Gotta survey somebody
New series 60
Gotta survey somebody

Methodological challenges in population studies of older people

Susanne Kelfve
Till Mimmi
Abstract

Conducting representative surveys of older people is challenging. This thesis aims to analyze a) the characteristics of individuals at risk of being underrepresented in surveys of older people, b) the systematic errors likely to occur as a result of these selections, and c) whether these systematic errors can be minimized by weighting adjustments.

In Study I, we investigated a) who would be missing from a survey that excluded those living in institutions and that did not use indirect interviews, b) how prevalence rates would be affected by these exclusions, and c) whether post-stratifying the data by sex and age (weighting adjustment) would correct for any systematic measurement error.

In Study II, we compared mortality and hospitalization rates in those who responded to a postal questionnaire with rates in the target population. In addition, we tested whether a weighting variable created with a number of auxiliary variables could correct for the differences.

In Study III, we followed a longitudinal cohort sample for 43 years. By recalculating baseline characteristics at each follow-up, we investigated how the sample changed after a) selective mortality and b) survey non-participation.

In Study IV, we investigated whether the systematic non-participation that is likely to occur in surveys of older people affects the association between education and health.

In sum, the results of these four studies show that people in the oldest age groups, women, those of low socioeconomic position, and those with the poorest health tend to be underrepresented in surveys of older people. This systematic underrepresentation might lead to an underestimation of poor health and function, a bias that is unlikely to be corrected by weighting adjustments, and to an underestimation of health inequality between educational groups. The results also show that the selective mortality that occurs in longitudinal samples might be compounded by selective non-participation among the most disadvantaged groups.
Sammanfattning

Att genomföra frågeundersökningar bland ett urval av äldre personer som ska representera en population, till exempel alla äldre personer över 80 år i Sverige, är en utmaning. Denna avhandling syftar till att undersöka a) vilka grupper av äldre person som riskerar att bli underrepresenterade i frågeundersökningar, b) vilka systematiska fel som kan uppkomma som en konsekvens av denna underrepresentation, samt c) om dessa systematiska fel kan minimeras genom att vikta data.

I den första delstudien, studie I, undersökte vi a) vilka grupper av äldre personer som skulle bli underrepresenterade i en undersökning som inte inkluderar personer som bor i särskilt boende och som inte använder indirekta intervjuer, det vill säga inte låter någon annan person som har god kunskap om urvalspersonen svara då urvalspersonen själv inte kan medverka b) hur prevalensen av ohälsa skulle påverkas av denna studiedesign, och c) om post-stratifiering utifrån kön och ålder (viktning av data) kan korrigera för dessa systematiska fel.

I den andra delstudien, studie II, jämförde vi personerna som svarat på en postenkät med populationen undersökningen var avsedd att spegla avseende sjukhusinläggningar under en tvåårsperiod, samt andelen döda ett år efter undersökningen. Anledningen var att vi ville se om svarande grupper var friskare eller sjukare än gruppen som inte deltog i undersökningen. Dessutom testade vi om en viktningsvariabel som skapats med ett antal hjälpvarianbler kunde korrigera för eventuella skillnader mellan svarande grupper och bortfallet i undersökningen.


I den fjärde och sista delstudien, studie IV, undersökte vi hur sambandet mellan utbildning och hälsa påverkas av den systematiska selektion som
ibland förekommer i undersökningar av äldre personer (att inte inkludera
personer i särskilt boende och att inte använda indirekta intervjuer).

Sammanfattningsvis visade resultaten i de fyra delstudierna att äldre perso-
ner som av olika anledningar exkluderas från urvalet eller inte deltar i under-
sökningar oftare är kvinnor än män, de har i genomsnitt lägre utbildnings-
nivå och högre medelålder, de har dessutom oftare hälsoproblem och funk-
tionsnedsättningar än personerna som deltar i undersökningar. Detta riskerar
att leda till att hälsoproblemen i den äldre befolkningen underskattas. Result-
taten kan således visa på en friskare och mera funktionsduglig grupp äldre än
vad som är fallet. Denna underskattning är dessutom svår att i efterhand
korrigera för. Resultaten visar också att mindre privilegierade gruppers häl-
soproblem tenderar att underskattas, liksom sambandet mellan utbildnings-
nivå och senare hälsa. Följaktligen kan tolkningen av förhållandet mellan
utbildning och hälsa på äldre dar bli felaktig. Slutligen visade resultaten att i
undersökningar där samma individer följs över längre tid, tenderar urvals-
personer med låg utbildning eller i typiska arbetarklassyrken i högre ut-
sträckning än personer med hög utbildning och tjänstemannayrken att falla
från undersöknings, dels genom en högre dödlighetsrisk, men också genom
lägre deltagande. Därmed förändras undersökningsgruppens karaktär och
med stor sannolikhet även de resultat som genereras utifrån longitudinella
data.
List of original publications

The thesis is based on the following studies referred to in the text by their respective Roman numerals.


IV Kelfve, S. (Manuscript). Underestimated health inequalities among the oldest old – A consequence of excluding the most disabled and disadvantaged.

The published articles are reprinted with the kind permission from the publishers.
Contents

1. Introduction .......................................................................................................................... 13

2. Background .......................................................................................................................... 16
   2.1 Survey Methodology ....................................................................................................... 16
   2.1.1 Weighting adjustment ............................................................................................... 19
   2.2 Surveying the older old ............................................................................................... 21
   2.3 Sociology and the role of the sample survey ............................................................... 26
       2.3.1 Welfare research .................................................................................................. 29
       2.3.2 Social stratification .............................................................................................. 31
       2.3.3 Sociology of aging ............................................................................................... 32
       2.3.4 Current and future challenges for the sample survey .......................................... 38

3. Aims ..................................................................................................................................... 45
   3.1 Specific objectives .......................................................................................................... 45

4. Materials and methods ....................................................................................................... 46
   4.1 The Swedish Level of Living Survey and the Swedish Panel Study of Living Conditions of the Oldest Old ........................................................................................................... 46
   4.2 The Stockholm Public Health Survey .......................................................................... 48
   4.3 Registers ........................................................................................................................ 48
   4.4 Measurements ............................................................................................................... 49
   4.5 Statistical analyses ........................................................................................................ 52

5. Overview of the studies ...................................................................................................... 55
   5.1 Study 1 .......................................................................................................................... 56
   5.2 Study 2 .......................................................................................................................... 57
   5.3 Study 3 .......................................................................................................................... 58
   5.4 Study 4 .......................................................................................................................... 59

6. Discussion ........................................................................................................................... 60
   6.1 Concluding remarks ...................................................................................................... 63

7. Acknowledgements ........................................................................................................... 65

8. References .......................................................................................................................... 68
1. Introduction

The sample survey is a cornerstone of empirical sociology and of disciplines such as epidemiology, economics, and public health. Statistical sampling strategies make it possible to investigate many phenomena without collecting data from all people in the population. Sample surveys have been used to investigate a variety of phenomena in populations, such as health, finances, welfare, and attitudes.

The results of these kinds of surveys can describe society. They can form our understanding of living conditions and attitudes in different groups in a population and changes in these factors over time. They can teach us about the relative and absolute advantages experienced by different sociodemographic groups and how these advantages are formed and changed. They can illuminate the effects of political reforms on the society as a whole and on specific groups. They can provide information about health and health behaviors in the population and sometimes about causal relationships between such factors.

The underlying assumption in a sample survey is that those individuals who participate are representative of the population to be studied. Any systematic non-participation may lead to biased results; i.e., results that do not accurately reflect the conditions in the population. Conducting representative surveys is always challenging. This thesis is about the specific challenges faced when surveying older people.

Most industrialized societies have experienced a dramatic demographic shift in the past century. The number of people who survive to old age has increased substantially, primarily because of improved living conditions and progress in medical care. Meanwhile, the birth rate has declined, and many couples choose not to have more than a few children. The result of increased life expectancy and low birth rates is an increase in the proportion of older people in the population (Christensen, Doblhammer, Rau, & Vaupel, 2009).

Despite increased life expectancy and better treatments for many diseases, most of us will experience a period at the end of life when we depend on others for financial support and care: our family, the welfare state, or both. Older people are thus one of the main target groups of the welfare state. Ac-
cordingly, they are vulnerable to political decisions and reforms (Biggs & Powell, 2001; Trydegård & Thorslund, 2010). Surveys have been an important tool for evaluating how effectively the welfare state works and how sociopolitical decisions affect different groups in society, including older people.

