.

Key words: interviewer effects; item nonresponse; interviewer survey; multilevel regression; theory of self-fulfilling prophecy; missing data

3.1 Introduction

Item nonresponse refers to "the failure to obtain information for one question within an interview" (Groves 1989, 135). This phenomenon often occurs when respondents are asked about their income and assets (Meyer, Mok, and Sullivan 2015). For example, examining the data collected in the US Survey of Income and Program Participation (SIPP), Meyer, Mok, and Sullivan (2015) found nonresponse rates of between 25% and 45% to questions relating to receipt of transfer payments such as assistance for needy families, disability insurance benefits, and unemployment insurance benefits. Schräpler (2006) reported item nonresponse rates of approximately 10% to questions on gross income in the British Household Panel Survey (BHPS). Item nonresponse in face-to-face surveys may be due to the respondent, the interviewer, or the interaction between the interviewer and the respondent (de Leeuw, Hox, and Huisman 2003).

From the respondent's perspective, income and asset questions may be difficult to answer. There are two main reasons for this difficulty. First, the question content is personal and intimate. Respondents may consider such questions to be an invasion of privacy, have concerns regarding data confidentiality, or both (Tourangeau, Rips, and Rasinski 2000). Second, answering such questions is cognitively demanding. Respondents may not answer because they have cognitive limitations, such as memory problems caused by aging or sickness (Colsher and Wallace 1989; Knäuper et al. 1997). As aging often coincides with physical and cognitive decline (Young 1997; Hayden et al. 2011; Milanović et al. 2013), surveys that target the elderly population could be affected more by respondents' reduced capabilities than surveys that target the general population. For instance, in a survey of elderly people (90.6% of respondents were over 70 years of age), Knäuper et al. (1997) found that respondents higher in cognitive ability were less likely to answer "don't know" to difficult questions than were respondents lower in cognitive ability. Therefore, surveys of the elderly may be at a higher risk of item nonresponse than surveys of younger age groups or of the general population.

In addition to the respondents, interviewers can affect item nonresponse (West and Blom 2017). For example, interviewers may skip a question or code the response inadequately (de Leeuw, Hox, and Huisman 2003), thereby producing item nonresponse (van der Zouwen, Dijkstra, and Smit 1991; Dykema, Lepkowski, and Blixt 1997; Josten and Trappmann 2016). Interviewers can also affect item nonresponse positively. For instance, after a "don't know" answer (Schaeffer 1991), they can probe in order to encourage the respondent to give a substantive answer on second thoughts. Survey practitioners and researchers train their interviewers to avoid item nonresponse (Billiet and Loosveldt 1988; Groves 1989; Fowler and Mangione 1990; Dahlhamer et al. 2010). Furthermore, interviewer training aims to reduce variability between interviewers (Fowler and Mangione 1990). However, despite proper training, some interviewers obtain higher item nonresponse rates than others (Bailar, Bailey and Stevens 1977). Previous studies (e.g., Pickery and Loosveldt 2004) have shown that interviewers in face-to-face surveys affect nonresponse to income items. These interviewer effects may arise from interviewer characteristics, experiences, attitudes, expectations, and/or behavior (Blom and Korbmacher 2013; West and Blom 2017).

Researchers have found significant effects of interviewers' sociodemographic characteristics on item nonresponse (e.g., Schräpler 2004; Riphahn and Serfling 2005; Essig and Winter 2009). For example, Schräpler (2004) found that female interviewers obtained notably more "don't know" answers and refusals than their male counterparts. However, it is difficult to apply these findings in survey practice. One reason for this unfeasibility is that sociodemographic characteristics are immutable. By contrast, interviewer expectations are influenceable and can thus be addressed — and modified — during interviewer training (Groves and Couper 1998). Positive changes in expectations could lead to desirable survey outcomes, such as fewer item nonresponses. Hence, findings about influenceable interviewer characteristics such as interviewer expectations could be considered in survey practice by implementing special training sessions that address these characteristics in a suitable way.

Like other interviewer characteristics, expectations differ among interviewers. These differences arise primarily from interviewers' different experiences in the past (Tolman 1932). One reason why interviewers' expectations matter when they ask respondents about their income and assets is that expectations drive verbal and non-verbal behavior during interactions, evoking behavior that make the expectations come true — a phenomenon that Merton (1948) called "self-fulfilling prophecy." Hyman (1954) argued that, before a survey, interviewers have a prior distribution of expected answers to questions in mind. These expectations influence the way they conduct the interviews, and thus the actual distribution of answers. Following the Thomas theorem —"If men define situations as real, they are real in their consequences" (Thomas 1928, 572)— interviewers (un)consciously act in accordance with their expectations in such a way that these expectations are fulfilled. For instance, if a respondent fails to provide a substantive answer to a question, interviewers who expect their respondents to answer may ask the same question again or try to obtain an adequate answer by probing. By contrast, interviewers who expect their respondents not to answer may fail to follow these rules of standardized interviewing.

These deliberations give rise to the hypothesis that the interviewers in the present study affected rates of nonresponse to income and asset questions. More specifically, it is predicted that respondents interviewed by interviewers who expected them to answer the income and asset questions were more likely to do so than respondents interviewed by interviewers who did not expect them to answer these questions.

Few studies have examined the effects of non-sociodemographic interviewer characteristics on nonresponse to income and asset questions. Bailar, Bailey, and Stevens (1977) investigated whether interviewers' opinions regarding the appropriateness of asking about sources and amounts of income had an impact on nonresponse to income questions. Sudman et al. (1977), Singer and Kohnke-Aguirre (1979), and Singer, Frankel, and Glassman (1983) examined whether the interviewer's judgement of the difficulty of obtaining sensitive information was associated with nonresponse to income questions. Overall, these studies were based on the theory of self-fulfilling prophecy but yielded little evidence that interviewer expectations are strongly related to nonresponse to income questions.

The results of these studies are difficult to generalize because the studies lacked conceptual, operational, statistical, and methodological power. For example, when examining interviewer effects, it is important to consider that respondents are nested within interviewers (Hox, de Leeuw, and Kreft 1991; Paterson and Goldstein 1991) — that is, they are at a lower level and are grouped around their interviewers at the upper level. Only when researchers consider this hierarchical data structure in their models can they clearly differentiate between respondent and interviewer effects and obtain reliable statistical estimates.

Two recent studies have considered both the hierarchical data structure of respondents nested within interviewers and interviewer characteristics that are related to item nonresponse to income questions (Pickery and Loosveldt 2001; Wuyts and Loosveldt 2017). Both studies examined whether, or to what extent, interviewer nonresponse to an income question was a predictor of respondent nonresponse to that question. Whereas Pickery and Loosveldt (2001) found no significant correlation between interviewer and respondent nonresponse to the income question, Wuyts and Loosveldt (2017) found that the odds of a respondent answering the income question doubled when the interviewer also answered that question. The inconsistency of these findings makes it clear that further research is required on the association between interviewer expectations and respondent nonresponse to income questions.

To contribute to filling this research gap, and to shed more light on effects related to interviewer expectations, this study examines interviewer effects on nonresponse to income and asset questions in a face-to-face survey by using two-level hierarchical logit regressions. The aim of the study is twofold: first, to determine the extent to which item nonresponse to financial questions is subject to interviewer effects; second, to examine whether interviewer expectations regarding the likelihood that respondents will provide a substantive answer to such questions matter in this context.

3.2 Data and Methods

3.2.1 Data

The present study used Austrian, Belgian, German, Spanish, and Swedish data from Wave 5 of the Survey of Health, Ageing and Retirement in Europe (SHARE)[4] (Börsch-Supan et al. 2013; Börsch-Supan 2017). SHARE is a multidisciplinary, cross-national panel survey that biannually collects microdata on the health, socioeconomic status, and social and family networks of individuals aged 50 years and older and their partners. Wave 5 of SHARE was conducted in 2013 in a total of 14 European countries (Austria, Belgium, Czech Republic, Denmark, Estonia, France, Germany, Italy, Netherlands, Luxembourg, Slovenia, Spain, Sweden, and Switzerland) and Israel. Samples from each country were

[4] This paper uses data from SHARE Wave 5 (10.6103/SHARE.w5.600); see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources. SHARE gratefully acknowledges additional funding (see www.share-project.org).

based on a probability sample that was representative of the non-institutionalized population aged 50 years and older (De Luca, Rossetti, and Malter 2013). The target persons and their partners were surveyed using computer-assisted personal interviewing (CAPI) (De Luca, Celidoni, and Trevisan 2013). All SHARE interviewers receive training. The training program aims to ensure reliability, consistency, generalization, and comparability of results across countries (Alcser and Benson 2005). During training, interviewers are instructed to probe after an initial "don't know" response (Groves et al. 2009).

A total of 27,038 individual interviews were conducted in the countries covered by the present study, Austria, Belgium, Germany, Spain, and Sweden (Bergmann et al. 2017). The response rate in Wave 5 for households that had participated in at least one previous wave ranged from 67% to 79% across these countries (Kneip, Malter, and Sand 2013). By contrast, the household response rate for refreshment samples, calculated in accordance with the standards set by the American Association for Public Opinion Research (AAPOR Response Rate 1), ranged from 34% to 60% across the above-mentioned five countries (Bergmann et al. 2017). A total of 642 interviewers were deployed in these countries.

As the present study investigated nonresponse to income and asset questions, only predefined subsamples numbering approximately 17,300 persons for the household income question and approximately 16,800 persons for the asset questions could be considered. If more than one individual was interviewed in a household, information on household income and assets was requested only from one person (De Luca, Rosetti, and Malter 2013). In addition, if physical and/or cognitive limitations made it too difficult for the respondent to answer question modules, a so-called proxy respondent could provide assistance or answer on the respondent's behalf (SHARE 2016). Proxy interviews were not included in the present analyses.

Before the interviewer asked the amount of a certain asset, such as a bank balance or interest or dividend income, a filter question was asked to ensure that the respondent or household possessed such an asset. Therefore, another 600 respondents, approximately, had to be excluded from the asset amount analyses because they did not have a bank account, mutual funds, bonds, or stocks.

In addition to data from the regular SHARE interviews, data on the SHARE interviewers were used. These data were collected within the framework of the SHARE Interviewer Survey (Blom and Korbmacher 2013) that was conducted in 2013 prior to the fieldwork for SHARE Wave 5 (Korbmacher et al. 2013). The data were collected in an online survey in six European countries (Austria, Belgium, Germany, Slovenia, Spain, and Sweden). All interviewers in these SHARE countries were invited by letter to participate in the survey after taking part in the national interviewer training sessions. Before fieldwork started, interviewers had to participate. Participation rates in 2013 ranged from 36% to 83% across countries. The SHARE interviewer ID, which was recorded in the regular SHARE interview and also provided in the SHARE Interviewer Survey, linked both datasets (Korbmacher et al. 2013). Slovenia, which had the lowest participation rate, was excluded from the present analyses.

When the respondent data were linked to the interviewer data, the sample sizes decreased, as only 421 of the 642 interviewers deployed in Austria, Belgium, Germany, Spain, and Sweden participated in the Interviewer Survey. Consequently, the subsample for the household income analyses decreased from 17,254 to 11,760 respondents; the subsample for the bank balance analyses decreased from 16,808 to 11,518; and the subsample for the interest and dividend income analyses decreased from 16,837 to 11,535. The number of observations reported in the Results section varies because certain information on the various respondent and interviewer covariates was missing.

3.2.2 Variables

Dependent variables. The dependent variables in the present study were nonresponse to one household income question and two asset questions. The first asset question was preceded by a question asking whether the respondent (and, if applicable, his or her spouse or partner) had a bank account, transaction account, saving account, or postal account. If yes, the respondent was asked to report the amounts that he or she (and, if applicable, his or her spouse or partner) currently had in these accounts. The respondent was then asked to report the amount of interest or dividend income received from savings in bank accounts, bonds, stocks, or mutual funds in the previous year. If the respondent did not know the answer or refused to answer, dichotomous unfolding bracket questions followed with different ranges of amounts. Only those individuals who did not answer the follow-up questions were classified as nonrespondents. Most of the respondents reported their household income, their bank balances, and their interest or dividend income (see table 3.1). However, some respondents were unable or refused to report the exact amounts. The share of these respondents across questions ranged from 8.5% to 22.3%.

Table 3.1. Response to household income and asset questions

Variable	Question text	Reported		Not reported	
		n	%	n	%
Household income	"How much was the overall income, after taxes and contributions that your entire household had in an average month in 2012?"	10,757	91.5	1,003	8.5
Bank balances	About how much do you[a] currently have in bank accounts, transaction accounts, saving accounts or postal accounts?	9,472	82.2	2,046	17.8
Interest/dividend income	Overall, about how much interest or dividend income did you[a] receive from your savings in bank accounts, bonds, stocks or mutual funds in 2012? Please give me the amount after taxes."	8,966	77.7	2,569	22.3

Note: Depending on the respondent's previous answers in the interview, the question text was adapted automatically. Hence, "you" may refer to the respondent or to the respondent and his or her spouse or partner. Numbers are based on linked respondents, excluding proxy respondents (n [total subsample] = 11,760, 11,518, and 11,535, respectively).

The dependent income and asset variables were coded as follows:

$$y_{ij}=\begin{cases}0 & amount\ reported \\ 1 & amount\ not\ reported\end{cases}, \quad (1)$$

where y_{ij} denotes a binary response variable of respondent *i* interviewed by interviewer *j*.

A group t-test showed that interviewer-specific item nonresponse rates did not differ significantly at the five percent level between respondents who could be linked to an interviewer and respondents who could not (Appendix, table A3.1).

Explanatory variables. The main explanatory variable in the present study was the interviewer's expectation as to the likelihood that his or her respondents would provide a substantive answer to questions about their income. To measure this expectation, interviewers were asked the following question in the online Interviewer Survey: "Social surveys very often ask for respondents' income. What do you expect, how many of your respondents (in percentage) in SHARE will provide information about their income?"

Interviewers were not specifically asked for their expectations regarding response to asset questions. However, the measurement of expectations regarding the likelihood that respondents would provide information about their income could serve as a proxy. A numerical answer ranging from 1% to 100% was requested. The answers were categorized for the analyses (table 3.2). The categories were chosen for two reasons: first, because interviewers entered mostly rounded values; second, to overcome the constraint caused by the fact that the distribution of the variable was left-skewed, with a mean of 75.5% and a median of 80%.

Table 3.2. Distribution of interviewers' expectations regarding income reporting

Percentage of respondents expected to report income	No.	%
50% or fewer	57	13.5
51—75%	112	26.6
76—90%	144	34.2
91—100%	76	18.1
Don't know or refuse to say	32	7.6

Note: N = 421 interviewers.

Control variables. The following control variables at respondent level were included in the analyses: gender, age, education, household composition, marital status, number of children, employment status, home ownership, area of residence, and country of interview. Another covariate that explained the outcomes at respondent level was the respondent's score on a recall test, which measured cognitive skills. This criterion was likely to be related to response to income and assets questions. In addition, two interview situation variables—the presence of others and the respondent's overall willingness to answer questions during the interview—which were coded by the interviewer, were considered because they may also have influenced the respondent's answers. Finally, because the underlying mechanisms of responding to financial questions may differ between respondents who had participated in SHARE at least once before and respondents who participated for the first time, a dummy variable was included that indicated whether the respondent belonged to the panel or the refreshment sample.

Covariates at interviewer level were gender, age, education, experience, and self-assessed health. Additionally, the following variables that reflected the interviewer's style in general and may have influenced the respondent's reporting behavior were included: self-assessed interview style, the reasons for being an interviewer, trust in people, and use of social networks and online banking. In SHARE, interviewers are assigned to individuals who reside in the same area. Such a non-interpenetrated sample hampers the distinction between area and interviewer effects (Schnell and Kreuter 2005). To address this problem, the models controlled additionally for the respondent's household and area characteristics. This modeling approach assumes that the additional control variables partially filter out potential area effects (Beullens and Loosveldt 2014). The full models are provided in Appendix, table A3.2.

3.2.3 Analytical Model

When respondents are nested within interviewers—that is, when they are at a lower level (level 1) and grouped around their interviewers at the upper level (level 2)—a multilevel approach is appropriate for addressing this underlying hierarchical structure (Hox, Moerbeek, and van den Schoot 2010). This technique enables a clear separation of interviewer effects from respondent effects by allowing the regression intercepts to differ randomly across groups.

Country-specific effects were modeled with country-fixed effects. An alternative approach would have been to consider country-specific effects by implementing a third level. However, several researchers have shown that including higher levels with too few units can lead to biased estimates and confidence intervals (Maas and Hox 2005; McNeish and Stapleton 2016).

The three dependent variables were treated as binary in the multilevel models. The response probabilities are denoted by Pr(y_{ij}=1) and are related to respondent *i* characteristics and interviewer *j* characteristics. Against this background, the logistic multilevel model was as follows:

$$\mathrm{Log}\left(\frac{\Pr(y_{ij}=1)}{\Pr(y_{ij}=0)}\right)= \gamma_{00}+\gamma_{10}X_{ij}+ \gamma_{01}Z_{j}+u_{0j} \qquad (2).$$

The dependent variables were explained by the random intercept γ_{00}, the explanatory variables of the respondent X_{ij} and the interviewer Z_j, and the residual error terms u_{0j} at level 2.

To quantify the extent to which the interviewers influenced the respondents' answers, the intraclass correlation coefficient (ICC) was calculated as follows for each dependent variable in the random intercept models without explanatory variables:

$$\text{ICC}=\sigma_u^2 \,/\, (\sigma_u^2 + \pi^2/3), \tag{3}$$

where σ_u^2 was defined as the variance at level 2. The variance at level 1 is fixed to $\pi^2/3$ in logistic multilevel regressions. The ICC ranges between 0 and 1. An ICC of 0 indicates that no variance is attributable to the interviewer, whereas a value of 1 means that all variance is attributable to the interviewer. Therefore, the higher that the value is, the more influence the interviewer had on the respondent's item nonresponse.

In multilevel logistic regressions, variance components cannot be compared across models with and without explanatory variables because of the fixed level 1 variance (Hox, Moerbeek, and van de Schoot 2010). Therefore, we followed Hox and colleagues' (2010) approach and calculated a scale correction factor for each model with explanatory variables. The scale-corrected variances allowed comparisons and further calculations.

3.3 Results

The analyses revealed that item nonresponse rates for all three financial questions (household income, bank balances, and interest or dividend income) were subject to interviewer effects (see table 3.3). This fact is reflected in the ICCs in the empty models, which is 0.39 for the household income, 0.41 for the bank balance, and 0.45 for the interest/dividend question. The rescaled explained variances at the interviewer level ranged from 16% to 25% after controlling for respondent characteristics, and from 28% to 41% after controlling for respondent and interviewer characteristics.

Table 3.3. Interviewer variance of different random intercept models on item nonresponse to household income and asset questions

	Household income			Bank balances			Interest/dividend income		
	σ_u^2	s.e.	explained σ_u^2 in %	σ_u^2	s.e.	explained σ_u^2 in %	σ_u^2	s.e.	explained σ_u^2 in %
Random intercept only	2.11	.27	N/A	2.54	.27	N/A	2.71	.29	N/A
Random intercept with level 1 variables	1.77	.24	16	2.01	.23	25	2.26	.25	21
Random intercept with level 1 and level 2 variables	1.35	.20	41	1.78	.21	33	2.06	.23	28

Note: Scale correction factors for the variances were .96, .97, and .97, respectively, in models with explanatory variables. s.e.= standard error.

Table 3.4 presents the coefficients for three multilevel logistic regression models showing interviewer effects only. The full models can be found in Appendix, table A3.2. Overall, none of the interviewers' sociodemographic characteristics were significantly correlated with item nonresponse (table 3.4). Self-assessed health status was significantly correlated at the five percent level with item nonresponse to the household income and bank balance questions. Interviewers who self-assessed their health as good or as very good or excellent obtained significantly higher item nonresponse rates than those who assessed their health as poor or fair. Moreover, only a few interviewer variables that reflected the self-assessed interviewer style or the reasons for being an interviewer were significantly correlated with "don't know" responses and refusals. For instance, interviewers who reported that they spoke faster if they noticed that the respondent was in hurry tended to obtain more item nonresponses to the household income question than those who reported that they did not adapt their speaking speed in that case. No other significant correlations at the five percent level were observed between self-assessed interviewer style or reasons for being an interviewer and item nonresponse. However, in all models that investigated item nonresponse to income and asset questions (table 3.4), interviewers who had positive expectations (more than 50%) as to the proportion of their respondents who would provide information about their income obtained lower nonresponse rates than interviewers with less positive expectations (fewer than 50%) in this regard.

Table 3.4. Estimated interviewer-level coefficients for the three logistic regression models on item nonresponse to financial questions

Interviewer-level variables	Household income	Bank balances	Interest/ dividend income
Gender: female (ref.)			
- male	−.16	–.15	−.35
	(.18)	(.18)	(.19)
Age in years	−.06	−.02	−.04
	(.05)	(.05)	(.06)
Age squared in years	.00	.00	.00
	(.00)	(.00)	(.00)
Education level: low (ref.)			
- medium	−.16	−.08	−.03
	(.30)	(.34)	(.32)
- high	−.19	.10	.17
	(.27)	(.33)	(.30)
Experience in years	−.06	−.02	−.04
	(.01)	(.01)	(.01)
Self-assessed health: poor or fair (ref.)			
- good	1.00**	.83*	.31
	(.34)	(.34)	(.34)
- very good or excellent	1.22***	.94**	.48
	(.33)	(.34)	(.34)
Interviewer style			
Explains question: not really/at all (ref.)			
- somewhat/perfectly	−.03	.16	.06
	(.18)	(.18)	(.18)
Shortens question: not really/at all (ref.)			
- somewhat/perfectly	−.05	−.11	−.04
	(.26)	(.31)	(.34)
Speaks faster: not really/at all (ref.)			
- somewhat/perfectly	.47**	.06	.26
	(.17)	(.17)	(.18)
Completes answers: not really/at all (ref.)			
- somewhat/perfectly	.39	.23	-.00
	(.48)	(.48)	(.39)
Sticks to instructions: not really/at all (ref.)			
- somewhat/perfectly	.19	-.23	−.02
	(.31)	(.33)	(.34)
Reasons for being an interviewer			
Interesting work: not that important (ref.)			
- very important	−.27	−.34	−.26
	(.24)	(.24)	(.26)
Interact with people: not that important (ref.)			
- very important	.37	.25	.06
	(.20)	(.21)	(.21)
Research involvement: not that important (ref.)			
- very important	−.19	.10	.15
	(.22)	(.25)	(.22)
Compensation: not that important (ref.)			
- very important	−.16	−.10	−.26
	(.19)	(.19)	(.20)

Table 3.4. (cont.)

	Household income	Bank balances	Interest/ dividend income
Trust in people	.07	.05	.01
	(.05)	(.05)	(.05)
Online banking: no (ref.)			
- yes	−.13	.01	.08
	(.19)	(.20)	(.20)
Social networks: no (ref.)			
- yes	−.34	−.18	−.18
	(.18)	(.18)	(.18)
Expected % of respondents reporting income: 50% or fewer (ref.)			
- 51–75%	−.76**	−.76**	−.98**
	(.27)	(.28)	(.31)
- 76%-90%	−.68**	−.64*	−.95***
	(.25)	(.26)	(.28)
- 91%–100%	−.60	−.83**	−.93**
	(.32)	(.29)	(.34)
Don't know/	.07	−.37	−.84*
Refuse to say	(.35)	(.41)	(.41)
N (respondents)	10,618	10,236	10,251
N (interviewers)	363	363	363
Log likelihood	−2,124.08	−3,462.20	−3,846.45
Chi-squared	319.60	346.59	375.91

Note: Standard errors in parentheses. ref. = reference category. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The robustness of the country-fixed effects in the models was tested by means of separate country-level analyses. The number of covariates had to be reduced to avoid the collapse of the models. Separate analyses could not be run for Sweden due to the small size of the interviewer sample and the number of covariates still in the reduced models. In sum, the separate country-level analyses yielded similar results and allowed similar conclusions. Coefficients of the interviewer expectations were all in the expected direction, although not all were significant at the five percent level for all countries and for the effects of all categories (see separate country-level analyses in table B.1 in the online supplementary materials, see Appendix A3.3). The size of the ICCs and the reduction of variance varied slightly by country. However, significant coefficients of the country-fixed effects in the full models support the observation that the influence of interviewers on item nonresponse varied slightly by country (Appendix, table A3.2). The reduction of the interviewer variance by including respondent and interviewer characteristics in the separate country-level analyses was mostly comparable to that in the models with country fixed effects. Overall, the results of the robustness tests yielded similar conclusions.

Average marginal effects (AME) were calculated to evaluate these correlations more appropriately. The standard coefficients in logistic models indicate only the effect direction; they provide no information on the effect size. AMEs represent the average change in probability when the variable predictor increases by one unit (Mood 2010). The probability that a respondent would not know the answers to, or refuse to answer, the questions about household income, bank balances, and interest or dividend income was lower if the interviewer had a positive expectation (more than 50%) regarding the proportion of respondents who would give a substantive answer to income questions than if the interviewer had a less positive expectation (50% or lower) in this regard (see table 3.5).

Table 3.5. Average marginal effects of interviewer expectations on item nonresponse

	Household income	Bank balances	Interest/dividend income
Percentage of respondents expected to report income			
50% or fewer (ref.)			
51–75%	−.051*	−.088**	−.129**
	(.020)	(.033)	(.042)
76–90%	−.047*	−.075*	−.125**
	(.019)	(.032)	(.040)
91–100%	−.042	−.094**	−.124**
	(.022)	(.034)	(.046)
Don't know/	.005	.046	−.113*
Refuse to say	(.030)	(.049)	(.053)
N (respondents)	10,618	10,236	10,251
N (interviewers)	363	363	363

Note: Standard errors in parentheses. ref. = reference category. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The strongest effects were found for the models that investigated interest or dividend income. The probability of a respondent not reporting interest or dividend income was 13 percentage points lower if the interviewer expected between 76% and 90% of respondents to provide information about their income than if the interviewer expected 50% or fewer respondents to do so. By contrast, the probability of a respondent not reporting household income was four percentage points lower if the interviewer expected between 76% and 90% of respondents to report their income than if the interviewer expected 50% or fewer respondents to do so.

3.4 Discussion

Item nonresponse in surveys of older populations (SHARE's target population consists of persons aged 50 years and older) may be more problematic than in surveys of the general population. Working memory declines with age, and this could lead to higher item nonresponse rates—especially when items are cognitively demanding and sensitive. In this study, the rate of nonresponse to financial questions ranged from 9% to 22%. The rate of nonresponse to the household income question in the present SHARE dataset was about the same as the rate of nonresponse to the gross income question in the British Household Panel Survey (BHPS), which targeted household members aged 16 years and over. This observation indicates that surveys that target middle-aged and elderly adults do not necessarily obtain higher item nonresponse rates than surveys that target the general population.

Multilevel analyses were conducted to determine the extent to which interviewers affected item nonresponse to the questions about household income, bank balances, and interest or dividend income. Overall, interviewers influenced the item nonresponse rates considerably. This large effect was reflected in the ICCs, which ranged from 0.37 to 0.45 across the various financial questions. The ICCs were slightly higher than those found in previous studies, but they were in line with the literature (Schräpler 2004; Riphahn and Serfling 2005; Essig and Winter 2009).

These slightly higher ICCs indicate that interviewers in the present study had more influence on respondents' probability to provide a substantive answer to income and asset questions than in other studies. This difference might be explained by SHARE’s specific target population (persons aged 50 years or older). This is supported by Groves and Magilavy (1986), who found greater interviewer-associated variance in telephone surveys among elderly respondents, suggesting that they were more susceptible to interviewer effects. Interviewers may play a particular role in surveys of older populations, because older respondents are more likely to make digressions than younger ones (Belli and Chardoul 1997) and to need more assistance in the question-and-answer-process. Moreover, research has shown that interviewers tend to deviate from the standardized interview script when interviewing older respondents (Belli et al. 1999). Both tendencies could increase interviewer-associated variance.

To explain the interviewer effects found in the present study, the interviewer data collected in the interviewer survey conducted prior to the beginning of the fieldwork were linked to the respondents’ data. Interviewer-level variance decreased when both the respondent and the interviewer data were considered, and between 35% and 40% of the interviewer variances could be explained.

Few of the sociodemographic interviewer characteristics were significantly correlated with item nonresponse to the income and asset questions. This result indicates that characteristics other than sociodemographic characteristics must be considered. Interviewer characteristics that were relevant in this context were interviewer expectations as to the proportion of respondents who would provide information about their income. For all income and asset questions, the level of interviewer expectations significantly influenced the respondent’s likelihood to provide substantive answers. These significant associations support the hypothesis that interviewers’ expectations regarding respondents’ likelihood to report their income are correlated with item nonresponse to financial questions. Respondents were more likely to report their income and assets when interviewed by an interviewer who expected more than 50% of his or her respondents to report their income than when interviewed by an interviewer who expected 50% or fewer of his or her respondents to do so. These associations, which indicate that interviewer expectations have a moderate influence in this context, are much stronger than those found in previous studies (Bailar, Bailey, and Stevens 1977; Singer and Kohnke-Aguirre 1979; Singer et al. 1983). In the present study, positive interviewer expectations that respondents would provide information on their income were found to reduce item nonresponse by up to 13 percentage points. With an average response rate of 85% for the three questions investigated, this effect is substantial. This result indicates that interviewers—and especially their expectations as to whether respondents will provide substantive answers—play a crucial role in complex surveys of older populations.

To reduce nonresponse to financial questions, interviewer training could address the association between interviewer expectations and item nonresponse and emphasize how interviewers should interact with respondents when asking questions about income and assets. This approach would increase interviewers' confidence in their ability to obtain substantive answers. In addition, information on data anonymity and confidentiality could help to change interviewers' attitudes. Being more informed about data protection could also help interviewers to handle respondents who have privacy concerns and do not want to report their income and assets. Although such training modifications may not help to align item nonresponse rates across interviewers, they may help to minimize item nonresponse rates generally.

These conclusions about interviewer training rely on observed correlations. The present study did not test in an experimental design whether high expectations on the part of interviewers that their respondents would provide information about their income led to fewer nonresponses to income and assets questions. Nor did it test whether interviewer training sessions could reduce item nonresponse and the variability of item nonresponse rates across interviewers. However, other studies have shown in experimental designs that interviewer training can help to reduce rates of item nonresponse (e.g., Billiet and Loosveldt 1988). Therefore, to validate our findings, future experimental studies should focus on examining the reduction in variability with respect to interviewer expectations.

The generalizability of the present results is subject to certain limitations. First, the expectations hypothesis proposed in the present study, which was based on self-fulfilling prophecy theory, could, strictly speaking, be tested only for item nonresponse to questions about income in general. In the Interviewer Survey, interviewers were asked what proportion of their respondents they expected would provide information about their income. They were not asked about their expectations as to whether respondents would report specific types of income or assets, such as bank balances, interest or dividend income. Hence, the variables used to test the hypothesis did not perfectly reflect the underlying theoretical concept. Second, the information on interviewers was collected before the fieldwork started. However, interviewers may change their expectations and attitudes during fieldwork and adapt their behavior accordingly. It was not possible to investigate possible discrepancies between the answers given in the Interviewer Survey and interviewers' actual behavior. Such discrepancies can lead to over- or underestimation of interviewer effects. Third, no information was available on the interview situation itself. It was not known how interviewers actually behaved during the question-and-answer-process because no audio-recorded data were available. For instance, the present results show that very few variables relating to self-assessed interviewer style were significantly correlated with item nonresponse. This led to the conclusion that, for some older respondents, tailoring the standardized script may avoid (or evoke) item nonresponse, whereas for other older respondents it may not. However, another reason for the very few significant correlations between item nonresponse and interviewer style could be that interviewers provided inaccurate or socially desirable self-assessments. Interviewers may have thought that they should provide certain answers, or they may have misjudged themselves or assessed themselves differently. Future studies should try to rely additionally on audio-recorded data to shed more light on the question-and-answer-process when respondents are asked about their income and assets.

In spite of its limitations, this study certainly adds to our understanding of the association between interviewer expectations and item nonresponse to income and asset questions. Compared to prior research, it used more robust information, the dataset was larger, and a suitable analytical approach was adopted that considered the underlying hierarchical structure of the data. In addition, information on the household and the area of residence were taken into account in order to separate respondent effects from interviewer and/or area effects. These advantages allow us to state that associations exist between nonresponse to financial questions and interviewer characteristics—in particular interviewer expectations regarding respondents' likelihood to provide information about their income.

Appendix

Table A3.1. Distribution of interviewer-specific item nonresponse rates by linkage status

	Linked	N	Mean	SD	*p*-value
Household income	yes	421	0.09	0.01	
	no	221	0.09	0.01	0.74
Bank balances	yes	421	0.19	0.01	
	no	221	0.22	0.02	0.13
Interest/dividend	yes	421	0.23	0.24	
income	no	221	0.26	0.25	0.10

Table A3.2. Estimated coefficients for the three logistic regression models on item nonresponse

	Household income	Bank balances	Interest/dividend income
Respondent-level characteristics			
Gender: female (ref.)			
- male	−.33***	−.38***	−.44***
	(.09)	(.07)	(.06)
Age in years	−.07	−.01	−.07
	(.06)	(.04)	(.04)
Age squared in years	.00	.00	.00*
	(.00)	(.00)	(.00)
Education level: low (ref.)			
- medium	.40***	.12	.13
	(.12)	(.09)	(.08)
- high	.50***	.11	.23**
	(.13)	(.09)	(.09)
Household size: single HH (ref.)			
- Two–person HH	−.89***	−.04	−.17
	(.16)	(.12)	(.12)
- multiple HH members	.03	.02	-.05
	(.11)	(.09)	(.09)
Marital status: - married/partnership (ref.)			
- separated/divorced	−.03	−.38**	−.40***
	(.16)	(.13)	(.11)
- never married	−.08	−.56***	−.09
	(.21)	(.16)	(.13)
- widowed	.21	-.14	−.06
	(.18)	(.14)	(.12)
Number of Children	−.07	−.07**	−.05
	(.04)	(.03)	(.03)
Employment status: unemployed (ref.)			
- employed	.22	.19*	.28***
	(.12)	(.10)	(.08)
Home ownership: no (ref.)			
- yes	.24*	.43***	.40***
	(.11)	(.08)	(.08)
Recall test: first quartile (ref.)			
- second quartile	−.15	−.03	−.12
	(.11)	(.10)	(.10)
- third quartile	−.09	.03	−.05
	(.13)	(.10)	(.10)
- fourth quartile	−.26	−.24	−.13
	(.17)	(.12)	(.12)

Table A3.2. (cont.)