Surveys are also used to increase our knowledge about the aging process. During the recent decades, new survey databases have been initiated, often with longitudinal, multidisciplinary, and sometimes international designs. These surveys often include a wide range of sociodemographic information, behavioral data, and biological markers related to the aging process (Börsch-Supan et al., 2013; Hofer & Piccinin, 2010). Such surveys frequently adopt a life-course approach in which researchers examine the connections between exposures from throughout the life span and later outcomes or chains of outcomes (Kuh, Cooper, Hardy, Richards, & Ben-Shlomo, 2014).

Conducting representative surveys of older people is challenging. Many of these challenges are due to the poor health and function that often accompany old age. For instance, poor health and function increase the likelihood of living in an institution, and it is not unusual to exclude people living in institutions from surveys for ethical, legal, or practical reasons (Lafortune & Balestat, 2007). Additionally, older people are sometimes unable to participate in surveys because of frailty or cognitive impairment (Chatfield, Brayne, & Matthews, 2005; Hardy, Allore, & Studenski, 2009; Manton & Suzman, 1992). Thus, in old age, both the probability of being included in a survey and the probability of being able to respond to it are correlated with health and functional ability. In addition, longitudinal surveys that follow individuals into old age face the problem of sample attrition due to mortality. That is, over time, longitudinal samples gradually became smaller and include only those selected individuals who survived to old age.

Survey response in general has been declining for decades, and today it is not unusual to have response rates below 50 percent (Curtin, Presser, & Singer, 2005; Galea & Tracy, 2007). Moreover, not all surveys use fieldwork strategies specifically designed to facilitate the inclusion of those with poor health and function. As a consequence of the likelihood of non-random survey participation in old age, selective mortality, and high non-response rates, many surveys might not accurately reflect the conditions among older people in the population.

Although researchers and others sometimes acknowledge that non-representative samples are a problem and discuss the implications of such samples, we often go on use such samples anyway – in research, population
surveillance at the national level, and cross-national comparative studies. This is often understandable and even unavoidable; achieving a non-selected sample is challenging and sometimes impossible. The question is the magnitude of effects a selected sample has on the study’s results and how well we can adjust for these effects. We hypothesized that the use of surveys with non-representative samples is a more important problem than typically acknowledged, both in terms of effects on study results and potential implications for public policy based on these results.

_Aim_

This thesis explores and illuminates the problem of estimating prevalence rates and calculating correlations with data from non-representative study populations in surveys of older people and evaluates methods designed to compensate for these consequences. The focus will be on the older old (those above 75 years), but some analyses include the younger old (those between 65 and 75 years). Both cross-sectional studies and studies with a longitudinal design will be covered. The results will be discussed in relationship to the ongoing debate about the importance of representative survey samples.
2. Background

2.1 Survey Methodology

In empirical sociology and other social sciences, the aim is often to investigate a specific population, such as all citizens in a country, all pregnant women in a specific region, or all students at a specific school. One way of investigating such a population is to conduct a sample survey. The aim of a sample survey is usually to estimate parameters in the population, such as income, the prevalence of a disease, the occurrence of alcohol consumption, or the occurrence of a certain opinion. Statistical probability sampling makes it unnecessary to investigate all people in the population. Instead, it is possible to explore a question by investigating a random sample of people from the population. From this sample, it is possible to estimate the population parameter of interest and to calculate the amount of uncertainty linked to the estimate (Scheaffer, Mendenhall III, & Ott, 2006). The sample survey is thus a cost-effective, time-saving, and powerful tool used by researchers from multiple disciplines, policy makers, and market researchers and for measuring public opinion.

The statistical inference methods used in probability sampling are based on the assumption that the sample has been selected from the population using a known probability mechanism. In other words, all people in the target population must have a known probability of being selected. With this knowledge, it is then possible to estimate the precision of a parameter; that is, how certain we can be of the true value of the parameter in the population. Tests of statistical significance build on the assumption of random deviation from the population. Any non-random deviation (from the target population) in the selected sample might result in biased estimates.

In a perfect survey, all people in the target population would have a known probability greater than zero of being included in the sample, the sample would be large enough to provide highly precise estimates, all people in the sample would respond to the survey, and all answers in the survey would accurately reflect reality. This is seldom if ever the case. Several different errors can occur when conducting a survey.
In this chapter, I have used de Leeuw, Hox, and Dillman’s (2008) survey terminology. They describe the statistical part of survey methodology as having four cornerstones: coverage, sampling, response, and measurement. Four types of data collection error stem from these concepts: coverage, sampling, non-response, and measurement error. The first three are errors of inference and sampling, whereas the fourth, measurement error, is a problem of validity and reliability. The four cornerstones and four errors are separate from the theoretical and conceptual aspects of survey methodology, such as construct validity (i.e., whether a survey measures what it is supposed to measure).

**Coverage and coverage error**

As previously noted, the purpose of a survey is to say something about a specific population, the target population. Inference; that is, drawing conclusions about a target population on the basis of data from a sample, builds on the assumptions of probability sampling. How probability samples can be drawn depends on the existence of suitable sampling frames, such as population registers. Coverage error occurs when the sampling frame does not exactly match the target population. If parts of the target population are not included in the sampling frame, their probability of being selected is zero (de Leeuw et al., 2008).

The size of the coverage error depends on the quality of the population register used as the sampling frame. Access to high quality registers varies over time and by country. Coverage (whether or not most people in the target population are included) can differ, and some registers can be up-to-date, whereas others are not. Countries like Sweden, which has a system of personal identification numbers, can generate sampling frames at the individual level, and these sampling frames have usually high coverage rates. Other possible sampling frames include registers of inhabitants in municipalities and lists of households’ postal addresses or telephone numbers. The kind of sampling frame used influences the likelihood that certain categories of people will not be covered. For example, the shift from landlines to cell phones has affected surveys that rely on landline telephone numbers. Young people are likely to be missing in these surveys, as a comparatively large number of them live in households without landlines (Blumberg & Luke, 2007). Marginalized groups, such as homeless people or unregistered immigrants, can
also be difficult to reach. These groups are often missing in sampling frames and therefore from surveys (Carr-Hill, 2013).

**Sampling and sampling error**

Sampling error is the error due to the probabilistic nature of the sampling procedure. The size of the sampling error depends on the size of the sample: the larger the sample, the smaller the sampling error. This means it is possible to calculate the size of the sampling error, and as a consequence, the precision of the survey estimate. When a simple random sampling procedure is used, the size of the sampling error is easy to calculate with standard statistical techniques. However, when more complex sampling methods are used, such as cluster or stratified sampling, the calculation of the sampling error becomes more complex (de Leeuw et al., 2008).

**Response and non-response error**

Non-response occurs when a sampled individual does not participate in a survey (unit non-response) or does not answer some of the questions (item non-response). Non-response errors arise when those who do not respond differ from the respondents regarding a factor that is related to the studied outcome. Statistical inference methods based on probability samples are not designed for non-response; instead they build on the assumption of 100 percent response. The error term that defines the precision of an estimate derives only from the sampling error.

Non-response occurs in all surveys, but not all non-response leads to non-response errors. If the non-response occurs at random (“missing at random” or MAR), nothing but the statistical power is violated. However, it is often difficult to evaluate whether missing data are MAR or “missing not at random” (MNAR). Because the non-response error depends on the correlation between the outcome variable and the reason for non-participation, there is no fixed level of acceptable non-response rate. Hence, there is no linear relationship between non-response rate and non-response bias, and the non-response bias could vary between estimates in a survey (Groves, 2006).

Non-response rates differ between countries, over time, and between different groups. There is a substantial literature about factors that influence response rates and about the impact of these factors on the results of surveys.
In addition to survey design, factors such as culture, incentives, and interviewer characteristics have been discussed as potentially affecting survey response. For an overview of non-response in surveys, including different types of non-response, differences in non-response between survey modes, the potential impact of non-response on the results, and methods that can be used to adjust for non-response, see (Groves, Dillman, Eltinge, & Little, 2002).

Although participation rates vary substantially between surveys, survey response rates have generally declined during the past few decades. This decline has been reported in all types of household surveys (Groves, 2006), including epidemiological surveys (Galea & Tracy, 2007), and surveys targeting older people (Gao et al., 2015). In the 1960s in Sweden, a non-response rate of 10 percent was discussed as a potential problem (S. Johansson, 1971). Today, many surveys have problems reaching 50 percent response rates (Galea & Tracy, 2007). Survey non-response is also unevenly distributed across different groups. In general, non-response is higher among people in lower socioeconomic groups, people who are unemployed, and people who are unmarried (Galea & Tracy, 2007).