	Household income	Bank balances	Interest/dividend income
Presence of others: no (ref.)			
- yes	.31**	.02	.05
	(.11)	(.07)	(.08)
Willingness to participate: low/fair/mixed (ref.)			
- (very) good	−1.45***	−1.24***	−1.13***
	(.19)	(.13)	(.14)
Area of residence: big city (ref.)			
- suburbs or outskirts	−.24	−.11	.12
	(.24)	(.17)	(.14)
- large town	-.06	.01	.25
	(.22)	(.16)	(.14)
- small town	.04	.08	.20
	(.19)	(.17)	(.15)
- rural area or village	.29	.25	.34*
	(.18)	(.16)	(.15)
- unknown	.47*	.38	.38
	(.22)	(.20)	(.21)
Panel member: no (ref.)			
- yes	−.32*	−.40***	−.40***
	(.13)	(.11)	(.10)
Country: Austria (ref.)			
Germany	.52	.83*	.67
	(.30)	(.36)	(.35)
Sweden	-.20	.36	.76
	(.41)	(.41)	(.41)
Spain	.87*	1.36***	1.43***
	(.36)	(.37)	(.38)
Belgium	.47	1.23***	1.48***
	(.33)	(.36)	(.35)
Interviewer-level characteristics			
Gender: female (ref.)			
- male	−.16	−.15	−.35
	(.18)	(.18)	(.19)
Age in years	−.06	-.02	−.04
	(.05)	(.05)	(.06)
Age squared in years	.00	.00	.00
	(.00)	(.00)	(.00)
Education level: low (ref.)			
- medium	−.16	−.08	−.03
	(.30)	(.34)	(.32)
- high	−.19	.10	.17
	(.27)	(.33)	(.30)
Experience in years	−.01	.00	.00
	(.01)	(.01)	(.01)
Self-assessed health: poor or fair (ref.)			
good	1.00**	.83*	.31
	(.34)	(.34)	(.34)
very good or excellent	1.22***	.94**	.48
	(.33)	(.34)	(.34)

Table A3.2. (cont.)

	Household income	Bank balances	Interest/dividend income
Interviewer style			
Explains question: not really/at all (ref.)			
- somewhat/perfectly	−.03	.16	.06
	(.18)	(.18)	(.18)
Shortens question: not really/at all (ref.)			
- somewhat/perfectly	−.05	−.11	−.04
	(.26)	(.31)	(.34)
Speaks faster: not really/at all (ref.)			
- somewhat/perfectly	.47**	.06	.26
	(.17)	(.17)	(.18)
Completes answers: not really/at all (ref.)			
- somewhat/perfectly	.39	.23	−.00
	(.48)	(.48)	(.39)
Sticks to instructions: not really/at all (ref.)			
- somewhat/perfectly	.19	−.23	−.02
	(.31)	(.33)	(.34)
Reasons for being an interviewer			
Interesting work: not that important (ref.)			
- very important	−.27	−.34	−.26
	(.24)	(.24)	(.26)
Interact with people: not that important (ref.)			
- very important	.37	.25	.06
	(.20)	(.21)	(.21)
Research involvement: not that important (ref.)			
- very important	−.19	.10	.15
	(.22)	(.25)	(.22)
Compensation: not that important (ref.)			
- very important	−.16	−.10	−.26
	(.19)	(.19)	(.20)
Trust in people	.07	.05	.01
	(.05)	(.05)	(.05)
Online banking: no (ref.)			
yes	−.13	.01	.08
	(.19)	(.20)	(.20)
Social networks: no (ref.)			
yes	−.34	−.18	−.18
	(.18)	(.18)	(.18)
Expected % of respondents reporting income: 50% or fewer (ref.)			
51–75%	−.76**	−.76**	−.98**
	(.27)	(.28)	(.31)
76%–90%	−.68**	−.64*	−.95***
	(.25)	(.26)	(.28)
91%–100%	−.60	−.83**	−.93**
	(.32)	(.29)	(.34)
Don't know/refuse to say	.07	−.37	−.84*
	(.35)	(.41)	(.41)
N (respondents)	10,618	10,236	10,251
N (interviewer)	363	363	363
Log likelihood	−2,381.90	−3,742.58	−4,148.46
Chi-squared	364.20	429.04	425.06

Note: Standard errors in parentheses. ref.= reference category. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

A3.3 Supplementary materials

Supplementary materials are available online at http://www.oxfordjournals.org/our_journals/jssam/.

References

Alcser, Kirsten H., and Grant Benson. 2005. "The SHARE Train-the-Trainer Program." In *The Survey of Health, Aging, and Retirement in Europe – Methodology*, edited by Axel Börsch-Supan and Hendrik Jürges, 70–74. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging (MEA).

Bailar, Barbara, Leroy Bailey, and Joyce Stevens. 1977. "Measures of Interviewer Bias and Variance." *Journal of Marketing Research* 14 (3): 337–343. https://doi.org/10.1177/002224377701400309.

Belli, Robert F., and Stephanie A. Chardoul. 1997. "The Digression of Survey Actors in a Face-to-face Health Interview." Paper presented at the 52nd Annual Conference of the American Association for Public Opinion Research, Norfolk, VA, May.

Belli, Robert F., Paul S. Weiss, and James M. Lepkowski. 1999."Dynamics of Survey Interviewing and the Quality of Survey reports: Age Comparisons." In *Cognition, Aging, and Self-Reports*, edited by Norbert Schwarz, Denise C. Park, Bärbel Knäuper and Seymour Sudman, 262–282. New York: Psychology Press.

Bergmann, Michael, Thorsten Kneip, Giuseppe De Luca, and Annette Scherpenzeel. 2017. *Survey Participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1–6*. SHARE Working Paper Series 31-2017. Munich, Germany: Munich Center for the Economics of Aging (MEA), Max Planck Institute for Social Law and Social Policy. Available at: http://www.share-project.org/uploads/tx_sharepublications/ WP_Series_31_2017_Bergmann_et_al.pdf (accessed February 2018).

Beullens, Koen, and Geert Loosveldt. 2014. "Interviewer Effects on Latent Constructs in Survey Research." *Journal of Survey Statistics and Methodology* 2 (4): 433–458. https://doi.org/10.1093/jssam/smu019.

Billiet, Jacques, and Geert Loosveldt. 1988. "Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training." *Public Opinion Quarterly* 52 (2): 190–211. https://doi.org/10.1086/269094.

Blom, Annelies G., and Julie M. Korbmacher. 2013. "Measuring Interviewer Characteristics Pertinent to Social Surveys: A Conceptual Framework." *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2013-00001. Available at: https://surveyinsights.org/?p=817.

Börsch-Supan, Axel 2017. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release Version: 6.0.0*. SHARE-ERIC. https://doi.org/10.6103/SHARE.w5.600.

Börsch-Supan, Axel , Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie M. Korbmacher, F.rederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology* 42 (4): 992–1001. https://doi.org/10.1093/ije/dyt088.

Colsher, Patricia L., and Robert B. Wallace. 1989. "Data Quality and Age: Health and Psychobehavioral Correlates of Item Nonresponse and Inconsistent Responses." *Journal of Gerontology* 44 (2): 45–52. https://doi.org/10.1093/geronj/44.2P45.

Dahlhamer, James M., Marcie L. Cynamon, Jane F. Gentleman, Andrea L. Piani, and Michael J. Weiler. 2010. "Minimizing Survey Error through Interviewer Training: New Procedures Applied to the National Health Interview Survey (NHIS)." Paper presented at the Joint Statistical Meetings, Vancouver, British Columbia, August.

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

De Luca, Giuseppe, Martina Celidoni, and Elisabetta Trevisan. 2013. "Item Nonresponse and Imputation Strategies in SHARE Wave 5." In *SHARE Wave 5: Innovations & Methodology*, edited by Frederic Malter and Axel Börsch-Supan, 85–100. Munich, Germany: MEA, Max Planck Institute for Social Law and Social Policy.

De Luca, Giuseppe, Claudio Rossetti, and Frederic Malter. 2013. "Sample Design and Weighting Strategies in SHARE Wave 5," In *SHARE Wave 5: Innovations & Methodology*, edited by Frederic Malter and Axel Börsch-Supan, 75–84. Munich, Germany: Munich Center for the Economics of Aging, Max Planck Institute for Social Law and Social Policy.

Dykema, Jennifer, James M. Lepkowski, and Steven Blixt. 1997. "The Effect of Interviewer and Respondent Behavior on Data Quality: Analysis of Interaction Coding in a Validation Study." In *Survey Measurement and Process Quality*, edited by Lars E. Lyberg, Paul P. Biemer, Martin Collins, Edith D. de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin 287–310. New York: John Wiley.

Essig, Lothar, and Joachim K. Winter. 2009. "Item Non-Response to Financial Questions in Household Surveys: An Experimental Study of Interviewer and Mode Effects." *Fiscal Studies* 30 (3–4): 367–390. https://doi.org/10.1111/j.1475-5890.2009.00100.x.

Fowler, Floyd J. Jr., and Thomas W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Thousand Oaks, CA: SAGE.

Groves, Robert M. 1989. *Survey Errors and Survey Costs*. New York: John Wiley.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview*. New York: John Wiley.

Groves, Robert M., Floyed J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. New York: John Wiley.

Groves, Robert M., and Lou J. Magilavy. 1986. "Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys." *Public Opinion Quarterly* 50 (2): 251–266. https://doi.org/10.1086/268979.

Hayden, Kathleen M., Bruce R. Reed, Jennifer J. Manly, Douglas Tommet, Robert H. Pietrzak, Gordon J. Chelune, Frances M. Yang, Andrew J. Revell, David A. Bennett, and Richard N. Jones. 2011. "Cognitive Decline in the Elderly: An Analysis of Population Heterogeneity." *Age and Ageing* 40 (6): 684–689. https://doi.org/10.1093/ageing/afr101.

Hox, Joop J., Edith D. de Leeuw, and Ita G. G. Kreft. 1991. "The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 439–461. New York: John Wiley.

Hox, Joop J., Mirjam Moerbeek, and Rens van de Schoot. 2010. *Multilevel Analysis: Techniques and Applications*. New York: Routledge.

Hyman, Herbert H. 1954. *Interviewing in Social Research*, Chicago: University Press.

Josten, Michael, and Mark Trappmann. 2016. "Interviewer Effects on a Network-Size Filter Question." *Journal of Official Statistics* 32 (2): 349–373. https://doi.org/10.1515/jos-2016-0020.

Knäuper, Barbara, Robert F. Belli, Daniel H. Hill, and A. Regula Herzog. 1997. "Question Difficulty and Respondents' Cognitive Ability: The Effect on Data Quality." *Journal of Official Statistics* 13 (2): 181–199.

Kneip, Thorsten, Frederic Malter, and Gregor Sand. 2013. "Fieldwork Monitoring and Survey Participation in Fifth Wave of SHARE," In *SHARE Wave 5: Innovations & Methodology*, edited by Frederic Malter and Axel Börsch-Supan, 101–157. Munich, Germany: Munich Center for the Economics of Aging, Max Planck Institute for Social Law and Social Policy.

Korbmacher, Julie M., Sabine Friedel, Melanie Wagner, and Ulrich Krieger. 2013. "Interviewing Interviewers: The SHARE Interviewer Survey," In *SHARE Wave 5: Innovations & Methodology*, edited by Frederic Malter and Axel Börsch-Supan, 67–74. Munich, Germany: Munich Center for the Economics of Aging, Max Planck Institute for Social Law and Social Policy.

Maas, Cora J. M., and Joop J. Hox. 2005. "Sufficient Sample Sizes for Multilevel Modeling." *Methodology* 1 (3): 86–92. https://doi.org/10.1027/1614-2241.1.3.86.

McNeish, Daniel, and Laura M. Stapleton. 2016. "Modeling Clustered Data with Very Few Clusters." *Multivariate Behavioral Research* 51 (4): 495–518. https://doi.org/10.1080/00273171.2016.1167008.

Merton, Robert K. (1948). "The Self-Fulfilling Prophecy." *The Antioch Review* 8 (2): 193–210. http://dx.doi.org/10.2307/4609267.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29 (4): 199–226. http://dx.doi.org/10.1257/jep.29.4.199.

Milanović, Zoran, Saša Pantelić, Nebojša Trajković, Goran Sporiš, Radmila Kostić, and Nic James. 2013. "Age-related Decrease in Physical Activity and Functional Fitness among Elderly Men and Women." *Clinical Interventions in Aging* 8:549–556. https://doi.org/10.2147/CIA.S44112.

Mood, Carina 2010. "Logistic regression: Why We Cannot Do What We Think We Can Do, And What We Can Do About It." *European Sociological Review* 26 (1): 67–82. https://doi.org/10.1080/0141192910170408.

Paterson, Lindsay, and Harvey Goldstein. 1991. "New Statistical Methods for Analysing Social Structures: An Introduction to Multilevel Models." *British Educational Research Journal* 17 (4): 387–393. https://doi.org/10.1080/0141192910170408.

Pickery, Jan, and Geert Loosveld. 2001. "An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse." *Journal of Official Statistics* 17 (3): 337–350.

Pickery, Jan, and Geert Loosveldt. 2004. "A Simultaneous Analysis of interviewer Effects on Various Data Quality Indicators with Identification of Exceptional Interviewers." *Journal of Official Statistics* 20 (1): 77–89.

Riphahn, Regina T., and Oliver Serfling. 2005. "Item Non-Response on Income and Wealth Questions." *Empirical Economics* 30 (2): 521–538. https://doi.org/10.1007/s00181-005-0247-7.

Schaeffer, Nora C. 1991. "Conversation with a Purpose—or Conversation? Interaction in the Standardized Interview." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 365–391, New York: John Wiley.

Schnell, Rainer, and Frauke Kreuter. 2005. "Separating Interviewer and Sampling-Point Effects." *Journal of Official Statistics* 21 (3): 389–410.

Schräpler, Jörg-Peter. 2004. "Respondent Behavior in Panel Studies. A Case Study for Income Nonresponse by Means of the German Socio-Economic Panel (SOEP)." *Sociological Methods & Research* 33 (1): 118–156. https://doi.org/10.1177/0049124103262689.

Singer, Eleanor, Martin R. Frankel, and Marc B. Glassman. 1983. "The Effect of Interviewer Characteristics and Expectations on Response." *Public Opinion Quarterly* 47 (1): 68–83. https://doi.org/10.1086/268767.

Singer, Eleanor, and Luane Kohnke-Aguirre. 1979. "Interviewer Expectation Effects: A Replication and Extension." *Public Opinion Quarterly* 43 (2): 245–260. https://doi.org/10.1086/268515.

Sudman, Seymour, Norman M. Bradburn, Edward Blair, and Carol Stocking. 1977. "Modest Expectations: The Effects of Interviewers' Prior Expectations on Responses." *Sociological Methods & Research* 6 (2): 171–182. https://doi.org/10.1177/004912417700600203.

SHARE. 2016. "SHARE Release Guide 5.0.0." Munich: Max-Planck-Institue for Social Law and Social Policy, Munich Center for the Economics of Aging.

Thomas, William I. 1928. "The Methodology of Behavior Study." In *The Child in America: Behavior Problems and Programs*, edited by William I. Thomas and Dorothy S. Thomas, 553–576. New York: Knopf.

Tolman, Edward C. 1932. *Purposive Behavior in Animals and Men*. Berkeley: University of California Press.

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

van der Zouwen, Johannes, Wil Dijkstra, and Johannes H. Smit. 1991. "Studying Respondent-Interviewer Interaction: The Relationship between Interviewing Style, Interviewer Behavior, and Response Behavior." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymor Sudman, 419–437, New York: John Wiley.

West, Brady T., and Annelies G. Blom. 2017. "Explaining Interviewer Effects: A Research Synthesis." *Journal of Survey Statistics and Methodology* 5 (2): 175–211. https://doi.org/10.1093/jssam/smw024.

Wuyts, Celine, and Geert Loosveldt. 2017. "The Interviewer in The Respondent's Shoes: What Can We Learn from the Way Interviewers Answer Survey Questions?" *Field Methods* 29 (2): 140–153. https://doi.org/10.1177/1525822X16663162.

Young, Archie. 1997. "Ageing and Physiological Functions." *Philosophical Transactions of the Royal Society B: Biological Sciences* 352 (1363): 1837–1843. https://doi.org/10.1098/rstb.1997.0169.

4. Explaining Interviewer Effects in Wave Nonresponse and Income Item Nonresponse: Evidence of Common Causes

Abstract

Missing data in one wave of panel data collection may imply that measures of change cannot be accurately estimated over multiple waves. In surveys that use face-to-face interviewing in data collection, interviewers may cause or prevent missing data from occurring on a unit and item level. Such influence of the interviewer may have common causes. To investigate interviewer effects, data from the Survey of Health, Ageing and Retirement in Europe (SHARE) is used to examine the extent to which interviewer characteristics can explain the interviewer variance in wave nonresponse and income item nonresponse simultaneously by applying a two-level sequential logit model. In addition, this study examines the extent to which the interviewer effects on wave nonresponse and income item nonresponse have common causes. The results showed the presence of interviewer effects, and that respondent and interviewer characteristics can explain 47.8 percent of the interviewer variance. Moreover, there were common causes of interviewer effects on wave nonresponse and income item nonresponse. The interviewer characteristics of agreeing with the statement "if you catch people at the right time, most will participate" and having high expectations about the share of respondents who will report their income had a significant effect on wave response rates and income item response rates that is in the same direction for both rates. The former characteristic had a negative effect and the latter had a positive effect.

Key words: interviewer effects; unit nonresponse; item nonresponse; multilevel regression; sequential modeling; missing data

4.1 Introduction

Explaining interviewer effects in missing data is essential because by understanding how missing data and interviewers are linked, we can help in improving interviewer trainings and survey designs in order to minimize missing data. Many studies have shown, that interviewer trainings can help to prevent missing data or reduce variation induced by the interviewer (Billiet and Loosveldt 1988; Fowler and Mangione 1990; Groves and McGonagle 2001; Fowler 2004; Daikeler and Bosnjak 2020). Only if we understand the links between missing data and the interviewer, we can train interviewers in such a manner that helps in preventing missing data caused by the interviewer.

Particularly in panel surveys, that collect data from the same respondent over time, it is worthwhile preventing missing data. If observing changes from the same respondents over time is challenged by missing data, caused by nonresponse to the entire survey or nonresponse to survey items, we risk aggregating trends imprecisely or even introducing bias in measures of change (Lynn and Lugtig 2017). Having accurate measures of changes over time is important because aggregated trends are often used for decision-making. One of these aggregated trends that are commonly used and are commonly challenged by missing data are for example trends in income (Yan, Curtin, and Jans 2010). Therefore, in order to predict income trends accordingly, it is important to minimize missing data.

There is a large body of literature that shows that interviewers affect missing data (e.g., Pickery, Loosveldt, and Carton 2001; Schräpler 2006). Many studies have examined the impact of interviewer characteristics in cross-sectional surveys on missing data such as nonresponse (e.g., Durrant et al. 2010; Blom et al. 2011; Lipps and Pollien 2011 Jäckle et al. 2013; Durrant and D’Arrigo 2014). Fewer studies have investigated the impact of interviewers on wave nonresponse in panel surveys (O'Muircheartaigh and Campanelli 1999;

Haunberger 2010; Lynn, Kaminska, and Goldstein 2014;) and on income item nonresponse (Singer and Kohnke-Aguirre 1979; Pickery and Loosveldt 2001; Essig and Winter 2009) (for an overview, see West and Blom 2017).

What the existing studies have in common is that they have demonstrated that interviewers affected missing data to a notable extent. But a key flaw the most studies share is that the interviewer characteristics they have studied (i.e., gender, age, and interviewer experience) had little explanatory power. Researchers who were able to collect and analyze richer interviewer data have provided evidence that psychological interviewer variables, such as optimism, self-confidence, positive expectations, and positive attitudes can contribute to explaining interviewer effects in missing data (Singer and Kohnke-Aguirre 1979; Durrant et al. 2010; Jäckle et al. 2013; Vassallo et al. 2015). Moreover, the question of whether interviewer effects on different error sources of missing data have common causes remains unanswered. The question of whether the same interviewer characteristics have an impact on multiple errors is important because the absence of one error does not mean the absence of all others (Groves 2006).

To my knowledge, only few studies have focused on multiple survey errors and interviewer effects (e.g., Brunton-Smith, Sturgis, and Williams 2012; Vercruyssen, Wuyts, and Loosveldt 2017; Ackermann-Piek 2018). The study by Brunton-Smith, Sturgis, and Williams (2012) showed that interviewers with very high or very low cooperation rates contribute more to the measurement error variance in survey estimates than interviewers who achieved middle cooperation rates. However, Brunton-Smith and his colleagues (2012) could not explain these interviewer effects with interviewer characteristics, because no interviewer characteristics were available. Another study that could use rich interviewer data was conducted by Ackermann-Piek (2018, chapter 7). This study examined the relationship of interviewers between achieved cooperation rates and mean scores of survey items. The results from her study indicated that there is a weak negative association between cooperation rates and mean scores. However, the hypothesis that interviewer effects have common causes, which would mean that the same interviewer characteristics can cause both nonresponse error and measurement error, cannot be supported by Ackermann-Piek's (2018, chapter 7) study, as none of the interviewer characteristics she examined were able to explain interviewer effects on both, cooperation rates and mean scores. A study that found recurring effects of interviewer characteristics on unit nonresponse and item nonresponse was conducted by Vercruyssen, Wuyts, and Loosveldt (2017). They investigated interviewer effects in unit nonresponse and item nonresponse with interviewer sociodemographics, such as gender, age, and education, and interviewer experience. This study indicated that gender (mis)matching can affect the probability to respond to the entire survey and to survey items. However, much of the variance at the interviewer level remained unexplained and moreover, unit nonresponse and item nonresponse were treated independently from each other. Nonetheless, these three studies show that investigating multiple survey errors and their connection to interviewers may be a promising area of research.

Therefore, in order to understand the link between interviewers and missing data, I study the effect of various interviewer characteristics on wave response and income item response simultaneously using data from the Survey of Health, Ageing and Retirement (SHARE). SHARE is a panel survey that collects individual-level data, as well as rich interviewer data. This data offers us the opportunity to contribute to the literature on interviewer effects on multiple survey errors by answering two research questions:

1. To what extent can interviewer characteristics explain variation in wave nonresponse and income item nonresponse?
2. To what extent do interviewer effects on wave nonresponse and income item nonresponse have common causes?

4.2 Background

4.2.1 Missing data in panel surveys

Two types of nonresponse can lead to missing data in longitudinal surveys: unit nonresponse and item nonresponse (Binder 1998). Unit nonresponse describes the failure to obtain data from the entire interview (Groves and Couper 1998). In longitudinal surveys, we can further classify unit nonresponse into three types of categories: initial nonresponse, wave nonresponse, and attrition. Initial nonresponse means that the sampled subject could not be recruited into the panel survey. Wave nonresponse means that the sampled subject was recruited into the panel survey but this panel member missed a particular wave. Attrition means that the sampled subject was recruited into the panel survey but this panel member dropped out completely from the panel survey (Bethlehem, Cobben, and Schouten 2011).

Researchers have studied all types of nonresponse in longitudinal surveys extensively in order to prevent missing data in future surveys or later waves. They often observed a large decline in participation in the second wave, and after the second wave, participation remains at a rather stable level (e.g., Lepkowski and Couper 2002; Watson 2003; Wooden and Watson 2004; Bergmann et al. 2019;). With regards to second wave nonresponse, researchers found that nonresponse is determined by individual factors that are related to moving, such as age and community attachment (e.g., Lepkowski and Couper 2002), and by individual factors that are related to the likelihood to be at home, such as age, gender, marital status, employment status, household size, and household composition (e.g., Watson and Wooden 2009). In addition to these characteristics, nonresponse in the second wave has been found to be determined by individual factors that are related to the likelihood to refuse to participate such as race/ethnicity, educational level, income, urbanity of the region of residence, and social integration (Lepkowski and Couper 2002; Bianchi and Biffignandi 2019;) and by individual factors that are related to the likelihood of being incapable of participating, such as senility, language barriers and severe illness (Groves 2004).

In contrast to unit nonresponse, which is the failure to obtain data from the entire interview, item nonresponse is the failure to obtain a response to a particular question in the panel interview (Dillman et al. 2002). Item nonresponse is caused by respondents who did not provide an answer (i.e., respondents who responded with “don’t know”, refused to respond, or overlooked some questions), respondents who provided unusable answers (i.e., responses that lie outside the range of sensible answer options) and by respondents who provided a valid answer but the answer got lost (i.e., technical transfer errors) (de Leeuw, Hox, and Huisman 2003).

Researchers have found that in surveys, the amount of missing data generated by item nonresponse is typically large for sensitive and complex survey questions compared to factual survey questions (Tourangeau and Yan 2007). Income questions are one class of sensitive and complex survey questions (Tourangeau and Yan 2007). Previous studies have shown that the amount of missing data in income questions is typically large compared to other survey questions (e.g., Fisher et al. 2019), and that income item nonresponse tends to be concentrated at the tails of the income distribution (Bollinger et al. 2019). Income item nonresponse is influenced by respondent characteristics, such as age, gender and education (Yan, Curtin, and Jans 2010; Bollinger et al. 2019), the variability and complexity of income composition (Frick and Grabka 2010), and memory recall ability (Colsher and Wallace 1989).

If respondents take a cognitive shortcut because the question is too complex or difficult, or if respondents answer in a socially desirable way because the question is too sensitive, they may give a "don't know" response instead of providing a substantive value to the question (Krosnick 2002; Krosnick et al. 2002). Therefore, item nonresponse can also be viewed as a measurement error (Tourangeau, Rips, and Rasinksi 2000). A measurement error is the difference between the response collected and its true value. Measurement errors are typically calculated based on respondents' substantive responses (Biemer et al. 2011). Because a "don't know" response can misreport the respondent's "true" value, item nonresponse is a source of measurement error that generates missing data.

According to the *response continuum model* by Yan and Curtin (2010), unit nonresponse and item nonresponse can be linked to each other. Respondents with high probabilities to respond to the entire survey have also high probabilities to answer survey items, whereas respondents with low probabilities to participate in the survey have low probabilities to respond to the survey items. Therefore, considering this direction of the interconnection between unit nonresponse and item nonresponse is useful when studying multiple types of nonresponse.

4.2.2 Link between missing data and interviewers

Interviewers play an important role in preventing the occurrence of missing data in face-to-face surveys, because they interact with the respondent, thereby influencing how the respondent behaves (Groves and Couper 1998). They are responsible for establishing contact the respondent, obtaining the respondent's cooperation in the survey, and conducting the interviews (Mneimneh et al. 2019). Some interviewers are more successful in accomplishing these tasks than other interviewers. Additionally, an interviewer may be more successful in accomplishing certain tasks than other tasks (e.g., Edwards, Sun, and Hubbard 2020; Ongena and Haan 2020). As a result, different amounts of missing data may be generated across respondents who are interviewed by different interviewers.

When data collected by the same interviewer are more similar than data collected by different interviewers, this is known as interviewer effects (Kish 1962). The causes of interviewer effects have been described in the literature (Groves and Couper 1998). They arise from differences in interviewers' sociodemographic characteristics, experience, personality, skills, expectations, behavior, and other observable attributes of the interviewers. Based on the literature, it seems that psychological characteristics may drive observed interviewer effects rather than the conventionally examined interviewer characteristics (i.e., gender, age, and interviewer experience). These differences across interviewers lead to differences in the way the interviewers interact with respondents, and therefore, to different survey outcomes across interviewers (Groves and Couper 1998), such as missing data on unit or item level.

4.3 Data and Methods

4.3.1 Data

I used data from the Survey of Health, Ageing and Retirement in Europe (SHARE) (Börsch-Supan 2017; Börsch-Supan et al. 2013)[5]. SHARE is a panel survey on ageing that biennially collects micro-data on health, socioeconomic status, and social networks from individuals aged 50 or older in Europe. Interviewers conduct face-to-face interviews using a laptop with an installed computer-assisted personal interviewing (CAPI) software.

SHARE also contains rich information about the interviewers who approached and collected data from the panel members. The interviewers were asked by the national survey coordinators to participate in a voluntary online survey prior to their regular fieldwork for SHARE. This online survey collected data on the interviewers' socioeconomic status, interviewer experience and personality. Moreover, these data provide information on interviewers about their attitudes towards reasons for being an interviewer and towards gaining cooperation, general trust in people and data privacy concerns. This online survey collected also data on self-reported behavior, such as behavior during the interview or using online banking and using social networks, and survey expectations. The respondent data is linked to interviewer data via a SHARE interviewer ID, which is first requested in the SHARE interviewer survey, then recorded in the sample management system, followed by the regular SHARE interview. I analyze data where the SHARE interviewer ID at the time

[5] This paper uses data from SHARE Waves 5 and 6 (DOIs: 10.6103/SHARE.w5.700, 10.6103/SHARE.w6.700, see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982) and Horizon 2020 (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C). SHARE gratefully acknowledges other various national funding sources (see www.share-project.org).

of data preparation was present in these three sources. This resulted in data for four countries (Sweden, Slovenia, Italy, and Germany). These countries have efficiently conducted the SHARE interviewer survey and they represent all European areas (North, East, South, and West).

I investigated wave nonresponse and item nonresponse in the second wave of SHARE using data from all respondents who were interviewed in the first wave and were eligible and approached for an interview in the second wave. I exclude those who died between the fieldwork of the first wave and the start of fieldwork for the second wave. I obtained covariates that I will use in understanding the correlates of missing data from the first wave. Interviewers are included in my analyses if they have interviewed at least one panel respondent in the second wave. This selection process resulted in a dataset of 9,008 panel respondents who were approached by 306 interviewers in the four countries. In my analyses, I used observations that have complete information on the respondent control variables and on the interviewer explanatory variables. Where item nonresponse rates of control variables or explanatory variables were higher than five percent, I created a separate variable category, labelled "missing value", instead of dropping observations. In my analyses, I used observations from 6,347 panel respondents who were approached by 266 interviewers for the second wave. Of these panel respondents, 4,917 were interviewed in the second wave by 266 interviewers.

4.3.2 Variables of interest

The variables of interest were second-wave nonresponse and income item nonresponse in the second wave. As previous studies have shown, both phenomena generate large amounts of missing data in surveys (Watson 2003; Yan, Curtin, and Jans 2010) and are prone to interviewer effects in interviewer-administered surveys (Pickery, Loosveldt, and Carton 2001; Schräpler 2006).

I defined wave nonresponse as a binary variable with the following values:

$$Y_{1ij}=\begin{cases} 0\ nonresponse\ in\ the\ second\ wave \\ 1\ participation\ in\ the\ second\ wave. \end{cases} \tag{1}$$

The outcome Y_{1ij} takes the value 0 if an eligible panel member *i*, was not interviewed by interviewer *j* and, therefore, did not participate to the second wave of the survey. The outcome Y_{1ij} equals 1 if an eligible panel member *i* was interviewed by interviewer *j* and, therefore, participated in the second wave.

For item nonresponse, I defined a binary outcome variable as follows:

$$Y_{2ij}=\begin{cases} 0\ item\ nonresponse\ to\ income\ question \\ 1\ substantive\ response\ to\ income\ question. \end{cases} \tag{2}$$

The outcome Y_{2ij} equals 0 if a panel respondent *i* was interviewed by interviewer *j* in the second wave but did not provide a substantive value for income. The outcome Y_{2ij} takes the value 1 if a panel respondent *i* was interviewed by interviewer *j* in the second wave and did provide a substantive income value.

The SHARE questionnaire asked respondents a series of detailed questions on income that distinguish between the different sources of income. The first question asked respondents whether they have a source of income. If respondents responded “yes”, they were immediately asked to enter the exact amount of this income. If a respondent responded “don’t know” or refused to answer the question, this triggered an unfolding sequence of bracket questions. Since the attempt at getting a precise value for income failed, the aim of the unfolding brackets was to obtain the second preferred type of answer, namely a range in which the respondent’s income is located (SHARE 2017). If respondents did not answer the unfolding bracket sequence, item nonresponse was recorded for the income question and was defined as there being no substantive answer regarding employment earnings, earnings from self-employment, public pension payments, and/or other regular payments (Appendix, table A4.1). Respondents without any sources of income (mainly housewives and househusbands) were filtered out of all income questions. In my analyses, a value of 0 Euro on the income variable was not defined as income item nonresponse.

4.3.3 Analytical strategy and empirical model

Respondents' decision to answer the income question was not a single choice but rather consists of a sequence of choices. First, the sampled subjects decided whether to respond to the survey. Once they have decided to respond to the survey, the respondents had to decide whether to answer the income question. Therefore, I defined wave nonresponse as the lack of wave response and income item nonresponse as the failure to respond to the income question, and modelled them sequentially. This allowed me to analyze them simultaneously in a sequence (Figure 4.1).