**Measurement and measurement error**

The fourth potential survey error is measurement error. In contrast to the previously discussed survey errors, measurement error is only about the validity and reliability of the estimate. That is, does the answer to a question correspond to the “true” value? For example, does the amount of alcohol that respondents state they have consumed in the last week correspond to the actual amount? Measurement error will always occur in a survey, but to a varying extent. Random measurement errors violate the reliability of the estimate, and systematic measurement errors, the validity (de Leeuw et al., 2008). For example, if the deviation from the actual amount of alcohol consumed is randomly distributed across respondents, the error is random, whereas if most respondents, or a certain group of respondents, underestimate their alcohol consumption, the error is systematic.

2.1.1 Weighting adjustment

One way to improve the precision of a population estimate and deal with some of the bias that stems from data collection errors is to weight the sur-
vey data; that is, to give more weight to some people than others in the analyses. In practice, this means that a new variable is created in which each individual is given a number. This number, referred to as the respondent weight, can be interpreted as the proportion of the people in the target population that the respondent represents. Weighting adjustment is commonly used in surveys, primarily to compensate for unequally distributed sampling in the population, and to reduce the bias in the survey estimates that is caused by non-response errors and non-coverage errors (Biemer & Christ, 2008).

There are three main types of survey weights. They are intended to compensate for 1) unequal sample inclusion probability, 2) non-response, and 3) discrepancies between the target population and the analytical sample (post-stratification). Sometimes these types of weights are discussed as the three stages of survey weighting (Kalton & Flores-Cervantes, 2003). They are referred to as stages because even though they are sometimes used separately, when they are used together, the final weight is created in the order of the stages listed above.

The first type of weight is related to the sampling procedure. That is, in some situations, weights can be used to correct for unequal probability that people will be selected for the sample. For example, if some people have had a higher probability of being selected for the sample because of oversampling of specific groups in the target population, unbiased population estimates could be derived by giving respondents weights that are inversely proportional to their selection probability. This type of weight is usually referred to as the base weight (Goedemé, 2013; Kalton & Flores-Cervantes, 2003; Winship & Radbill, 1994).

The second type of weight is related to non-response. Non-response weighting adjustment is a procedure that aims to correct the sample for unit non-response. The main idea is to find respondents similar to the non-respondents with regard to factors that are known for both groups (auxiliary variables). If there are only a few such factors, then a simple cell weighting procedure can be done. That is, respondents and non-respondents can be sorted into different cells depending on their auxiliary variable characteristics. More weight is then given to the respondents in the cells with too few respondents. Sometimes a number of auxiliary variables are available, such as sociodemographic factors (e.g., sex, age, and income) or other information that may have been collected during field work (e.g., health or reason for non-response). These variables could be used for a more complex
weighting adjustment in which different regression models are performed (Biemer & Christ, 2008; Kalton & Flores-Cervantes, 2003).

Finally, the third type of weight is related to discrepancies between the target population and the analytical sample. This weight is constructed by post-stratifying the sample so that it better represents the population; that is, by giving more weight to participants with characteristics that are underrepresented in the study sample. These post-stratifying weights are usually created with known population demographic factors, such as sex and age. Post-stratification is used to compensate for non-coverage and also sometimes for non-response (Goedemé, 2013; Kalton & Flores-Cervantes, 2003).

Although the first type of weight, the base weight, is only related to the sampling procedure, non-response weighting and post-stratification depend on the quality of the auxiliary variables that are available and on the methods used to create the weight variable. To be effective, a weighting variable must predict both the outcome and the probability of survey participation. It is rare to find one single effective variable; usually several variables must be used (Groves, 2006; Kreuter & Olson, 2011).

Non-response weighting adjustments partly reduce non-response bias, although the effect varies with the outcome (Hall, Brown, Nicolaas, & Lynn, 2013; Van der Heyden et al., 2014). In addition to having an uncertain effect on the outcome, non-response weighting often increases parameter variance (Goedemé, 2013). Hence, the benefit of using non-response weighting adjustment must be balanced against the increased parameter variance that is likely to occur as a result of the adjustment (Höfler, Pfister, Lieb, & Wittchen, 2005; Kreuter & Olson, 2011).

2.2 Surveying the older old

Conducting high-quality surveys with few errors is always challenging. However, some groups are more difficult to survey than others. First, as mentioned above, people from some groups, such as homeless people or unregistered immigrants, are likely to be absent from sampling frames. Gathering information about these groups usually requires research methods other than surveys based on probability sampling. Second, some groups, such as children, are difficult to interview because of legislation or parents who act as gatekeepers. Third, some groups are difficult to survey because they have a hard time responding to questionnaires or answering interview questions. For example, language problems may make it difficult for immi-
grants to participate in a survey, whereas physical or psychological impairments might prevent others from participating (Tourangeau, Edwards, Johnson, Wolter, & Bates, 2014).

This thesis is about the specific problems that occur in surveys of the older old. Surveying the older old poses special coverage, response, and measurement challenges.

Coverage

In old age, it is common to live in an institution, such as a nursing home. In 2015, 13 percent of those 80 years or older in Sweden were living in an institution (NBHW, 2015c). It is not unusual to exclude people living in institutions from surveys for ethical or practical reasons (Gudex & Lafortune, 2000; Lafortune & Balestat, 2007). Such reasons can include the lack of a register of people in institutions to use as sampling frame or legal obstacles that prevent researchers from contacting those who live in institutions.

The reason for excluding people living in institutions affects whether or not their exclusion will bias the survey results. If the target population is restricted to community-dwelling people, excluding those who live in institutions is unproblematic. However, if the aim is to say something about the total population 80 years and above and the sample frame only includes community-dwelling people, excluding those who live in institutions creates a substantial discrepancy between the target population and the sample frame. This discrepancy results in coverage error, as the characteristics of those living in the community are unlikely to reflect the characteristics of a random sample of the target population.

The probability of living in an institution in old age has been associated with a number of factors, such as age, health status, chronic conditions, education, marital status, smoking, and alcohol consumption (Asakawa, Feeny, Senthilselvan, Johnson, & Rolfson, 2009); disability, hearing and visual impairment, and sex (Rodgers & Herzog, 1992; Wallace, Kohout, & Colsher, 1992); family situation (Noël-Miller, 2010); and socioeconomic factors such as income and home ownership (Martikainen, Nihtilä, & Moustgaard, 2008). Hence, excluding people who live in institutions might generate biased estimates of all factors that correlate with the probability of living in an institution in old age.

The probability of living in an institution also varies between countries and over time. The threshold for an older individual to become eligible for
institutional care depends on a variety of factors, including the prevailing welfare system, current social policies, the availability of institutional facilities, and the individual’s family situation (Daatland & Lowenstein, 2005; L. Johansson, Sundström, & Hassing, 2003). This means that even if they are restricted to community-dwelling people, survey samples from different countries are not necessary comparable. For example, if the threshold for institutional care is lower in country A than country B, a survey might find a lower prevalence rate of poor health in the community-dwelling people in country A than country B, although there are no real differences in the prevalence of poor health in the two populations. If the threshold for institutional care is raised, a survey based only on community-dwelling people will show an increase in the prevalence of poor health even though the health status of the population has remained constant. Accordingly, excluding people living in institutions might affect results in ways that vary between populations and over time. One factor examined in this thesis is how the exclusion of people living in institutions affects the results of studies of older people.

Response

As discussed earlier, non-response error occurs when the non-response is non-random and the non-respondents differ from the respondents with regard to a factor that is related to the outcome or with regard the outcome itself. In surveys of the older old it is unlikely that non-response is random. In older people, frailty and cognitive impairment become increasingly prevalent with age. As a result, it is difficult for some older people to participate in surveys. A frail person might find it too exhausting to cope with an interview, a person with cognitive impairment might find it difficult to understand and answer some questions, a person with impaired vision might have trouble filling in a questionnaire, and a person with impaired hearing might have trouble participating in a telephone interview. Thus, survey non-response among older people is likely to be related to poor health (Chatfield, Brayne, & Matthews, 2005; Hardy, Allore, & Studenski, 2009; Manton & Suzman, 1992), and a lower participation rate indicates selection on the basis of better health (Vestergaard et al., 2015). In addition, a review has reported that cognitively impaired individuals are systematically under-represented in most community surveys (Riedel-Heller, Busse, & Angermeyer, 2000). This health-related non-response is systematic and cannot be compared with the
non-response that occurs as a result of choosing not to participate. One aim of this thesis is to examine which characteristics are associated with non-response in surveys of older people and how this non-response can affect survey results.