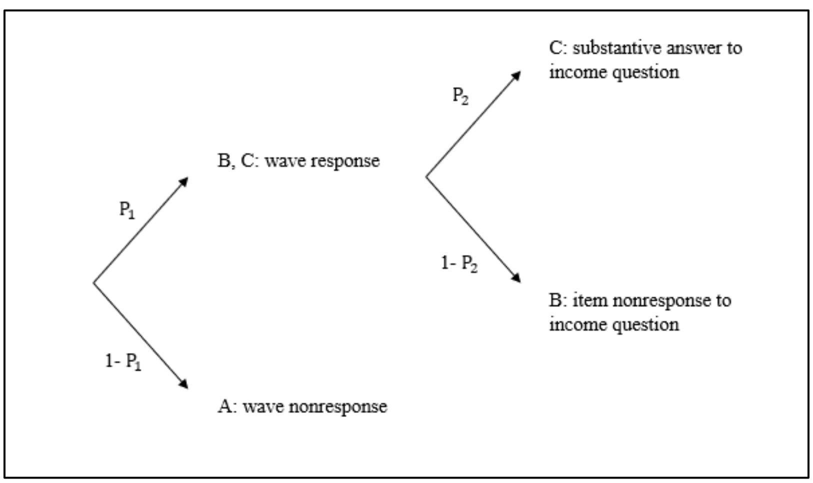

Figure 4.1. Simplified model of survey response process.

The survey response process in my model consisted of two transitions (k=2) with three states: A, B, and C. The conditional probability that a panel member passes transition

k is given by P (see fig. 4.1). In our study, these transitions were second-wave response and a substantive response to the income question in the second wave. The sequential approach allows the examination of the relationship between explanatory and the probabilities of passing each transition point. Moreover, it allows an integrated examination of the relationship between the explanatory variables and the probabilities of passing both transitions simultaneously (Buis 2017).

Based on the procedure outlined by Buis (2017), I formulated the estimation of the response process as follows:

$$E(L_{ij})=(1-p_{1ij})\, l_A + p_{1ij}\,(1-p_{2ij})\, l_B + p_{1ij}\, p_{2ij}\, l_C, \tag{3}$$

where L_{ij} is the final outcome, i.e. an interview with a substantive response to the income question of respondent i, who is interviewed by interviewer j, p_{kij} is the probability of respondent i, interviewed by interviewer j, to pass transition k, where explanatory and control variables are accounted for in p_{kij}, ls are the outcome of each potential states (0 or 1), with state A being wave nonresponse, B income item nonresponse, and C substantive response to the income question (see fig. 4.1). Thus, the final outcome is weighted by the probabilities of attaining a transition k. Equation (3) shows the nonlinear relationship between explanatory variables, as well as control variables, and the final outcome L_{ij}.

Based on the literature (West and Blom 2017), I included from different dimensions a large set of explanatory variables for interviewers. These six dimensions were: sociodemographic characteristics, experience, personality, attitudes, self-reported behavior, and survey expectations (Appendix, table A4.2).

The variables of the sociodemographic dimension were gender, age, and education. The experience variables were: general interviewer experience, prior experience in SHARE, and working hours. The variables of the personality dimension were the personality traits

(Big Five according to Rammstedt and John (2007)). The variables of the attitude dimension were: attitudes towards the relative importance of reasons for being an interviewer (payment, interesting work, opportunity to interact, gaining insights in other's life, involvement in scientific research, and flexible working hours), data privacy concerns, trust in people, opinion on respondents' reason to participate (contributing to scientific research, serving society, interacting with someone, inability to say no, and receiving incentives), and attitudes towards gaining respondents' cooperation (based on de Leeuw et al. (1998)). The variables of the self-reported behavior dimension were: interviewer behavior if difficulties from the respondent side arise during the interview (explaining the question if respondent has trouble to understand the question, rereading exactly the question if respondent has trouble with the question, shortening the question if respondent has trouble to listen the interviewer, speaking slowly if respondent has trouble to understand the question, speaking faster if respondent is in hurry, completing answers by interviewer if answers are known, and sticking to interviewer instructions, even if interviewer do not consider them as sensible), and other self-reported behavior such as using online banking, using social networks, and own survey participation, own item nonresponse to income question. The variable of the survey expectation dimension were survey expectations about the percentage of respondents who will report income. The operationalization and distribution of interviewer variables are described in the Appendix, tables A4.2–A4.4).

When analyzing the sequential process, I considered several control variables for respondent characteristics based on the literature on respondent characteristics that correlate with wave nonresponse (e.g., Lepkowski and Couper 2002; Groves 2004; Watson and Wooden 2009; Bianchi and Biffignandi 2019). I included gender, age, education, household size, marital status, employment status, number of children, self-assessment of health, limitations of daily activities (limitation of instrumental activities of daily living (IADL) scale according to Lawton and Brody (1969)), smoking habits, number of chronic diseases, score of recall memory test, number of depression symptoms (EURO-D scale according to Prince et al. in 1999), giving any kind of help to others, getting any kind of help from others, maximum score of grip strength test, type of area and country of residence.

To investigate the common causes of interviewer effects on wave responses to the income question in the second wave, I used a two-level sequential logistic regression model. This modelling choice allows me to consider the potential correlation of interviewer variation with wave response rates and interviewer variation in income item response rates. I started the analyses with a model that included only the final outcome and the *k* transition, which I used as the null model (M_0). The null model is useful as a benchmark to which other models can be compared (Hox, Moerbeek, and van de Schoot 2010). I then added all covariates at the respondent level to the model (M_1). This model is able to explain variance at the respondent level. In addition, it may affect the variance at the interviewer level. This is because if the distribution of respondents is not the same for all interviewers (i.e., interviewer assignments differ in their composition), this variation can explain some of the interviewer-level variance in the average final outcome between interviewers (Hox, Moerbeek, and van de Schoot 2010). This model also included three dummy variables for the four countries (DE, IT, SE, and SI). These dummies were fitted as fixed covariates in order to reduce the variance of the residual error terms and to control for any unexpected differences between the countries. Finally, I added all the explanatory variables at the interviewer level to the model (M_2). These interviewer characteristics are gender, age, education, interviewer experience, personality, attitudes, and behavior (Appendix, tables A4.2–A4.4).

Based on the approach of Hox and his colleagues (2010), I also calculated corrected estimates for models with control variables or explanatory variables. This correction is necessary because variance components cannot be compared across models with and without independent variables because the level-one variance in logistic regression models is fixed to $\frac{\pi^2}{3}$.

4.4 Results

The analyses showed that 23 percent of the first wave panel members did not participate in the second wave and five percent of the second wave panel respondents did not give their income. The mean of the wave response rates by interviewer was 79.4% and the median by interviewer was 82.3%. The 10th and 90th percentiles were 58.6% and 95%, respectively. The mean of the income response rates by interviewer was 94.3% and the median by interviewer was 97.6%. The 10th and 90th percentiles were 83.3% and 100%, respectively.

To evaluate the extent to which our measured interviewer characteristics can explain second wave cooperation and income item response, we inspected, in a multilevel setting, the proportion of the variance that was attributable to the interviewer compared to the total variance. The null model (M_0) showed a systematic difference between interviewers in the final outcome (see table 4.1).

Table 4.1. Estimates of variances in two-level logistic regression models on second wave response and income item response

Variance component	M_0	M_1	M_2
$\sigma^2_{Respondent}$	3.29	2.45	2.45
$\sigma^2_{Interviewer}$	.46	.50	.24

Note: For full model details see Appendix, table A4.5. n(respondent)=11,264 and n(interviewer)=266. No. of respondents: sum of observations for wave response (6,347) and item response (4,917). M0: Null model. M1: Model with control variabels at respondent level. M_2: Model with control variables at respondent level and explanatory variabels at interviewer level. According to Hox, Moerbeek, and van de Schoot (2010), M_1 and M_2 were estimated with a scale correction factor of 0.86.

Thus, 12.3% of the total variance of wave response with substantive response to the income question was at the interviewer level. Adding variables at the respondent level (M_1) increased the interviewer variance from .46 to .50 (table 4.1). When we added interviewer variables (M_2), the variance component at the interviewer level was reduced. The last model (M_2) showed an interviewer variance component of .24. Therefore, 47.8% of the variance at the interviewer level was explained by the control and explanatory variables, as (.46 – .24) / .46 equals .478.

In the analyses of the common causes of interviewer effects on different types of nonresponse, I found that several interviewer characteristics significantly correlated with wave response and income item response at a significance level of 95% or higher in M_2. These were interviewer gender, age, education, experience, attitudes, behavior and expectations (see table 4.2). The results showed that out of 45 interviewer characteristics, 20 significantly correlated with wave response or item response at a significance level of 95% or higher. These significant correlations occurred more frequently with item response than with wave response.

Table 4.2. Significant interviewer effects in two-level logistic regression model (M_2) on wave and income item response

	Wave response		*Income item response*	
	Log-odds	*s.e.*	*Log-odds*	*s.e.*
Interviewer characteristics				
Gender: male (ref.)				
- female	–.01	.11	–.48*	.20
Age in years	–.00	.01	–.04***	.01
Education level: low (ref.)				
- medium	.03	.19	–1.18**	.37
- high	–.22	.18	–.73*	.37
Experience				
Prior worked for SHARE: no (ref.)				
- yes	–.01	.14	–.75**	.24
Working h*ours in hrs./week*: 0–10 (ref.)				
- 11–20 hrs.	.03	.14	.02	.25
- 21–30 hrs	.61***	.17	.50	.32
- 31+ hrs	.06	.16	–.17	.27
- missing value	.64	.47	.97	.88

Table 4.2. (cont.)

	Wave response		Income item response	
	Log-odds	*s.e.*	*Log-odds*	*s.e.*
Attitude scores towards importance of reasons for being an interviewer				
Flexible working hrs.: not that important (ref.)	.13**	.04	–.11	.09
- very important				
Data privacy concerns: no/few (ref.)				
- some/many	.25	.13	.30	.25
- missing value	–1.40	.88	–3.64**	1.24
Opinion scores on respondents' reason to participate				
- contributing to scientific research	.09	.07	.53***	.12
- interacting with someone	–.08	.05	–.26**	.10
- expressing opinion	.05	.05	.25*	.10
- inability to say no	.04	.04	.15*	.07
Attitudes towards gaining cooperation				
Always persuade reluctant respondents: disagree (ref.)				
- agree	–.08	.12	.71**	.23
Caught at right time, most people participate: disagree (ref.)				
- agree	–.35*	.16	–.85**	.31
Self-reported interviewing behavior				
Explains question: not really/at all (ref.)				
- somewhat/perfectly	.21	.12	–.95***	.21
Speaks faster: not really/at all (ref.)				
- somewhat/perfectly	–.00	.11	–.42*	.19
Completes answers: not really/at all (ref.)				
- somewhat/perfectly	–.57	.33	–1.23*	.52
Using online banking: no (ref.)				
- yes	.29*	.13	–.45*	.23
Using social networks: no (ref.)				
- yes	–.28*	.13	.18	.23
Income reported: no (ref.)				
- yes	.13	.12	.75***	.21
Expectations towards percentage reporting income: 0–50 % (ref.)				
- 51–75%	.25	.17	.70*	.27
- 76–90%	.34*	.17	.96***	.26
- 91–100%	.16	.21	.73*	.36
- missing value	.21	.25	–.11	.41

Note: Table shows only significant effects for wave response and/or item response at a significance level of .05 or lower. For full model details see Appendix, table A4.5. M_2= model with control variables at respondent level and explanatory variabels at interviewer level. n(respondent)=11,264 and n(interviewer)=266. No. of respondents: sum of observations for wave response (6,347) and item response (4,917). According to Hox, Moerbeek, and van de Schoot (2010), M_2 was estimated with a scale correction factor of 0.86. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To find out whether there were common causes for interviewer effects on wave response and on income item response in the second wave, we inspected these 20 statistically significant observed correlations in more detail. Only three of these 20 significant correlations are significant for both transitions (wave response and item response). These are the use of online banking, attitude towards gaining cooperation (agreeing with the statement "if you catch people at the right time most will participate"), and having high expectations about the share of respondents who will report their income. Two of these three interviewer characteristics correlations in wave response and in item response were in the same direction for both types of nonresponse. These were the attitude towards gaining cooperation and the high expectations about income reporting. In other words, the respondent's probability to cooperate and provide a substantive answer to the income questions decreased with interviewers who agreed that just catching the respondent at the right time determined the choice of participation. Moreover, if interviewers expected more than 75 percent but less than 90 percent of their respondents to report their income, they obtained significantly higher wave response rates and income item response rates than those interviewers who expected less than 50 percent of their respondents to report their income.

To inspect the effect size of these last findings further, we predicted the respondents' probabilities for each transition by these two interviewer characteristics (see table 4.3). The respondents' predicted probabilities varied with the interviewer's attitude towards the statement "if you catch people at the right time most will participate" by five percentage points for wave response and by three percentage points for the response to the income question. The average rates of both transitions were lower if all respondents would have had interviewers who agreed that "if you catch people at the right time most will participate". The predicted probabilities with respect to interviewer expectations showed that if all respondents would have had interviewers that expected more than 75% but less than 90% of their respondents to report their income, average rates of both transitions would have been five percentage points higher than compared with the case if all respondents would have had interviewers that expected a maximum of 50% of their respondents to report their income.

Table 4.3. Predicted probabilities for each transition by interviewer attitude towards gaining cooperation and interviewer expectations towards reporting income

	Wave response		Income item response	
	Mean	*s.e.*	Mean	*s.e.*
Caught at right time, most people participate:				
disagree	.83***	.02	.97***	.01
agree	.78***	.01	.94***	.01
Percentage of respondents expected to report income:				
50% or fewer	.75***	.03	.91***	.02
51%–75%	.79***	.02	.95***	.01
76%–90%	.80***	.02	.96***	.01
91%–100%	.77***	.02	.95***	.01
Missing value	.78***	.04	.90***	.04

Note. n(respondent)=11,264 and n(interviewer)=266. *** $p < 0.001$.

4.5 Conclusion and Discussion

This study showed that once interviewers gained cooperation from the panel members, they were very likely to collect substantive income data from their respondents. The average wave response rates by interviewer was 79.4% and the average income item response rate by interviewer was 94.3%. These rates are in line with previous findings of second wave response rates (e.g., Watson 2003; Buck et al. 2006; Lipps 2007) and income item response rates in European panel surveys (e.g., Riphahn and Serfling 2005; Schräpler 2006). These findings suggest that it may be worthwhile to invest into recruiting panel members for the second wave interview in order to reduce the risk of missing data due to wave nonresponse and item nonresponse because once respondents decided to participate, they are likely to report their income. The opposite is also true, as Yan, Curtin and Jans (2010, 155) have shown that “those who were hard to recruit are more likely to refuse individual items once recruited into the sample.” Therefore, we suggest that researchers conduct qualitative and quantitative research that examines the relationship between respondents who are hard to recruit and those who do not answer survey questions in order to develop measures that can decrease the likelihood of missing data in wave response and income item response.

After modelling wave response and income item response sequentially, we analyzed the variance of wave response and income item response in the second wave simultaneously. In the two-level sequential logistics regression models, the null model showed that the proportion of the interviewer-level variance compared to the total variance was 0.123. Thus, 12.3% of the variance of wave response with substantive answer to the income question was attributable to the interviewer. Moreover, in the model with respondent control variables (M_1), the variance component at the interviewer-level was larger than in the null model (M_0). I assume that the variance at the interviewer-level in M_1 increased because, contrary to M_1, M_0 did not consider the clustering effect of respondent characteristics. Since interviewers in SHARE were not allocated randomly to the respondent, this study shows that interviewer effects may be slightly larger if the allocation of respondents was taken into consideration.

To answer the first research question “To what extent can interviewer characteristics explain variation in wave nonresponse and income item nonresponse?”, I inspected the variance components and the variables used to explain the interviewer variance. The variation between interviewers in the final model (M_2), which included explanatory variables at the interviewer-level, was reduced by the statistically significant correlations. In this study, 48% of the variance at the interviewer level using explanatory variables of the interviewers and control variables of the respondents was explained. Despite using rich data on interviewers, this study was only able to partially explain interviewer variance using measures of interviewer characteristics. I advise future researchers to study other potentially causes of interviewer effects, as a large amount of variance at the interviewer level remained unexplained. To shed more light into this black box of interviewer effects on multiple error sources of missing data, we propose conducting qualitative studies that investigate the causes of these interviewer effects and subsequently use this information to collect corresponding quantitative interviewer data. Future researchers could use these newly collected quantitative data to examine whether the new measurements can explain interviewer effects on nonresponse and measurement error. Only on the basis of interviewer data that has an explanatory power regarding interviewer tasks and survey outcomes can we train interviewers in such a manner that helps reduce missing data caused by nonresponse error and measurement error.