**Measurement**

As a result of the challenges of coverage and response in surveys of older people, extensive fieldwork is usually required to achieve a representative study population (Lundberg & Thorslund, 1996). However, some of the fieldwork strategies that are necessary might influence measurement error. This section will cover the possible measurement errors that might occur as an effect of two of these fieldwork strategies, the use of indirect interviews and mixed-mode data collection.

Surveying people who are frail, cognitively impaired, or very ill requires adjusted survey methods. For example, how can you get survey information about a frail old person who lives in a nursing home and has severe dementia? Even if you overcome the possible obstacles to get access to people in nursing homes, the person might not be able to participate in an interview. The option in this case is to conduct an indirect interview; that is, an interview with a substitute respondent (a proxy), such as a close relative, who responds on behalf of the person who cannot participate. The use of indirect interviews is central in population studies of older persons, as it makes it possible for even very frail and ill persons to be represented in the results. On the other hand, it is reasonable to question whether a proxy can provide reliable answers and how much the choice of proxy matters.

To answer these questions, it is first necessary to choose a gold standard; that is, to decide which perspective the proxy informant should provide. Is the gold standard the hypothetical answer that would be given by the older person, or is it some objective assessment, such as a clinical assessment from a professional, that the answers of either a proxy respondent or an older person himself should be compared with? The answer to that question affects the direction and the size of the measurement error.

The literature on the use of proxy reports in surveys is extensive. Much of the literature focuses on the level of agreement between an older person’s self-assessment and the proxy assessment with regard to different outcomes, and regards the subject’s response (in this case, the older person’s response) as the gold standard (Kane et al., 2005; Snow, Cook, Lin, Morgan, &
Magaziner, 2005). In general, there is better support for the use of proxy reports on objective questions that are easy to assess, such as physical function, than for more subjective questions, such as psychological outcomes (Magaziner, Bassett, Hebel, & Gruber-Baldini, 1996; Neumann, Araki, & Gutterman, 2000). Furthermore, the direction of the measurement error varies by type of question. Whereas proxies tend to over-report functional limitations (Magaziner, Simonsick, Kashner, & Hebel, 1988; Neumann et al., 2000), they tend to under-report depressive symptoms (Shardell, Alley, Miller, Hicks, & Magaziner, 2012) and rate the subjects global self-rated health (SRH) as better than the subjects rate it (Vuorisalmi, Sarkeala, Hervonen, & Jylhä, 2012). Characteristics of the proxy, as well as his or her relationship with the older person are also factors shown to affect the quality of the answers. The ideal proxy is a person who has lived with the older person for a longer period (Shardell et al., 2012).

The second fieldwork strategy that should be considered in surveys of older people is the use of a mixed-mode data collection. The same interview mode does not fit everyone. For example, although some older people who live alone might enjoy being visited by a new person, a visit by a stranger might be uncomfortable to others. For the latter group of people, a telephone interview might be a better option than a face-to-face interview. Busy proxy respondents might also prefer telephone interviews or postal questionnaires to face-to-face interviews. In contrast, telephone interviews are probably a bad option for respondents with impaired hearing.

As with the indirect interviews, collecting data with a mixed mode approach might affect the measurement error. The extensive literature on the subject typically focuses on differences between the results generated by the various data collection modes. For an informative review of how the quality of data might vary by data collection mode, see Bowling (2005).

The potential measurement error that might occur because of a mixed-mode approach and the use of indirect interviews must be balanced against the error that stems from non-response. The varying modes of data collection in a mixed-mode approach lead to the inclusion of people who otherwise would not have responded to the survey, which reduces non-response error (Fowler Jr et al., 2002; Peytchev, Baxter, & Carley-Baxter, 2009). The use of indirect interviews, for example, is important in avoiding attrition bias in studies of the distribution of cognitive abilities in old age (Weir, Faul, & Langa, 2011).

In general, the measurement error in studies of older people is considered less of a problem than the non-response bias that is the alternative (Corder,
Woodbury, & Manton, 1996). Analyses of measurement error are a complex issue that requires further research, but such analyses were not undertaken in the studies included in this thesis.

2.3 Sociology and the role of the sample survey

History and views

The sample survey has a long tradition in sociology and other social sciences. The first attempt to study social regularities using probabilistic theory was made by the Belgian astronomer Adolphe Quetelet between approximately 1830 and 1860 (Goldthorpe, 2000). Quetelet developed an interest in applying the probabilistic methods he had used as an astronomer to other kinds of data, generating demographic statistics and later “moral statistics,” which included statistics of marriage, suicide, and other social phenomenon. Quetelet thought that in addition to studying the mean of social actions (such as marriage) in different populations, he could use probabilistic methods to study the individual’s propensity for these social actions. He used the term *physique sociale* to describe this new science, which he thought should be the basis of statistical analyses of social data.

Quetelet’s work was controversial. Auguste Comte, the founder of positivist philosophy, strongly doubted that probability theory should or could be used as a valid method in studies of society. Comte had used the term *physique sociale* prior to Quetelet. As a reaction to Quetelet’s use of the term and to distinguish his work from the “the kind of empirical enquiry” that Quetelet was engaged in, Comte coined the term *sociology* (*sociologie*).

Emile Durkheim’s studies of suicide were inspired by Comte’s methods, rather than by Quetelet’s. Although Durkheim is famous for being one of the earliest empirical sociologists, he was skeptical of the use of probabilistic thinking. In his legendary work on rates of suicide in different societies, he used deterministic rather than probabilistic analytical methods. Durkheim and Comte were reluctant to let sociology become dependent on statistics. Instead, their vision was for sociology to develop its own research methods (Goldthorpe, 2000).

Regardless of the skepticism in France about using statistical methods in studies of social actions, probabilistic methods developed in synergy with sociology from the early 1830s to the 1930s, in what has been called “the
probabilistic revolution” (Goldthorpe, 2000). Today, quantitative research methods such as the sample survey are viewed as “the core practice and content” of sociology, not only as techniques used by sociologists and other social scientists (Payne, 2014, p. 85).

During the twentieth century, Paul Lazarsfeld, one of the most influential empirical sociologists, called “the father of the modern empirical sociology” (Jeřábek, 2001), contributed significantly to the development of what he and others labeled “survey research.” His work was not limited to sociology but impacted several disciplines, including unemployment research, public opinion and market research, mass media and communications research, political sociology, and applied sociology (Jeřábek, 2001). Lazarsfeld distinguished survey research from other methodological techniques, such as experimental methods (as in the natural sciences), the use of historical data to study large-scale social change (as in historical sociology), and the use of existing institutional data for social inference (as in Durkheim’s work). According to Lazarsfeld, survey research has three distinctive features. He writes that surveys:

“…are quantitative rather than qualitative; the researcher designs and uses his own data-gathering devices rather than depending on available historical or institutional records; and they concern people’s behavior and attitudes in real-life situations rather than constructed experimental situations.” (Lazarsfeld, 1958, p. 99)

For a comprehensive description of the development and use of social surveys as a tool for social investigation in Britain and the United States, see Bulmer, Bales, and Sklar (1991).

_Cameral sociology vs. “sociology that really matters”_

Sociology is a heterogeneous discipline. It covers a wide range of topics and research areas in which many different research methods are used and different research perspectives are applied. Sociologist Raymond Boudon (1934–2013) suggested that the discipline should be divided into four ideal types: _expressive, cameral, critical_, and _cognitive_ sociology, all of which are different but important (Boudon, 2002).

_Expressive_ (or _aesthetic_) sociology is closer to literature than science. Boudon used the work of Erving Goffman as an example and claims that the strength of Goffman's work lies in the way he wrote rather than in the work's scientific value. Expressive sociologists “are – or were for a while – visible
because they expressed in an original and effective fashion feelings which many people experience in their everyday social lives, such as the feeling that they are manipulated by anonymous forces, or that hypocrisy is a dominant feature of social interaction” (Boudon, 2002, p. 372).

*Cameral (or descriptive)* sociology is characterized by the intention to describe society and to improve the quality of policy-making. Cameral sociology investigates a broad range of topics, such as living conditions, crime, suicide, health inequality, and public opinion. It includes both qualitative studies that aim to describe topics such as the life of marginal people and quantitative studies that aim to answer questions of “how much” or “how many,” often via sample surveys, which are a staple method of this type of sociology. Cameral sociology has a long history but has grown enormously since the second half of the twentieth century, probably as a consequence of a growing demand for social data.

In addition to cameral sociology, Boudon also discusses *critical (or committed)* sociology, in which the main objective is to influence political processes. The political intention of critical sociology is sometimes clearly visible and sometimes less so.

The fourth type of sociology is *cognitive (or scientific)* sociology. Deriving from the view that a good theory is a theory that, in principle at least, can be reduced to a set of uncontroversial statements, cognitive sociology tries to explain puzzling social phenomenon. Boudon considered much of the work of early sociologists, such as Tocqueville, Weber, and Durkheim, to be cognitive sociology and claimed that some of their theories are as valid as theories from the natural sciences. In modern sociology, cognitive sociological theories cover a wide range of subjects, such as crime, social stratification, and collective actions. These theories produce cumulative knowledge and change how we perceive social phenomena.

Although Boudon considered all types of sociology to be important and justified, he viewed cognitive sociology as “the sociology that really matters” and claimed that the cameral orientation had expanded and become the dominant form of sociology at the expense of cognitive sociology (Boudon, 2002).

British sociologist John Goldthorpe supports Boudon’s views that there are four major types of sociology and that cognitive sociology or “sociology as social science” (SSS) is the sociology that really matters. However, in contrast to Bourdon’s view that the discipline of sociology has suffered from the rise of the cameral orientation, Goldthorpe believes that cognitive sociology can gain from a relationship with cameral sociology, in particular re-
Regarding data collection, a field historically pioneered by cameral sociologists:

“... while cameral sociology can certainly be differentiated from SSS in its typical concerns, it does not diverge from SSS in any way or to any extent that would entail fundamental conflict. To the contrary, there seems no major obstacle to SSS and cameral sociology operating in a co-operative and complementary fashion.” (Goldthorpe, 2004, p. 103).

Regardless of the scientific value of cameral sociology or its role in sociology as a discipline, the sample surveys that are characteristic of cameral sociology are widely used and shape our perception of how society works, and as a consequence, how society will develop.

2.3.1 Welfare research

The main objective of many surveys is to measure the welfare and living conditions in a population. Welfare research has strong roots in empirical sociology and is central to the development of a society, as its results can serve as a basis for political decisions and social policy. However, welfare is a concept that can be both difficult to define and to measure. Gross domestic product (GDP) per capita is an aggregated measure that is sometimes used to compare welfare across countries. The GDP measure is limited, though, as it is aggregated and restricted to one dimension. It does not capture what is meant by “welfare” in political debate and in research on welfare states, and the discussion of potential measures of economic and social performance is thus ongoing (Fritzell & Lundberg, 2007; Stiglitz, Sen, & Fitoussi, 2009).

The need to find new welfare measures grew in the 1960s, a period when many industrialized countries underwent intensive urbanization and population concentration that led to a variety of important social and structural consequences. Advocates of a new welfare measure searched for social indicators that could be used to measure of social development the way certain economic indicators were used to measure economic development. Several countries began the process of designing theories and creating tools to enable descriptions of social reality (S. Johansson, 1971, 1979). Sweden was one of the countries that pioneered this new type of welfare research.

The distinctive characteristic of the welfare research initiated in Sweden during the 1960s was the focus on measuring living conditions with regards to people’s “... command over resources in terms of money, possessions, knowledge, psychological and physical energy, social relations, security and
so on by means of which the individual can control and consciously direct her conditions of life” (Fritze & Lundberg, 2007, p. 5). Welfare was defined as “…living conditions in the areas where citizens seek to influence through collective decisions and through commitments in institutional forms, i.e. through politics” (Fritze & Lundberg, 2007, p. 5). This tradition had its roots in Richard Titmuss’ ideas about command over resources and of T.H. Marshall’s writings on the development of social citizenship rights in the 20th century.

In 1965, the government of Sweden established a low-income commission to provide empirical data that could serve as the basis for the discussion about low incomes that was going on in the country. The commission was tasked with carrying out a survey that would provide the basis for a policy to improve the living conditions of low-income earners. The study would include three parts: the first on Swedish people's income, the second on purchasing power, and the third on standard of living (Erikson, 1993). This task resulted in the first Level-of-living Survey, which was carried out in 1968 via interviews of a representative sample of the Swedish adult population (approximately 6500 people). Information gathered via the interviews was supplemented with register data (S. Johansson, 1971). The response rate was 91 percent, a figure that decreased in the following survey waves (Erikson & Åberg, 1987).

The decision to use a survey was not a foregone conclusion; nor was the decision about how to measure the concept of welfare. In the debate that preceded the first Level-of-living Survey, a number of doubts were raised: Would it be possible to measure welfare by simply asking people about their actual living conditions, and could they be trusted to tell the truth? How might possible non-response affect the results (Fritze & Lundberg, 2007)?

The potential effects of non-response on survey results, already briefly discussed in a previous section, are a main topic of this thesis. The question of how best to measure welfare is, however, outside the scope of this thesis. It is nevertheless important to note that Swedish welfare research has also been criticized. One critique concerns the connection between the Swedish welfare research and politics. For instance, Bo Rothstein has discussed welfare researchers’ inability to understand decreasing psychosocial wellbeing in children and young people – a problem that has arisen despite improvement in living conditions. Rothstein suggests that researchers are too dependent on the good will of politicians to do the research that would be necessary to understand the reasons for this problem and refrain from challenging societal reforms that have strong political support (Rothstein, 2009).
2.3.2 Social stratification

One key concept in sociology and in welfare research is social stratification. Resources and life chances are unequally distributed between individuals in a society, and this distribution is not random.

The social structure of a society can be viewed as a map of different positions that individuals can occupy, of places “in a given social structure” (Davis, 1942, p. 309). An individual’s position in the social structure affects her life chances and is associated with multiple outcomes, such as living conditions, health, and mortality. Social structure is more or less stable and is independent of the individuals who hold the positions. As Schumpeter puts it, “each class resembles a hotel … always full, but always of different persons” (Schumpeter, 1955, p. 126).

In his 1942 article, “A conceptual analysis of stratification,” sociologist and demographer Kingsley Davis outlined a theoretical framework for social stratification: Individuals occupy several positions in a given social structure. A stratum consists of many individuals with similar combinations of positions, which give them roughly the same privileges. These strata can be distinguished by different sorts of positions, such as sex, education, social class, and income.

Each of these positions in the different strata is connected to varying resources and life chances. Individuals with less advantageous positions have poorer living conditions than individuals with more advantageous positions, and these poorer living conditions are manifested in worse health and shorter life expectancy. The stratification process starts at birth. Poor childhood conditions are associated with increased mortality in adult life, and this association is mediated by socioeconomic achievements and lifestyle factors during adulthood (Hayward & Gorman, 2004). Social stratification is present in all societies, and a social gradient in health is usually present regardless of which social indicator is used: education, social class, financial resources, or other factors.

Welfare research based on the GDP measure only provides an aggregated picture of the welfare in a country. The new welfare research that developed during the 1960s (which resulted in surveys such as the Level-of-living Survey), on the other hand, provides the opportunity to compare multiple factors related to resources and living conditions between different groups in society and to examine how these conditions change over time. However, to be effective and accurately reflect conditions in society, surveys must include individuals from all strata in the society. The exclusion of disadvantaged groups might result in incorrect conclusions regarding these people’s living
conditions. Hence, analyses of how well individuals from different socioeconomic groups are represented in surveys of older people will be a central focus of this thesis.

2.3.3 Sociology of aging

Understanding the life of older people, the aging process, and societal change in an aging population often requires an approach that connects multiple disciplines such as demography, epidemiology, public health, sociology, and economics. A quick glance at the table of contents of the *Handbook of Aging and the Social Sciences* (Binstock & George, 2011) reveals the multidisciplinary character of aging research, which includes a broad range of research questions and perspectives. The book covers aging and time (e.g., life course perspectives), aging and social structure (e.g., health, diversity, and stratification), social factors and social institutions (e.g., intergenerational transfers, religion, and quality of life), and aging and society (e.g., politics, long-term care, and justice). In addition, the handbooks of aging consist of two more volumes: the *Handbook of the Biology of Aging* and the *Handbook of the Psychology of Aging*.

It follows that it might not be fruitful to distinguish the sociology of aging from other perspectives on aging. Gerontology includes research on aging from several disciplines and has been described as “a multi-disciplinary and (increasingly) interdisciplinary approach drawing upon the behavioral, natural, and social sciences” (Dannefer & Phillipson, 2010, p. xxi). Social gerontology is a research area that includes research on aging from disciplines such as demography, economics, and sociology.