To answer the second research question “To what extent do interviewer effects on wave nonresponse and income item nonresponse have common causes?” I inspected the statistically significant correlates of both types of nonresponse. The results showed that there is a tendency for common causes of interviewer effects on wave response and item response. However, the reduction of interviewer variance with the examined interviewer characteristics was caused by significant correlations with income item response rather than with wave response. The findings related to the variable capturing interviewers’ opinion on the respondents’ reasons to participate are particularly interesting. The opinion scores significantly correlated with item response but not with wave response, even though the effect was in the same direction. For instance, interviewers who scored high on the opinion scale that respondents participated in order to contribute to scientific research obtained higher (but statistically insignificant) cooperation rates but achieved higher (and statistically significant) substantive response rates to the income questions. This finding is in accordance with a study by Ackermann-Piek (2018, chapter 7), which showed that interviewer characteristics may affect multiple survey errors differently. These findings suggest that the interviewers’ attitude towards the reasons why people participate in surveys may be, in survey practice, related to the success of the interviewer in obtaining substantive response in single survey questions, but not necessarily to the success in obtaining respondent cooperation in general.

Moreover, two out of the 45 measured interviewer characteristics significantly correlated with both transitions in the same direction for both types of nonresponse. These were interviewer expectations towards income reporting and interviewer attitude towards gaining cooperation ("if you catch people at the right time most will participate"). On the one hand, positive interviewer expectations towards income reporting positively correlated with wave response and income item response. On the other hand, the interviewer attitude towards gaining cooperation ("if you catch people at the right time most will participate") negatively correlated with wave response and income item response. Thus, interviewer survey expectations in our study and interviewers' attitude towards gaining cooperation turned out to be promising in explaining interviewer effects for both transitions but although this study provided evidence of common causes for interviewer effects, our results cannot tell us how the interviewer opinions, attitudes, and expectations that explained these effects, were formed, i.e., the causal mechanisms at work when interviewers approach respondents and interview them. Yet, the mechanisms behind these correlations are unknown.

I speculate that there are two possible explanations. First, these associations might arise from an (non) optimistic self-assessment. This means that interviewers who achieved higher wave response rates and item nonresponse rates due to high self-confidence levels may achieve high cooperation rates and high income item nonresponse rates. This idea is supported by an early qualitative study by Snijkers, Hox, and de Leeuw (1999). The authors concluded that projecting an image of self-confidence and trust can help the interviewers fight nonresponse. If this is true, survey researchers should invest in interviewer trainings boosting the interviewers' self-confidence.

In addition to the possibility that self-confidence might be the mechanism behind these results, the second possibility is that interviewer expectations or beliefs reflect self-perception due to prior knowledge. In other words, these expectations and beliefs might have been good predictors of nonresponse because interviewers were able to correctly anticipate the panel respondents' willingness to cooperate or respond due to some prior knowledge of the respondents they were likely to be assigned to interview. If that is the case, then trainings that focus on increasing self-confidence may be less helpful to reduce missing data and we would need to design trainings that change interviewers' self-perception in terms of their capability to secure response, regardless of the respondent's characteristics and circumstances. Therefore, in order to adapt interviewer trainings accordingly, future research needs to disentangle the mechanism that are at work.

To conclude, my research has provided insights into the common causes and explanatory power of interviewer effects on wave response and income item response in the second wave. I encourage researchers to investigate interviewer effects in multiple survey errors, because, first, research in this area is scarce and, second, our results cannot be generalized to other studies, since interviewer effects can be study-specific (e.g., see study by Ackermann-Piek and her colleagues in 2020).

References

Ackermann-Piek, Daniela. 2018. "Interviewer effects in PIAAC Germany 2012." PhD diss., 119–136, Mannheim, Germany: University of Mannheim. Available at: https://ub-madoc.bib.uni-mannheim.de/42772 (accessed August 2020).

Ackermann-Piek, Daniela, Julie M. Korbmacher, and Ulrich Krieger. 2020. "Explaining Interviewer Effects on Survey Unit Nonresponse: A Cross-survey Analysis." In *Interviewer Effects from a Total Survey Error Perspective*, edited by Kristen Olson, Jolene D. Smyth, Jennifer Dykema, Allyson Holbrook, Frauke Kreuter, and Brady T. West, 192–203. New York: Chapman and Hall/CRC.

Bergmann, Michael, Thorsten Kneip, Giuseppe De Luca, and Annette Scherpenzeel. 2019. *Survey Participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1–7*. SHARE Working Paper Series 41-2019. Munich, Germany: Munich Center for the Economics of Aging (MEA), Max Planck Institute for Social Law and Social Policy. Available at: http://www.share-project.org/uploads/ tx_sharepublications/WP_Series_41_2019_Bergmann_et_al.pdf (accessed August 2020).

Bethlehem, Jelke, Fannie Cobben, and Barry Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. New York: John Wiley.

Bianchi, Annamaria, and Silvia Biffignandi. 2019. "Social Indicators to Explain Response in Longitudinal Studies." *Social Indicators Research* 141 (3): 931–957. https://doi.org/10.1007/s11205-018-1874-7.

Biemer, Paul P., Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. 2011. *Measurement Errors in Surveys*. New York: John Wiley.

Billiet, Jacques, and Geert Loosveldt. 1988. "Improvement of The Quality of Responses to Factual Survey Questions by Interviewer Training." *Public Opinion Quarterly* 52 (2): 190–211. https://doi.org/10.1086/269094.

Binder, David A. 1998. "Longitudinal Surveys: Why are these Surveys Different from All Other Surveys?" *Survey Methodology* 24:101–108.

Blom, Annelies G., Edith D. de Leeuw, and Joop J. Hox. 2011. "Interviewer Effects on Nonresponse in the European Social Survey." *Journal of Official Statistics* 27 (2): 359–377.

Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. 2019. "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch."*Journal of Political Economy* 127 (5): 2143–2185. https://doi.org/10.1086/701807.

Börsch-Supan, Axel. 2017. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5 and Wave 6. Release version: 6.0.0.* SHARE-ERIC. https://doi.org/10.6103/SHARE.w5.600 and https://doi.org/0.6103/SHARE.w6.600.

Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie M. Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology* 42 (4): 992–1001. https://doi.org/10.1093/ije/dyt088.

Buck, Nicholas, Jonathan Burton, Heather Laurie, Peter Lynn, and S.C. Noah Uhrig. 2006. *Quality Profile: British Household Panel. Survey Version 2.0: Waves 1 to 13: 1991–2003*. Colchester, UK: University of Essex, Institute for Social and Economic Research.

Brunton-Smith, Ian, Patrick Sturgis, and Joel Williams. 2012. "Is Success in Obtaining Contact and Cooperation Correlated With the Magnitude of Interviewer Variance?" *Public Opinion Quarterly* 76 (2): 265–286. https://doi.org/10.2307/41684564.

Buis, Maarten L. 2017. "Not All Transitions Are Equal: The Relationship between Effects on Passing Steps in a Sequential Process and Effects on the Final Outcome." *Sociological Methods & Research* 46 (3): 649–680. https://doi.org/10.1177/0049124115591014.

Colsher, Patricia L., and Robert B. Wallace. 1989. "Data Quality and Age: Health and Psychobehavioral Correlates of Item Nonresponse and Inconsistent Responses." *Journal of Gerontology* 44 (2): 45–52. https://doi.org/10.1093/geronj/44.2.P45.

Daikeler, Jessica, and Michael Bosnjak. 2020. "How to Conduct Effective Interviewer Training: A Meta-Analysis and Systematic Review." In *Interviewer Effects from a Total Survey Error Perspective*, edited by Kristen Olson, Jolene D. Smyth, Jennifer Dykema, Allyson Holbrook, Frauke Kreuter, and Brady T. West, 46–56. New York: Chapman and Hall/CRC.

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

de Leeuw, Edith D., Joop J. Hox, Ger Snijkers, and Wim de Heer. 1998. "Interviewer Opinions, Attitudes and Strategies Regarding Survey Participation and Their Effect on Response." *Proceedings of the Eighth International Workshop on Household Survey Nonresponse*, 239–248. Mannheim, Germany: Zentrum für Umfragen, Methoden und Analysen (ZUMA). Available at: https://www.ssoar.info/ssoar/bitstream/ handle/document/49722/ssoar-1998-leeuw_et_al-Interviewer_opinions_attitudes_ and_strategies.pdf?sequence=1&isAllowed=y&lnkname=ssoar-1998-leeuw_et_al-Interviewer_opinions_attitudes_and_strategies.pdf (accessed August 2020).

Dillman, Don A., John L. Eltinge, Robert M. Groves, and Roderick J. A. Little. 2002. "Survey Nonresponse in Design, Data Collection, and Analysis." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 3–26. New York: John Wiley.

Durrant, Gabriele B., and Julia D'Arrigo. 2014. "Doorstep Interactions and Interviewer Effects on the Process Leading to Cooperation or Refusal." *Sociological Methods & Research* 43:490–518. https://doi.org/10.1177/0049124114521148.

Durrant, Gabriele B., Robert M. Groves, Laura Staetsky, and Fiona Steele. 2010. "Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys." *Public Opinion Quarterly* 74: 1–36. https://doi.org/10.1093/poq/nfp098.

Edwards, Brad, Hanyu Sun, and Ryan Hubbard. 2020. "Behavior Change Techniques for Reducing Interviewer Contributions to Total Survey Error." In *Interviewer Effects from a Total Survey Error Perspective*, edited by Kristen Olson, Jolene D. Smyth, Jennifer Dykema, Allyson L. Holbrook, Frauke Kreuter, and Brady T. West, 76–87. New York: Chapman and Hall/CRC.

Essig, Lothar, and Joachim K. Winter. 2009. "Item Non-Response to Financial Questions in Household Surveys: An Experimental Study of Interviewer and Mode Effects." *Fiscal Studies* 30 (3–4): 367–390 https://doi.org/10.1111/j.1475-5890.2009.00100.x.

Fisher, Paul, Laura Fumagalli, Nick Buck, and Silvia Avram. 2019. *Understanding Society and its income data*. Understanding Society Working Paper Series No. 2019–08. Colchester, UK: University of Essex. Available at: https://www.understandingsociety.ac.uk/ sites/default/files/downloads/working-papers/2019-08.pdf (accessed August 2020).

Fowler, Floyd J. Jr. 2004. "Reducing Interviewer-Related Error Through Interviewer Training, Supervision, and Other Means." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymor Sudman, 259–278. New York: John Wiley.

Fowler, Floyd J. Jr., and Thomas W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Thousand Oaks, CA: SAGE.

Frick, Joachim R., and Markus Grabka. 2010. "Item Nonresponse and Imputation of Annual Labor Income in Panel Surveys from a Cross-national Perspective." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith, 355–374. New York: Wiley.

Groves, Robert M. 2004. *Survey Errors and Survey Costs*. New York: John Wiley.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70 (5): 646–675. https://doi.org/10.1093/poq/nfl033.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley.

Groves, Robert M., and Katherine A. McGonagle. 2001. "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation." *Journal of Official Statistics* 17 (2): 249–265.

Haunberger, Sigrid 2010. "The Effects of Interviewer, Respondent and Area Characteristics on cooperation in Panel Surveys: A Multilevel Approach." *Quality & Quantity* 44 (5): 957–969. https://doi.org/10.1007/s11135-009-9248-5.

Hox, Joop J., Mirjam Moerbeek, and Rens van de Schoot. 2010. *Multilevel Analysis: Techniques and Applications*. New York: Routledge.

Jäckle, Annette, Peter Lynn, Jennifer Sinibaldi, and Sarah Tipping. 2013. "The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-Operation with Face-to-Face Surveys." *Survey Research Methods* 7 (1): 1–15. https://doi.org/10.18148/srm/ 2013.v7i1.4736.

Kish, Leslie 1962. "Studies of Interviewer Variance for Attitudinal Variables." *Journal of the American Statistical Association* 57 (297): 92–115.

Krosnick, Jon A. 2002. "The Causes of No-Opinion Responses to Attitude Measures in Surveys: They are Rarely What They Appear to Be." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 87–100. New York: John Wiley.

Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert C. Mitchell, et. al. 2002. "The Impact of "No Opinion" Response Options on Data Quality: Non-attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66 (3): 371–403. https://doi.org/10.1086/341394.

Lawton, M. Powell, and Elaine M. Brody. 1969. "Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living." *The Gerontologist* 9 (3): 179–86. https://doi.org/10.1093/geront/9.3_Part_1.179.

Lepkowski, James M., and Mick P. Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 259–272. New York: John Wiley.

Lipps, Oliver. 2007. "Attrition in the Swiss Household Panel." *methods, data, analyses* 1 (1): 45–68. https://doi.org/10.12758/mda.2007.003.

Lipps, Oliver, and Alexandre Pollien. 2011. "Effects of Interviewer Experience on Components of Nonresponse in the European Social Survey." *Field Methods* 23 (2): 156–172. https://doi.org/10.1177/1525822X10387770.

Lynn, Peter, Olena Kaminska, and Harvey Goldstein. 2014. "Panel Attrition: How Important is Interviewer Continuity?" *Journal of Official Statistics* 30 (3): 443–457. https://doi.org/10.2478/jos-2014-0028.

Lynn, Peter, and Peter Lugtig. 2017. "Total Survey Error for Longitudinal Surveys." In *Total Survey Error in Practice*, edited by Paul P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West, 279–298. New York: Wiley.

Mneimneh, Zeina, Lars E. Lyberg, Sharan Sharma, Mahesh Vyas, Dhananjay B. Sathe, Frederic Malter, and Yasmin Altwaijri. 2019. "Case Studies on Monitoring Interviewer Behaviour in International and Mutinational Surveys." In *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A. L. Stoop, and Brita Dorer, 731–770, Hoboken, NJ: John Wiley.

O'Muircheartaigh, Colm, and Pamela Campanelli. 1999. "A Multilevel Exploration of the Role of Interviewers in Survey Non-Response." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*162 (3): 437–446. https://doi.org/10.1111/1467-985X.00147.

Ongena, Yfke, and Marieke Haan. 2020. "Differences in Interaction Quantity and Conversational Flow in CAPI and CATI Interviews." In *Interviewer Effects from a Total Survey Error Perspective*, edited by Kristen Olson, Jolene D. Smyth, Jennifer Dykema, Allyson L. Holbrook, Frauke Kreuter, and Brady T. West, 164–75. New York: Chapman and Hall/CRC.

Pickery, Jan, and Geert Loosveldt. 2001. "An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse." *Journal of Official Statistics* 17 (3): 337–350.

Pickery, Jan, Geert Loosveldt, and Ann Carton. 2001. "The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys: A Multilevel Approach." *Sociological Methods & Research* 29 (4): 509–523. https://doi.org/10.1177/0049124101029004004.

Prince, Martin J., Friedel M. Reischies, Aartjan T. F. Beekman, Rebecca Fuhrer, Cees Jonker, Sirkka-Liisa Kivelä, Brian A. Lawlor, et. al. 1999. "Development of the EURO–D Scale– a European Union Initiative to Compare Symptoms of Depression in 14 European Centres." *The British Journal of Psychiatry* 174 (4): 330–338. https://doi.org/10.1192/bjp.174.4.330.

Rammstedt, Beatrice, and Oliver P. John. 2007. "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German." *Journal of Research in Personality* 41 (1): 203–212. https://doi.org/10.1016/j.jrp.2006.02.001.

Riphahn, Regina T., and Oliver Serfling. 2005. "Item Non-Response on Income and Wealth Questions." *Empirical Economics* 30 (2): 521–558. https://doi.org/10.1007/s00181-005-0247-7.

Schräpler, Jörg-Peter. 2006. "Explaining Income Nonresponse – A Case Study by Means of the British Household Panel Study (BHPS)." *Quality and Quantity* 40 (6): 1013–1036. https://doi.org/10.1007/s11135-005-5429-z.

SHARE. 2017. "SHARE Release Guide 6.0.0." Munich: Max-Planck-Institue for Social Law and Social Policy, Munich Center for the Economics of Aging.

Singer, Eleanor, and Luane Kohnke-Aguirre. 1979. "Interviewer Expectation Effects: A Replication and Extension." *Public Opinion Quarterly* 43 (2): 245–260. https://doi.org/10.1086/268515.

Snijkers, Ger, Joop J. Hox, and Edith D. de Leeuw. 1999. "Interviewers' Tactics for Fighting Survey Nonresponse." *Journal of Official Statistics* 15 (2): 185–198.

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

Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133 (5): 859–883. https://doi.org/10.1037/0033-2909.133.5.859.

Vassallo, Rebecca, Gabriele B. Durrant, Peter W. F. Smith, and Harvey Goldstein. 2015. "Interviewer Effects on Non-Response Propensity in Longitudinal Surveys: A Multilevel Modelling Approach." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178 (1): 83–99. https://doi.org/10.1111/rssa.12049.

Vercruyssen, Anina, Celine Wuyts, and Geert Loosveldt. 2017. "The Effect of Sociodemographic (Mis)match between Interviewers and Respondents on Unit and Item Nonresponse in Belgium." *Social Science Research* 67:229–238. https://doi.org/10.1016/j.ssresearch.2017.02.007.

Watson, Dorothy. 2003. "Sample Attrition Between Waves 1 and 5 in the European Community Household Panel." *European Sociological Review* 19 (4): 361–378. https://doi.org/10.1093/esr/19.4.361.

Watson, Nicole, and Mark Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 157–181. Chichester, UK: John Wiley.

West, Brady T., and Annelies G. Blom. 2017. "Explaining Interviewer Effects: A Research Synthesis." *Journal of Survey Statistics and Methodology* 5 (2): 175–211. https://doi.org/10.1093/jssam/smw024.

Wooden, Mark, and Nicole Watson. 2004. "Sample Attrition in the HILDA Survey." *Australian Journal of Labour Economics* 7 (2): 293–308.

Yan, Ting, and Richard Curtin. 2010. "The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective." *International Journal of Public Opinion Research* 22 (4): 535–551. https://doi.org/10.1093/ijpor/edq037.

Yan, Ting, Richard Curtin, and Matthew Jans. 2010. "Trends in Income Nonresponse Over Two Decades." *Journal of Official Statistics* 26 (1): 145–164.

Appendix

Table A4.1. SHARE Income questions

Label	Question wording
Employment earnings	"Have you had any wages, salaries or other earnings from dependent employment in 2014?" If so: "After any taxes and contributions, what was your approximate annual income from employment in the year 2014? Please include any additional or extra or lump sum payment, such as bonuses, 13 month, Christmas or Summer pays."
Earnings from self-employment	"Have you had any income at all from self-employment or work for a family business in 2014?" If so: "After any taxes and contributions and after paying for any materials, equipment or goods that you use in your work, what was your approximate annual income from self-employment in the year 2014?"
Public pension payments	"Have you received income from any of these sources in the year 2014?" 1. Public old age pension 2. Public old age supplementary pension or public old age second pension 3. Public early retirement or pre-retirement pension 4. Main public sickness benefits 5. Main public disability insurance pension 6. Secondary public disability insurance pension 7. Secondary public sickness benefits 8. Public unemployment benefit or insurance 9. Main public survivor pension from your spouse or partner 10. Secondary public survivor pension from your spouse or partner 11. Public war pension 12. Public long-term care insurance 13. Social assistance 14. None of these If so: "After taxes, about how large was a typical payment of your [public old age pension/public old age supplementary pension or public old age second pension/public early retirement or pre-retirement pension/ main public sickness benefits/ main public disability insurance pension/ secondary public disability insurance pension/ secondary public sickness benefits/ public unemployment benefit or insurance/ main public survivor pension from your spouse or partner/ secondary public survivor pension from your spouse or partner/ public war pension/ public long-term care insurance/ social assistance] in 2014"
Other regular payments	„Did you receive any of the following regular payments or transfers during the year 2014" 1. Life insurance payments from a private insurance company 2. Regular private annuity or private personal pension payments 3. Alimony 4. Regular payments from charities 5. Long-term care insurance payments from a private insurance company 96. None of these 1. Life insurance payments from a private insurance company 2. Regular private annuity or private personal pension payments 3. Alimony 4. Regular payments from charities 5. Long-term care insurance payments from a private insurance company 6. None of these If so: "After any taxes and contributions, about how large was the average payment of your [life insurance payments from a private insurance company/ private annuity or private personal pension payments/ alimony/ regular payments from charities/ long-term care insurance payments] in 2014?"

Table A4.2. SHARE Interviewer Survey items and operationalization

Dimension	Indicators		Item	Original responses	Operationalization
Sociodemographics	Gender		Are you male or female?	Female Male	Female=1 Male = 0
	Age		In which year were you born?	[1900-1997]	2015 – [Year]
	Education		Which is your highest level of education?	1) Graduated from lower-level secondary school 2) Graduated from medium-level secondary school 3) Advanced technical college entrance qual. or graduated from upper-level secondary school 4) University degree	1 = 1 (low) 2 = 2 (middle) 3/4 = 3 (high)
Experience	General		In what year did you first start working as an interviewer?	[1900 – 2015]	2015 – [Year]
	SHARE		Have you worked as an interviewer on any previous wave of SHARE?	Yes No	Yes =1 No = 0
	Working time		Approximately, how many hours a week do you currently work as an interviewer?	[0 – 100] *Don't know/ Refuse*	Quartiles + item nonresponse [0-10; 11-20; 21-30; 31-70; missing value]
Personality	Big 5 personality traits	Openness:	I see myself as someone who has few artistic interests(R). Do you... I see myself as someone who has an active imagination. Do you...	For each item: Disagree strongly Disagree a little Neither agree nor disagree Agree a little Agree strongly *Don't know/ Refuse*	Sum score [2-10]
		Conscientiousness:	I see myself as someone who tends to be lazy (R). Do you... I see myself as someone who does a thorough job. Do you...	For each item: Disagree strongly Disagree a little Neither agree nor disagree Agree a little Agree strongly *Don't know/ Refuse*	Sum score [2-10]
		Extroversion:	I see myself as someone who is reserved (R). Do you... I see myself as someone who is outgoing, sociable. Do you...	For each item: Disagree strongly Disagree a little Neither agree nor disagree Agree a little Agree strongly *Don't know/ Refuse*	Sum score [2-10]

Table A4.2. (cont.)

Dimension		Indicators	Item		
Personality	Big 5 personality traits	Agreeableness:	I see myself as someone who is generally trusting. Do you...	For each item: Disagree strongly Disagree a little Neither agree nor disagree Agree a little Agree strongly *Don't know/ Refuse*	Sum score [2-10]
			I see myself as someone who is considerate and kind to almost everyone. Do you...		
		Neuroticism:	I see myself as someone who is relaxed, handles stress well (R). Do you...	For each item: Disagree strongly Disagree a little Neither agree nor disagree Agree a little Agree strongly *Don't know/ Refuse*	Sum score [2-10]
			I see myself as someone who gets nervous easily. Do you...		
Attitude		Towards reasons being an interviewer:	Payment	For each item: 1 (Not important at all) – 7 (Very important) *Don't know/ Refuse*	For each item: [1-7]
			Interesting work		
			Opportunity to interact with people		
			Gaining insight into other peoples' social circumstances		
			Involvement in scientific research		
			Involvement in research that serves society		
			Flexible working hours		
		Towards cooperation:	Reluctant respondents should always be persuaded to participate.	For each item: Strongly agree Somewhat agree Somewhat disagree Strongly disagree *Don't know/ Refuse*	For each item: Strongly agree/ Somewhat agree =1 Somewhat disagree/ Strongly disagree =0
			With enough effort, even the most reluctant respondent can be persuaded to participate.		
			An interviewer should respect the privacy of the respondent.		
			If a respondent is reluctant, a refusal should be accepted.		
			One should always emphasize that participation is voluntary.		

Table A.4.2. (cont.)

Dimension	Indicators	Item		
Attitude	Towards cooperation:	It does not make sense to contact reluctant target persons repeatedly.	For each item: Strongly agree Somewhat agree Somewhat disagree Strongly disagree *Don't know/ Refuse*	For each item: Strongly agree/somewhat agree =1 Somewhat disagree/ Strongly disagree =0
		If you catch them at the right time, most people will agree to participate.		
	Towards respondents' reason to participate in survey:	Contributing to research	For each item: 1 (Almost none of the respondents) – 7 (Most of the respondents)	For each item: [1–7]
		Taking part in order to serve society		
		Opportunity to interact with someone		
		Expressing their opinions		
		Inability to say 'no' to the interviewer		
		Receiving incentives or gifts		
	Trust in People	Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people? Please use the scale from 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted	You can't be too careful (0) - Most people can be trusted (10)	[0 –10]
	Data privacy concerns	How concerned are you about the safety of your personal data?	Very concerned Quite concerned A little concerned Not concerned at all *Don't know/ Refuse*	A little concerned/ Not concerned at all =1 Very concerned/ Quite concerned =2 Missing value=3
Self-reported behavior	Conducting interviews:	If the respondent doesn't understand a question, I explain what the question really means.	For each item: Perfectly Somewhat Not really Not at all *Don't know/ Refuse*	For each item: Perfectly/ Somewhat =1 Not really/ Not at all =0
		If the respondent has trouble with a question, I reread the exact question.		
		If I notice that the respondent has trouble listening to me, I shorten long question texts.		

Table A4.2. (cont.)

Self-reported behavior	Conducting interviews:	If I notice that the respondent has trouble understanding the question, I speak more slowly.	For each item: Perfectly Somewhat Not really Not at all *Don't know/ Refuse*	For each item: Perfectly/ Somewhat=1 Not really/ Not at all =0
		If I notice that the respondent is in a hurry, I speak faster.		
		If I know from the course of the interview what an answer will be, I complete the answer myself		
		I stick to the interviewer instructions, even if I don't consider them sensible.		
	Own survey participation	In the last 5 years, how often have you taken part in a survey as a respondent (not counting this survey)?	[0 – 150] *Don't know/ Refuse*	Quintiles + item nonresponse [0; 1; 2–3; 4–9; 10–150; missing value]
	Own survey income response	All in all, approximately what was the average monthly income of your household after taxes in the last year?	[0 – ∞] *Don't know/ Refuse*	Income reported =0 Income item nonresponse =1
	Using online banking	Do you use the internet for online-banking?	Yes No	Yes =1 No = 0
	Using social networks	Do you use social networks on the internet like Facebook or Twitter?	Yes No	Yes =1 No = 0
Survey expectations	Towards respondents' response to income question	Social surveys very often ask about respondents' income. What do you expect, how many of your respondents (in percentage) in SHARE will provide information about their income?	[0 – 100] Percentage *Don't know/ Refuse*	[0–50; 5 –75;76–90; 91–100; missing value]

Table A4.3. Descriptive statistics of categorical interviewer variables

Dimension	Variable	Category	N	%
Sociodemographics	Gender	female	172	64.7
		male	94	35.3
	Education	low	24	9.0
		middle	70	26.3
		high	172	64.7
Experience	SHARE	no	84	31.6
		yes	182	68.4
	Working hours	1-10 hrs.	72	27.1
		11-20 hrs.	77	28.9
		21-30 hrs.	52	19.6
		31+ hrs.	62	23.3
		Missing value	3	1.1
Attitudes towards cooperation	Always persuade reluctant respondents	disagree	84	31.6
		agree	182	68.4
	With effort, respondent persuadable	disagree	109	41.0
		agree	157	59.0
	Respect privacy of respondent	disagree	4	1.5
		agree	262	98.5
	If reluctant, accept refusal	disagree	51	19.2
		agree	215	80.8
	Emphasize voluntary nature of participation	disagree	19	7.1
		agree	247	92.9
	If reluctant, do not contact repeatedly	disagree	137	51.5
		agree	129	48.5
	Caught at right time, people participate	disagree	44	16.5
		agree	222	83.5
Attitude	Data privacy concerns	no/few	70	26.3
		some/many	195	73.3
		missing value	1	0.4
Conducting interviews	Explaining question if respondent misunderstood	no	114	42.8
		yes	152	57.1
	Rereading exact question if respondent misunderstood	no	29	10.9
		yes	237	89.1
	Shortening question if respondent has troubles	no	224	84.2
		yes	42	15.8
	Speaking slowly if respondent has troubles	no	8	3.0
		yes	258	97.0
	Speaking faster if respondent in hurry	no	148	55.6
		yes	118	44.4
	Completing answers if known	no	255	95.9
		yes	11	4.1
	Sticking to instructions, even if not considered as sensible	no	31	11.7
		yes	235	88.4

Table A4.3. (cont.)

Dimension	Variable	Category	N	%
Self-reported behavior	Own survey participation	0	40	15.0
		1	25	9.4
		2–3	52	19.6
		4–9	52	19.6
		10–150	46	17.3
		missing value	51	19.2
	Own income reported	no	79	29.7
		yes	187	70.3
	Using online banking	no	65	24.4
		yes	201	75.5
	Using social networks	no	130	48.87
		yes	136	51.1
Survey expectations	Towards respondents' response to income question	0–50%	36	13.5
		51–75%	64	24.1
		76–90%	110	41.4
		91–100%	36	13.5
		missing value	20	7.5

Note: n(interviewer)=266.

Table A4.4. Descriptive statistics of continuous interviewer variables

Dimension	Variable	Mean	SD	Min.	Max.
Socio-demographics	Age	56.2	11.2	22	78
Experience	General	10.8	10.7	0	56
Personality	Openness	7.4	1.9	2	10
	Conscientiousness	8.8	1.4	2	1.9
	Extroversion	7.5	1.7	2	10
	Agreeableness	8.3	1.4	3	10
	Neuroticism	4.7	1.8	2	8
Attitudes towards reasons being an interviewer	Payment	5.8	1.3	2	7
	Interesting work	6.2	1.0	2	7
	Opportunity to interact with people	5.7	1.5	2	7
	Gaining insight into other peoples' social circumstances	5.2	1.6	1	7
	Involvement in scientific research	5.8	1.3	1	7
	Involvement in research that serves society	5.9	1.2	1	7
	Flexible working hours	6.3	1.2	1	7
Attitudes towards respondents' reason to participate in survey	Contributing to scientific research	6.3	1.2	1	7
	Serve society	4.8	1.3	2	7
	Interacting with someone	4.6	1.3	1	7
	Expressing own opinions	5.2	1.3	1	7
	Inability to say no	3.3	1.5	1	7
	Receiving incentives or gifts	4.6	1.7	1	7
Attitude	Trust in People	6.6	1.8	1	10

Note: n(interviewer)=266.

Table A4.5. Full sequential two-level logistic regression model

	M$_0$		M$_1$		M$_2$	
	Wave response	Income item response	Wave response	Income item response	Wave response	Income item response
Transition	1.35***	3.06***	1.42**	4.47***	.60	6.92**
	(.05)	(.08)	(.44)	(.92)	(1.08)	(2.12)
Respondent characteristics						
Country: Germany (ref.)						
Sweden			.26	.29	.53*	–1.20**
			(.14)	(.22)	(.25)	(.44)
Italy			.32*	–.45*	.48	–1.02*
			(.14)	(.22)	(.24)	(.44)
Slovenia			.32	–1.02***	.84**	–1.70***
			(.19)	(.23)	(.29)	(.48)
Residence: big city (ref.)						
Suburbs/ outskirts of a big city			–1.70***	–.32	–.10	–.24
			(.48)	(.30)	(.13)	(.32)
large town			.16	–.10	.16	.12
			(.13)	(.29)	(.13)	(.32)
small town village			.44***	–.30	.42***	–.16
			(.12)	(.25)	(.12)	(.27)
rural area			.44***	–.29	.30*	–.07
			(.12)	(.24)	(.12)	(.27)
unknown/ not coded			–.88***	–.55	–.90***	–.05
			(.18)	(.45)	(.18)	(.48)
Gender: male (ref.)						
female			–.05	.16	–.05	.14
			(.09)	(.19)	(.09)	(.20)
Age in yrs.			–.01	–.02	–.01*	–.02
			(.00)	(.01)	(.00)	(.01)
Education level: low (ref.)						
medium			.06	.13	.05	–.14
			(.08)	(.09)	(.08)	(.18)
high			.14	–.39*	.13	–.44*
			(.09)	(.18)	(.09)	(.19)
Employment status: retired (ref.)						
working			–.43***	.05	–.42***	.11
			(.09)	(.20)	(.09)	(.20)
other/unknown			–.12	.11	–.11	.10
			(.10)	(.20)	(.10)	(.22)
Marital status: married/ partnership (ref.)						
separated /divorced			.09	–.28	.10	–.31
			(.11)	(.21)	(.10)	(.22)
never married			.08	–.41	.08	–.48
			(.14)	(.29)	(.13)	(.30)
widowed			.19	–.16	.21	–.21
			(.11)	(.20)	(.11)	(.21)
Having children: no (ref.)						
yes			.12	–.31	.11	–.28
			(.10)	(.24)	(.10)	(.25)

Table A4.5. (cont.)