Although they span several disciplines, in principal, quantitative studies of older people emerge from one of two main research perspectives. The first perspective aims to describe the life of older people—for example their living conditions, health, and the presence of formal and informal care. This descriptive (cameral) research is essential for social policy decisions as it provides information about living conditions among older people in general, living conditions among older people from different sociodemographic groups, and information on how these living conditions change over time. The second perspective aims to investigate the causal relationships between different exposures and different outcomes, such as the mechanisms behind the influence of biological, environmental, and/or societal exposures on living conditions, diseases, and the aging process. Although this type of research primarily aims to identify causal mechanisms, in reality such research
studies statistical associations between different points in time, which in turn can imply causality.

Whereas the first perspective can use cross-sectional data (data on individuals from the population who are interviewed at one point in time), the second requires longitudinal data (data on individuals who are followed over time). A life-course approach is often used to study many of the research questions that stem from the second perspective. In the life-course approach, health and living conditions in old age are measured as outcomes of accumulated living conditions and other exposures during the lifespan (Kuh et al., 2014).

Although there are other types of longitudinal data, such as register-based data or retrospective questions in cross-sectional surveys, longitudinal survey data are the best option for answering many research questions. Longitudinal surveys give researchers the ability to analyze inter- and intra-individual change and the onset and duration of different exposures and outcomes. They make it possible to establish the temporal order of events and they minimize the problem of recall bias (Lynn, 2009, pp. 1-18). However, longitudinal surveys face problems that cross-sectional surveys do not. In addition to the challenges of financing and administering a longitudinal survey that runs over many years, sample attrition is of major concern in longitudinal surveys. Sample attrition can be divided into dropout due to non-response and dropout due to mortality. Sample attrition gradually drains individuals from studies, and this attrition can be systematic. Systematic attrition means that the characteristics of the individuals who stay in the sample will differ from the characteristics of those included in the sample at the beginning of the study and/or from those that the survey intends to represent (Deeg, 2002). Systematic attrition in longitudinal surveys of older people is another topic that will be analyzed in this thesis.

**Demographic developments**

Survival has increased dramatically during the last century. A few decades ago, people 80 years and older constituted a small fraction of the population, but in many societies today, this group is a growing sector of the population. In Sweden, almost one of ten people are expected to be 80 years and older in 2050 (Figure 1). People 80 years and older cannot be viewed as a peripheral part of the population, a group of exceptional individuals that have the opportunity to live a little bit longer than normal. Today, most of us can expect
to live a long life and to experience many years of old age (Christensen et al., 2009).

![Figure 2.1](image)

Figure 2.1. Previous and projected proportions of men and women 80 years and older in the Swedish population. Source: Statistics Sweden.

**Social stratification in old age**

Social stratification is also present in old age, and older people are a heterogeneous group (Fors, Lennartsson, & Lundberg, 2008). Both the aging process and living conditions in old age vary between individuals and between different groups in society. One question that has captured the interest of researchers and led to debate in the field of aging research in recent decades is whether health inequality increases or decreases with age.

The “divergence-convergence hypothesis” (also known as the “age as a leveler hypothesis”) suggests that the magnitude of health inequality differs by age. Specifically, the hypothesis suggests that health inequality is greatest in middle age and diminishes in old age. One proposed explanation of the hypothesis is that a delayed onset of health problems among highly educated people leads to declining health inequalities in old age. Another is the “social
welfare effect”: the idea that retirement and the welfare state have changed the social context for individuals in their sixties, which might affect their health and diminish inequalities (Herd, 2006).

The “accumulation hypothesis,” on the other hand, posits that health inequality increases with age. This hypothesis builds on the idea that disadvantages and advantages associated with early-life lifestyle and disparate material and psychosocial resources accumulate over time, causing heterogeneity in health to increase with age (Lynch, 2003; Prus, 2007). Over time, this accumulation might lead to cohort inversion, in which inequality appears to decrease as a result of selective mortality (premature mortality among the most disadvantaged people) (Ferraro, Shippee, & Schafer, 2009; Lynch, 2003).

Another study suggests that health inequality, measured as mortality differences, is stable across ages, but declines with decreasing health—the study found that the transition from poor health to death was not associated with socioeconomic status (SES). Thus, the author suggests that the “age as a leveler hypothesis” should be revised, as it is “illness and not age that levels SES mortality differences” (Hoffmann, 2011, p. 374).

It is difficult to compare the diverging results of studies on health inequality in old age, as researchers used different measures of health, different measures of inequality, and, maybe more importantly, different survey designs. The studies use different modes of data collection. Some use indirect interviews and others do not, and some include people living in institutions whereas others do not. One of the aims of this thesis is to investigate how differences in survey design might impact measures of health inequality in old age.

The development of aging surveys

Although population surveys are commonly undertaken in many countries, the older old have often been neglected in these studies. Traditional level-of-living surveys, including but not limited to the LNU in Sweden, have upper age limits or do not use sampling or fieldwork methods adjusted for older people (Lundberg & Thorslund, 1996). There have been exceptions; as early as the 1950s, some surveys included very old individuals. One example is the Duke Longitudinal Studies of Normal Aging, a series of longitudinal studies of middle-class community-dwelling people 60 to 96 years old that was initiated in 1955. Another is the Alameda County Study, a longitudinal
study based on 20- to 94-year old community residents in a California county (Deeg, 1989).

As with the level-of-living surveys, which developed in response to the need to describe living conditions in the adult population, aging studies have developed in response to the need to describe living conditions among the older old—a need that arose because of the changed structure of the population. In the 1980s and 1990s, when the interest and demand for surveys of the older old began to grow, researchers started to focus more on the methodological challenges involved in these types of surveys. The challenges included problem of locating the older old, as many of the oldest people live in institutions; problems with gaining access to the older individuals; response rates; respondent burden; and data quality (Herzog & Dielman, 1985; Wallace et al., 1992). A number of researchers suggested procedures for conducting surveys among the older old. Suggestions included sampling procedures that allowed for the inclusion of people living in institutions, the use of proxy interviews, limitations on interview time and respondent burden, suggestions about characteristics of the interviewers, and suggestions about interviewer training and support (Lundberg & Thorslund, 1996; Rodgers & Herzog, 1992).

Today, many countries have their own longitudinal aging studies. Conditions for conducting population studies differ between countries, and the aims of the studies differ somewhat. Thus, longitudinal studies vary in sampling procedures, response rates, interview modes, and topics included in the surveys. For reviews, see Erten-Lyons et al. (2012) and Gudex and Lafontune (2000). In to conducting such national surveys, researchers and others have focused much effort and many resources on cross-national research on aging. During the past several decades, initiatives have been started that involve a number of countries. In Europe, two large surveys span many countries and involve older people. One is designed to include individuals over 50, and the other aims to gather data representative of the total adult population.

*The Survey of Health, Ageing and Retirement in Europe (SHARE)*

SHARE is a longitudinal survey of health and living conditions in people 50 years and older in Europe. Baseline data from the first participating countries was collected 2004. Subsequent waves have been carried out 2006 and 2010, and six more waves are planned (up to 2024). The ultimate goal of SHARE
“is to provide high-quality micro-level panel data of economic, social and health factors that accompany and influence ageing processes at the individual and societal levels” (Börsch-Supan et al., 2013, p. 2).

Since SHARE is a multinational survey, the sampling design is complex. At baseline in 2004, SHARE researchers chose sampling frames for each country with care to ensure that each country used the frame that would allow the best possible random selection (Börsch-Supan et al., 2013). Because not all countries have access to population registers, the target population differs between the countries. In about half of the countries, individuals in institutions are not included in the target population. In addition, the response rates vary between the countries, from above 70 percent to below 40 percent at baseline in 2004. SHARE has made a great deal of effort to harmonize the surveys from the different countries. In addition to using a harmonized questionnaire, sampling weights are employed to adjust for the unequal selection probabilities of the target sample units. Proxy reports are allowed (Börsch-Supan & Juerges, 2005).

The European Union Statistics on Income and Living Conditions (EU-SILC)

EU-SILC is a general data source that provides statistical information to Eurostat, the European Union’s statistical office. EU-SILC, which is not primarily focused on older people, collects microdata on income and living conditions from all member states in the European Union and from some non-member countries. Eurostat describes the EU-SILC as “the main source for comparable indicators for monitoring and reporting on living conditions and social cohesion at the EU level” (Eurostat, 2010, p. 3), and its data are used for research and policy-making at both the national and EU levels. Statistics Sweden is responsible for collecting microdata for EU-SILC in Sweden.

Like the sampling procedures in SHARE, the sampling procedures in EU-SILC differ between countries. Whereas the Nordic countries, the Netherlands, and Slovenia sample individuals, other countries use addresses or households as sample units, which means that those living in institutional settings are not included. Proxy reports are allowed, but Eurostat recommends using them sparingly. EU-SILC has been criticized because of unclear reporting regarding which countries use weighting adjustments and poor information about whether any of the weights are also adjusted for non-response (Goedemé, 2013; Iacovou, Kaminska, & Levy, 2012).
2.3.4 Current and future challenges for the sample survey

In this section I will discuss some of the current and future challenges for sample surveys and describe the ongoing discussion about the value of representative sample surveys. Because the challenges are just as relevant to other disciplines such as social epidemiology and public health, the section requires a context broader than sociology.

Three eras of survey research

The development of survey research has been divided into three eras (Groves, 2011). The years between 1930 and 1960 are known as the era of invention. Sample methodology developed during this time. Statisticians were excited about the unbiased estimates and measurable sampling errors offered by probability sampling. Between 1960 and 1990, the era of expansion, the spread of telephones and later of computers enabled new sample frames, interview modes, and kinds of data processing. During this period, the number of surveys in the social sciences and in the private sector greatly increased. Meanwhile, however, non-response became a subject of concern. The period after 1990 has been characterized by increasing non-response rates and higher data collection costs, but also by new technology. Technological development offers new opportunities for data collection, such as Internet panels and information from social media. However, it also creates new obstacles to data collection. In particular, the gradual transition to cellular phones has led to coverage problems and difficulties in reaching people. One reason is that cellular phone numbers are less likely to be included in the centralized registers of telephone numbers traditionally used as sample frames (Groves, 2011).

The crisis (?) of the sample survey

In 2007, Savage and Burrows predicted that the glory days of sample surveys were over. They claimed that the effectiveness of the sample survey is questionable in modern society:
“It is unlikely, we suggest, that in the future the sample survey will be a particularly important research tool, and those sociologists who stake the expertise of their discipline to this method might want to reflect on whether this might leave them exposed to marginalization or even redundancy. … the repertoires of empirical sociology need to be rethought in an age of knowing capitalism.” (Savage & Burrows, 2007, pp. 892, 895)

Their argument is primarily based on three different developments. The first is decreasing response rates. Achieving an acceptable response rate today is usually difficult and resource-demanding. It is common for surveys to have response rates below 50 percent, a development that has given rise to a debate about the value of the results of such surveys. The second development is the advent of new forms of data collection that offer alternative options for collecting information.

Many of our actions, including medical consultations, Google searches, use of social media, and consumption behaviors (such as credit card use) leave electronic tracks. In the Internet era, enormous amounts of data are thus available on most individuals. Some social scientists have suggested that such data could be a more effective and up-to-date alternative for researchers than the information collected through surveys.

The third development is the increased migration that has occurred as a result of globalization. Many people move between nations for work, for safety, or just for the opportunity for a different life. Mass migration in some parts of the world makes national boundaries questionable as natural sample frames (Savage & Burrows, 2007). For a thorough discussion about the future of survey sampling, see Brick (2011).

Whereas the discussion about decreasing response rates and the development of new forms of data collection gives rise to questions about how to best collect data, the discussion about unclear sample frames calls attention to the issue of representativeness; that is, how important is it that the results of a survey are generalizable to a specific population.

**The importance or unimportance of representativeness**

Rothman, Gallacher, & Hatch have asserted that representativeness should not be a goal of scientific studies—only of descriptive studies such as “opinion polls, or for public health applications” (2013, p. 1012). They take the natural sciences as their prototype and equate epidemiology with the search for universal relationships in the natural sciences. Their argument is that in
science, the golden standard should be a controlled experiment in which the study sample is selected to minimize confounding that interferes with the results. For example, when studying the association between health and age, one should not try to select a sample that is representative of a population. Instead, one should try to minimize the diversity between the study units without taking away the variation in the variables that one is studying. In the study about health and age, for instance, the researchers should select individuals that vary somewhat in age but are as much alike as possible with regard to other variables. One example of a study with this type of design is the Nun Study of Aging and Alzheimer's Disease, a longitudinal study of 678 aging Roman Catholic sisters living in convents. The nuns undergo repeated extensive examinations and have also agreed to donate their brains for examination after death. The study provides a unique possibility to study aging and the causes of Alzheimer’s disease because the nuns in the study live their lives under similar environmental and behavioral conditions (Snowdon, 2001).

Of surveys and representative sampling, Rothman, Gallacher, & Hatch write:

“As initial steps, surveys may help to seed hypotheses and give a push toward scientific understanding, but the main road to general statements on nature is through studies that control skillfully for confounding variables and thereby advance our understanding of causal mechanisms. Representative sampling does not take us down that road.” (Rothman et al., 2013, p. 1014)

Not all epidemiologists share the opinion of Rothman, Gallacher, & Hatch. In a commentary, Ebrahim and Davey Smith (2013) state that assessing causal hypotheses is not the only goal of epidemiology; another goal is to provide descriptions of living conditions, health, and health behaviors in the population as the basis for sociopolitical decisions. However, the issue of representativeness also may be crucial in research on causal relationships between exposures and outcomes, because without a representative sample, bias may influence associations. This problem occurs if a factor is associated both with the outcome and the probability of participating in the survey. Rothman and his colleagues aim for a science in which all confounding factors are under control; i.e., in which the causal mechanism that is being studied can be isolated. However, people live in contexts, and it is difficult if not impossible to control for all factors that influence both the outcome and selective response (Ebrahim & Davey Smith, 2013).

Although it originated in epidemiology, the debate started by Rothman and his colleagues is also of interest outside that discipline. There is no wa-
tertightly bulkhead between epidemiology and sociology. Many population surveys are used as databases for researchers from a variety of fields, and many research questions are addressed using a multidisciplinary approach. Like epidemiologists, empirical sociologists are interested both in descriptive analyses of populations and in assessing causal mechanisms (Boudon, 2002; Goldthorpe, 2004).

It is the research question that determines whether and how selective participation will affect study results. In national population surveys, the explicit aim is to achieve a sample representative of the nation's population. In other surveys in which the aim is to investigate a specific mechanism, the need for representativeness is less clear. For example, in research that focuses on intra-individual change, it might not be crucial that the study sample is generalizable to a specific population, but if a factor related to the outcome is associated with selective participation, this selective participation might affect the results.

Imagine a survey of people 80 years and older. Selective participation results in a sample in which those with high levels of education are represented regardless of health and smoking behavior. However, among those with low levels of education, smokers with the poorest health are underrepresented; that is, smokers with low levels of education who participate in the survey consist mainly of those resistant individuals that did not develop poor health.

Assume further that the population consists of equal parts individuals with high levels of education and individuals with low levels of education, that 50 percent of the individuals in each educational group have been smokers, that the prevalence of poor health in those with high levels of education is 30 percent and the prevalence of poor health in those with low levels of education is 60 percent. If 75 percent of the smokers with low levels of education do not participate, this would result in the following errors in the estimates of proportions:

1. The proportion smokers in the population would be underestimated: 41 percent instead of 50 percent.
2. The proportion smokers in the group of people with low levels of education would be underestimated: 29 percent instead of 50 percent.
3. The prevalence of poor health in the population would be underestimated: 35 percent instead of 45 percent.
4. The prevalence of poor health among those with low levels of education would be underestimated: 43 percent instead of 60 percent.

If the research question is whether smoking is associated with poor health, then representativeness may be less crucial. The results from this hypothetical, selected sample would still show an association between smoking and poor health, but the association would be underestimated.

5. The Relative risk (RR) of poor health among smokers (compared with non-smokers) would be underestimated: 1.43 instead of 2.0.

However, this assumes that the association is stronger than the selection. If people with low levels of education have worse health than those with high levels of education regardless of smoking, then the sample might give the false impression that smoking is not harmful.

Finally, if the research question is whether the association between smoking and health is differs by level of education, the results would also be wrong. The result would give the false impression that smoking is more harmful for health among people with high than people with low levels of education.

6. The RR of poor health among people with low levels of education who smoked would be underestimated: 1.25 instead of 2.0.

7. The RR of poor health among people with high levels of education who smoked would be correctly estimated: 2.0.

8. The relative risk ratio of poor health in people with low as compared with people with high levels of education would be underestimated: 0.62 instead of 1.0.