	M_0		M_1		M_2	
	Wave response	Income item response	Wave response	Income item response	Wave response	Income item response
Self–reported health status: poor (ref.)						
fair			.04	.08	.04	.10
			(.13)	(.26)	(.12)	(.26)
good or better			.00	–.07	.01	–.03
			(.13)	(.26)	(.13)	(.26)
Instrumental activities of daily living limitations (IADL)						
IADL score			.30***	–.17	.30***	–.19
			(.06)	(.09)	(.06)	(.10)
Giving help to others: no (ref.)						
yes			.06	–.07	.06	–.00
			(.08)	(.16)	(.08)	(.16)
Receiving help from others: no (ref.)						
yes			.09	.20	.09	.17
			(.07)	(.15)	(.07)	(.15)
Depression score (EURO–D)			–.01	.05	–.00	.05
			(.02)	(.03)	(.02)	(.04)
Memory recall ability test: recalled less than half of the words (ref.)						
more than the half			.28***	–.09	.25***	–.05
			(.07)	(.14)	(.06)	(.14)
Missing value			–1.48***	–.11	–1.47***	–.15
			(.13)	(.37)	(.13)	(.37)
Maximum grip strength: 1st Quartile (ref.)						
2nd Quartile			–.02	.10	–.01	.07
			(.09)	(.18)	(.09)	(.18)
3rd Quartile			–.02	.32	–.03	.29
			(.10)	(.22)	(.10)	(.22)
4th Quartile			–.09	.23	–.08	.22
			(.13)	(.27)	(.13)	(.27)
Missing value			–1.61***	–.64	–1.59***	–.56
			(.16)	(.36)	(.16)	(.38)
Smoking: no (ref.)			–.10	–.01	–.08	–.04
yes			(.08)	(.17)	(.08)	(.18)
Drinking: not at all (ref.)						
rarely			.01	.46	.01	.44
			(.11)	(.27)	(.11)	(.27)
sometimes			.00	–.02	–.00	.10
			(.10)	(.21)	(.10)	(.22)
often			.14	–.06	.14	–.00
			(.09)	(.19)	(.09)	(.19)
daily			.10	.05	.09	.14
			(.09)	(.17)	(.09)	(.18)

Table A4.5. (cont.)

	M_0		M_1		M_2	
	Wave response	Income item response	Wave response	Income item response	Wave response	Income item response
Interviewer characteristics						
Gender: male (ref.)						
female					–.01	–.48*
					(.11)	(.20)
Age in yrs.					–.00	–.04***
					(.01)	(.01)
Education level: low (ref.)						
medium					.03	–.22
					(.19)	(.18)
high					–1.18**	–.73*
					(.37)	(.37)
Experience:						
General interviewer experience in yrs.					.01	–.01
					(.01)	(.01)
Prior worked for SHARE: no (ref.)						
yes					–.01	–.75**
					(.14)	(.24)
Working hours in hrs./week: 0–10 (ref.)						
11–20 hrs.					.03	.02
					(.14)	(.25)
21–30 hrs.					.61***	.50
					(.17)	(.32)
31+ hrs.					.06	–.17
					(.16)	(.27)
missing value					.64	.97
					(.47)	(.88)
Personality traits score:						
– Extroversion					–.00	.03
					(.03)	(.06)
– Conscientiousness					.02	–.04
					(.04)	(.07)
– Neuroticism					.05	–.07
					(.03)	(.05)
– Agreeableness					–.03	–.05
					(.04)	(.08)
– Openness					.03	.10
					(.03)	(.05)
Attitude scores towards importance of reasons for being an interviewer:						
– Payment					–.07	–.03
					(.04)	(.08)
– Interesting work					–.01	.17
					(.07)	(.12)
– Opportunity to interact with people					–.04	.08
					(.05)	(.09)

Table A4.5. (cont.)

	M_0		M_1		M_2	
	Wave response	Income item response	Wave response	Income item response	Wave response	Income item response
Attitude scores towards importance of reasons for being an interviewer gaining insights in other's life:					.03	.10
					(.04)	(.07)
– Involvement in scientific research					–.11	–.08
					(.09)	(.15)
– Involvement in research that serves society					.09	–.02
					(.09)	(.15)
– Flexible working hours					.13**	–.11
					(.04)	(.09)
Data privacy concerns: no (ref.)						
yes					.25	.30
					(.13)	(.25)
missing value					–1.40	–3.64**
					(.88)	(1.24)
Trust in people score					–.01	.06
					(.03)	(.05)
Opinion scores on respondents' reason to participate:						
– Contributing to scientific research					.09	.53***
					(.07)	(.12)
– Serving society					–.09	–.19
					(.06)	(.10)
– Interacting with someone					–.08	–.26**
					(.05)	(.10)
– Inability to say no					.05	.25*
					(.05)	(.10)
– Expressing own opinions					.04	.15*
					(.04)	(.07)
– Receiving incentives					–.05	.10
					(.04)	(.07)
Interviewer self–reported behavior:						
– Explaining question if R misunderstood: no (ref.)						
yes					.21	–.95***
					(.12)	(.21)
– Rereading exact question if R misunderstood: no (ref.)						
yes					–.36	–.04
					(.19)	(.33)
– Shortening question if R has troubles: no (ref.)						
yes					–0.06	0.43
					(0.16)	(0.29)

Table A4.5. (cont.)

	M0		M1		M2	
	Wave response	Income item response	Wave response	Income item response	Wave response	Income item response
Interviewer self–reported behavior:						
– Speaking slowly if R has troubles: no (ref.)						
yes					–.25 (.30)	–1.35 (.73)
– Speaking faster if R in hurry: no (ref.)						
yes					–.00 (.11)	–.42* (.19)
– Completing answers if known: no (ref.)						
yes					–.57 (.33)	–1.23* (.52)
– Sticking to instructions, even if not considered as sensible: no (ref.)						
yes					–.15 (.17)	–.43 (.31)
Attitudes towards gaining cooperation:						
– Always persuade reluctant respondents: disagree (ref.)						
agree					–.08 (.12)	.71** (.23)
– With effort, respondent persuadable: disagree (ref.)						
agree					.14 (.13)	–.26 (.24)
– Respect privacy of respondent: disagree (ref.)						
agree					.72 (.37)	.64 (.78)
– If reluctant, accept refusal: disagree (ref.)						
agree					–.07 (.14)	.02 (.25)
– Emphasize voluntary nature of participation: disagree (ref.)						
agree					–.13 (.22)	–.27 (.42)
– If reluctant, do not contact repeatedly: disagree (ref.)						
agree					–.08 (.12)	.01 (.21)
– Caught at right time, people participate: disagree (ref.)						
agree					–.35* (.16)	–.85** (.31)

Table A4.5. (cont.)

	M_0		M_1		M_2	
	Wave response	Income item response	Wave response	Income item response	Wave response	Income item response
Using online banking:						
no (ref.)						
Yes					.29*	–.45*
					(.13)	(.23)
Using social networks:						
no (ref.)						
yes					–.28*	.18
					(.13)	(.23)
Own survey participation						
last 5 yrs.: no surveys (ref.)						
1–3 surveys					.18	–.02
					(.23)	(.31)
4–7 surveys					.06	–.14
					(.19)	(.32)
7+ surveys					–.12	–.01
					(.19)	(.36)
missing value					.12	.01
					(.18)	(.32)
Income reported: no (ref.)						
yes					.01	.75***
					(.32)	(.21)
Expectations towards percentage reporting income: 0–50 % (ref.)						
51–75%					.25	.70*
					(.17)	(.27)
76–90%					.34*	.96***
					(.17)	(.26)
91–100%					.16	.73*
					(.21)	(.36)
missing value					.21	–.11
					(.25)	(.41)
Log. likelihood	–4,251.5732		–3,981.98		–3,841.56	
Chi-squared	1,509.08		1,674.00		1,776.91	

Note: n(respondent)=11,264 and n(interviewer)=266. No. of respondents: sum of observations for wave response (6,347) and item response (4,917). M_0: Null model. M_1: Model with control variabels at respondent level M2: Model with control variabels at respondent level and explanatory variables at interviewer level. According to Hox, Moerbeek, and van de Schoot (2010), M_1 and M_2 were estimated with a scale correction factor of 0.86. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5. Summary, Conclusion and Outlook

Data that was not collected in longitudinal surveys can lower the precision of estimates and can bias study results. Therefore, in my thesis, I focused on missing data with the aim to improve the longitudinal survey data by understanding and preventing missing data that is likely to occur and to introduce bias in analyses based on longitudinal data. In particular, my work contributed to the knowledge of multiple error sources of missing data (attrition, wave nonresponse, and income item nonresponse) in a longitudinal survey – the Survey of Health, Ageing and Retirement in Europe (SHARE).

With my thesis I highlighted further that these types of nonresponse reflect two components of the *Total Survey Error (TSE) framework*: nonresponse error and measurement error, with income item nonresponse reflecting both. To evaluate the total quality of survey statistics, we would need to investigate all kinds of survey errors which calls for future research because each component of the TSE has its own specialties and needs a careful evaluation (see Groves et al. 2009).

I investigated three types of nonresponse: attrition, wave nonresponse, and income item nonresponse. All these types of nonresponse are likely to introduce bias into the survey statistics because those who respond may be likely to differ systematically from those who do not respond. Moreover, if we place these types of nonresponse within the *response continuum model* the interconnection between unit nonresponse and item nonresponse is particularly important to be considered in longitudinal studies. If one type of nonresponse can predict another type of nonresponse, we should treat them as connected survey errors in order to prevent missing data in longitudinal surveys effectively.

In this thesis, I answered three research questions with three studies. My first study aimed to answer the question "How many initially recruited individuals for a longitudinal survey drop out over 12 years of data collection and do those who drop out differ systematically from those who do not?" (see chapter 2). My study shows that initially recruited individuals drop out especially after their first participation. In the second wave, about 70% of the initially recruited sample participated again, and after 12 years in the sixth wave, about 40% of the initially recruited sample remained. This drop out of approximately 60% over years may bias estimates, especially in measures of change, because those who dropped out are, to some extent, systematically different from those who remained. The results show that the sample composition changed after the first wave but remained relatively stable after the second wave. In my study, the sample composition over waves changed with regard to respondent characteristics, such as birth in survey country, area of residence, education, and social activities. Those who were born in the survey country, live in a small town or rural village, are highly educated, and are socially active are less likely to attrite after their first interview than those who were born outside the survey country, live in a city or large town, are low educated, and are not socially active. Therefore, researchers who want to study economic, social, and health aspects of sub-groups of the older population in Europe from a longitudinal perspective should consider these constraints in order to be able to draw correct inferences. Finally, this study shows that it is important to differentiate between the respondents who dropped out because they died and those who dropped out for other reasons, because, they are very likely to be different from each other.

To prevent missing data, it is necessary to identify its determinants and to understand the relationship first. In face-to-face surveys, apart from the respondent, the survey design, and the social environment, interviewers play an important role. According to the *conceptual framework for survey cooperation* they are the key agents in face-to-face studies and they can positively and negatively affect missing data. Despite well-designed survey, effective interviewer trainings, and study monitoring, any survey is at risk of failing to collect data that was intended to be collected. However, in my thesis I bring forward new insights on the link between interviewers and missing data which can be harnessed to adapt trainings in order to reduce the probability of missing data and the variation in missing data induced by interviewers.

Therefore, one promising determinant of missing data that can help to prevent missing data in longitudinal studies with face-to-face data collection is the interviewer. Thus, my second study (see chapter 3) aimed to answer the research question "To what extent do interviewers contribute to the occurrence of missing data in income and can we explain this link between income item nonresponse and interviewers?". My study shows that interviewers contribute to the variance in missing data and that interviewer characteristics can partially explain this link. In my study, interviewer characteristics explained on average about one third of the total variances of income item nonresponse rates. Moreover, interviewers' having optimistic expectations about income reporting has a positive effect on respondents' propensity to report income. My study results show that item nonresponse rates could be decreased by up to 13 percentage points with interviewers who expect more than 90% of their respondents to report their income. Thus, in order to prevent income missing data, research practitioners should design new interviewer training methods that allow to boost interviewers' expectations.

According to the *response continuum model* unit nonresponse and item nonresponse are interconnected. Respondents who lie on the left-hand side of the continuum are very unlikely to participate in a survey and those who lie on the right-hand side of the continuum are very likely to participate and to answer all survey questions. In my third study (see chapter 4), I considered this interconnection as well as the role of the interviewer as a key agent in the data collection process that can prevent missing data in unit nonresponse and income item nonresponse at the same time. Thus, the third article answers the research question "Does missing data caused by unit nonresponse and income item nonresponse have common causes that can be located with interviewers collecting the data?". My study shows that there are a few common causes of interviewer effects. (Non) optimistic attitudes explain interviewer effects in wave nonresponse and income item nonresponse. Again, optimistic expectations towards income reporting can decrease item nonresponse rates, and moreover, decrease wave nonresponse rates. In contrast, interviewers' attitude towards gaining cooperation (agreeing with "if you catch people at the right time most will participate") can increase both, wave nonresponse rates and income item nonresponse rates. Thus, this study shows that new training designs that boost interviewer expectations could increase both income response rates and participation rates, and that addressing non-optimistic attitudes in interviewer trainings could be effective in preventing missing data as well.

As my studies explained the link between nonresponse and interviewers only partially, further research is needed to fully uncover the attitudinal and behavioral mechanisms of interviewers and their connection to survey statistics and survey outcomes. Only if we know which behavioral and attitudinal mechanisms of the interviewers are related to missing data, we can address interviewer behavior and attitude properly in interviewer trainings. Furthermore, there is need for tests in experimental settings which can effectively reduce interviewer variance and missing data are. Therefore, I suggest to survey the nature of missing data in other studies in order to find the most effective approach for missing data prevention. Furthermore, I advise researchers to inform about the extent and the impact of missing data in the data they use and to make assumptions about the missing data before using survey data to provide as accurate survey statistics and estimations for the target population as possible.

Another idea to overcome the occurrence of and variation in missing data caused by interviewers could be to move face-to-face surveys to online self-completion surveys. Studies have shown, that item nonresponse, one type of missing data, can be reduced by switching from a face-to-face mode to an online mode (de Leeuw 2005; Dillman 2007). However, other studies have also shown that online surveys are prone to high breakoff rates, which would generate even a larger extent of missing data (Peytchev 2009).

Moreover, it is well-known that internet access is limited or simply unavailable in many regions (Digital Economy and Society Index). Also, not all people use actively the internet (Statista 2020). In EU, 11% of the population does not use the internet (Internet World Stats 2020), and those who do not use the internet are more likely to be older, unemployed, and have a lower educated level than those who use the internet (Helsper and Reisdorf 2017). As SHARE covers the target population 50+ in Europe, switching from face-to-face to online might be a greater challenge. Older people belong the class of people with lower levels of digital affinity (Herzing and Blom 2019) and thus, are less likely to use the internet. In addition, panel respondents get older with every survey wave and their skills, cognitive ability and health will potentially decrease. This further minimizes the probability of them participating in online surveys. Since the access and use of the internet differs across EU countries this may be likely the case. Further, as Yan and Curtin (2010, 536) acknowledged that "any change in survey design features such as the mode of administration, interviewer assignment, and interviewer behavior will change a sample person's propensity to take part in a survey and/or to answer a survey question." For some respondents the switch to online might increase their propensity to respond whereas for others it decreases. SHARE with interviewing for example also very old people (80+) may risk to lose this group of respondents by switching.

Furthermore, SHARE is still collecting data, and switching from one mode to another (i.e., face-to-face vs. online) between waves may induce mode effects, i.e. answers given differ by mode. For longitudinal survey data this means that "it is hard to decide whether a change over time is a real change for the person surveyed or if it is caused by a change in mode" (de Leeuw 2005, 239–240). In addition, SHARE collects numerous valuable additional data, such as physical performance tests, cognitive tests and, biomarkers. The feasibility of these tests and the data collection without interviewers may be challenging or even impossible (de Leeuw 2005). Therefore, I suggest to carefully think about the survey design and potential errors that may occur, and decide which design might be best for the target population researchers want to survey.

In closing, this thesis investigated missing data in longitudinal face-to-face surveys. There are many causes for missing data. By focusing on nonresponse as one cause of missing data and investigating the link between the interviewer and nonresponse, I opened up a plethora of avenues for future research. It is up to future researchers to implement new methods of preventing missing data. In my thesis, I provided some insights into the prevention of missing data and I look forward to applying this new knowledge in future data collection procedures.

References

de Leeuw, Edith D. 2005. "To Mix or Not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics* 21 (5): 233–255.

Dillman, Don A. 2007. *Mail and Internet Surveys: The tailored Design Method--2007 Update with new Internet, Visual, and Mixed-Mode Guide.* New York: John Wiley.

Digital Economy and Society Index. 2020. Digital Economy and Society Index (DESI). https://ec.europa.eu/digital-single-market/en/desi.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. New York: John Wiley.

Helsper, Ellen J., and Bianca C. Reisdorf. 2017. "The Emergence of a “Digital Underclass” in Great Britain and Sweden: Changing Reasons for Digital Exclusion." *New Media & Society* 19 (8): 1253–1270. https://doi.org/10.1177/1461444816634676.

Herzing, Jessica M. E., and Annelies G. Blom. 2019. "The Influence of a Person’s Digital Affinity on Unit Nonresponse and Attrition in an Online Panel." *Social Science Computer Review* 37(3): 404–424. https://doi:10.1177/0894439318774758

Internet World Stats. 2020. "Internet Usage in the European Union." https://www.internetworldstats.com/stats9.htm.

Peytchev, Andy. 2009. "Survey Breakoff." *Public Opinion Quarterly* 73 (1): 74–97. https://doi.org/10.1093/poq/nfp014.

Statista. 2020. "Global digital population as of July 2020." https://www.statista.com/statistics/617136/digital-population-worldwide/.

Yan, Ting, and Richard Curtin. 2010. "The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective." *International Journal of Public Opinion Research* 22 (4): 535–551. https://doi.org/10.1093/ijpor/edq037.