Sometimes the interaction between genetic and environmental factors must be studied to find mechanisms behind poor health and the aging process. One example is the way lifestyle factors and apolipoprotein E (apoE) ε4 allele interact in the development of dementia. Whereas the apoE ε4 allele has been identified as an independent risk factor for dementia, lifestyle factors such as smoking and alcohol consumption are risk factors for dementia primarily among apoE ε4 carriers (Kivipelto et al., 2008). Hence, a selected sample that mainly consists of individuals who tend to have good lifestyle habits might misinterpret the association between apoE ε4 and dementia.
Representativeness in aging studies

As mentioned before, the ability to draw representative survey samples, fieldwork strategies, and response rates vary between surveys. Although this might be less problematic in some studies, SHARE and EU-SILC aim to provide a picture of the life, health, and living conditions of people in Europe. This means that our understanding of the life of older individuals and of the aging process builds on the results of these surveys, and the results of these surveys also serve as a basis for policy makers’ decisions. Although there are problems of representativeness in SHARE and EU-SILC, the data are used for research question in which lack of representativeness might be a problem. I will provide three examples of research reports in which the samples are non-representative. The last two are based on studies that used data from SHARE and EU-SILC.

The first two examples are from studies about “successful aging,” a concept that has garnered a great deal of attention lately in the field of aging research. Multiple factors are relevant to successful aging, including disease, disability, and cognitive and physical functioning (Rowe & Kahn, 2015). These factors are also closely related to living in an institution and to the inability to participate in a survey without the help of a proxy (Lundberg & Thorslund, 1996). In a study from the United States, researchers used data from the Health and Retirement Study, a national study on aging in United States, to estimate the prevalence of successful aging. Only people who lived in the community were included in the analyses, and proxy interviews conducted during the study were excluded. The authors noted this as a limitation and concluded that the prevalence of successful aging would have been lower if people living in institutions and those interviewed by proxy had been included. The authors also concluded that the inclusion of such individuals “may alter the pattern of successful aging observed over time” (McLaughlin, Connell, Heeringa, Li, & Roberts, 2010).

This study was replicated in Europe (Hank, 2011). The authors used baseline data from SHARE and included data on all participants 65 years and older. As in the US study, proxy interviews (3 percent of the interviews) were excluded. However, the article does not describe how the researchers treated data on individuals living in institutions. The SHARE study profile is indicating that people living in institutions at baseline were excluded, but those who participated at baseline are followed when they move to institutions. The sampling procedure differs between the countries participating in
SHARE, though. It is thus impossible to know how the selection to each participating study was done unless it is described in the article. Although the authors described and discussed non-response rates and indicated that the samples were comparable, information about exclusion criteria was lacking (Hank, 2011). Hence, it is impossible to tell whether or not the study's results were likely biased by the exclusion of people living in institutions. If people living in institutions were excluded in some countries, the estimates of the prevalence of poor health might be lower in these countries than in the countries that included people living in institutions.

The third example is from an article describing a study that used EU-SILC data to compare healthy life years at age 50 by country socioeconomic indicators across EU member states (Fouweather et al., 2015). The authors used survey data on the age-specific prevalence of activity limitation from EU-SILC as the health measure. It seems likely that the varied sampling procedures in the different countries (i.e., only some of the countries included people living in institutions) affected the country comparison of health. That is, if living in an institution was associated with poor health, the prevalence of poor health would be lower in those countries where only people living in the community were included in the sample. In addition, if the threshold for living in an institution differed between countries, the size of the bias might differ between countries. The researchers did not mentioned whether or not the survey included indirect interviews.

The results of all these studies were likely affected by the lack of representative samples. In contrast to the US study, however, the two latter studies did not discuss the implication of the exclusions. These implications include the possibility that the estimated levels of poor health might be wrong and that the between-country comparisons of health might be inaccurate.

Although studies of older people have reported problems that may occur as the result of sampling exclusions and non-response (Hardy et al., 2009; Lundberg & Thorslund, 1996; Riedel-Heller et al., 2000; Vestergaard et al., 2015), few studies have estimated the potential size of the problem. This thesis therefore focuses on problems related to sampling exclusions and non-response in studies of older people.
3. Aims

This thesis describes and explores the consequences of estimating prevalence rates and calculating correlations on the basis of non-representative study populations in surveys of older people. It also evaluates methods used to compensate for analyses undertaken with such non-representative data. Four empirical studies are included; these studies address four specific objectives.

3.1 Specific objectives

- To analyze the characteristics of individuals at risk of being underrepresented in cross-sectional studies (Study I and II) and in longitudinal studies (Study III).

- To analyze the systematic errors in health-related prevalence rates that are likely to occur in face-to-face interview surveys (Study I), and postal questionnaire surveys (Study II) of older people.

- To analyze the systematic errors likely to occur in estimates of the associations between education and health in old age (Study IV).

- To analyze whether bias in estimates, caused by non-representative study populations, can be minimized by weighting the data for non-random non-response or exclusion (Study I and Study II).
4. Materials and methods

4.1 The Swedish Level of Living Survey and the Swedish Panel Study of Living Conditions of the Oldest Old

The Swedish Level of Living Survey (LNU) is a nationally representative longitudinal survey of the Swedish adult population. It was initiated in 1968 and is still running. LNU is a multipurpose study, carried out in face-to-face interviews, with the aim of assessing the welfare of the adult population of Sweden. The questionnaires cover a broad range of topics, including living conditions, family situation, working conditions, health, health behaviors, and financial resources. The sample is representative of the Swedish population 15 to 75 years old (the lower age limit was raised to 18 years in 1991). The 1968 survey included approximately 6000 individuals and was drawn from a larger random sample used in the Low Income Investigation of 1967 (Erikson & Åberg, 1987). The original sample has been followed up in 1974, 1981, 1991, 2000 and 2010. For each new wave after 1968, immigrants and younger cohorts have been added to the sample to keep the sample representative of the population. The response rates in LNU have ranged between 91 percent (in 1968) and 77 percent (in 2000). Data from the 2010 wave has not been used in this thesis.

The Swedish Panel Study of Living Conditions of the Oldest Old (SWEOLD) was initiated 1992 as an extension of LNU. The sample included all individuals previously interviewed in LNU who had passed the age limit of 75 years. In the 2002, 2004, and 2011 waves, the SWEOLD inclusion criteria were extended to include those in the LNU sample regardless of whether or not they had been interviewed in LNU or not. In the 2004 wave, however, the lower age limit was 69. The response rates in SWEOLD have ranged between 95 percent in 1992 and 84 percent in 2002 and 2011.

Together, LNU and SWEOLD provide a longitudinal database that spans a period of 43 years. Except for the 2004 wave of SWEOLD, which used telephone interviews, both LNU and SWEOLD primarily used face-to-face
interviews to gather data. In SWEOLD, indirect interviews were used when the older person was too frail or cognitively impaired to participate in an interview. If possible, close relatives or health care personnel who knew the respondent well were used as informants. In 2004 and 2011, telephone interviews and postal questionnaires were also offered as an alternative to a face-to-face interview in SWEOLD. For a more thorough description of LNU and SWEOLD, see Fritzell & Lundberg (2007) and Lennartsson et al. (2014). Studies I and IV use data from SWEOLD. Study III uses data from a cohort followed in both LNU and SWEOLD.

Figure 4.1. A schematic picture of the Swedish Level of Living Survey (LNU) and the Swedish Panel Study of Living Conditions of the Oldest Old (SWEOLD).
4.2 The Stockholm Public Health Survey

The Stockholm Public Health Survey (SPHS) is a population-based questionnaire survey of the people living in the Stockholm County. SPHS was initiated by the Stockholm County Council and is conducted by Statistics Sweden in collaboration with researchers at Karolinska Institutet. The survey has been conducted in 2002, 2006, and 2010. It includes both cross-sectional representative samples for each year and subsamples that are followed over time. The aim of the survey is surveillance of health and risk factors, and the data are used for research, policy-making, and health care planning. The postal questionnaire includes questions on health status, lifestyle, living conditions, and social relations (Svensson et al., 2012).

In 2010, the upper age limit of 84 was removed. To improve the inclusion of the older population, a special questionnaire was used for people 65 years and older. In addition, researchers sent an introductory letter prior to sending the questionnaire. The fieldwork also included three reminders. As the sampling frame was all people registered as residents of Stockholm County, people living in institutions were also included in the sample. However, no information is available about the number of participants who lived in institutions. The total response rate to the questionnaire sent to those 65 years and older in 2010 was 74 percent. Of the respondents, 10.1 percent reported that they had help with filling out the questionnaire. The 2010 survey data on those 65 years and older were used in Study II.

4.3 Registers

In Sweden, like the other Nordic countries, all people have a unique personal identification number. This system of identification numbers makes it possible to maintain relatively comprehensive population registers from which individual-level data can be sampled, to link the data from different registers at the individual level, and to link survey data to information from the registers (also at the individual level). After linkage all data are rendered anonymous by replacing the personal numbers with a serial number. Three national registers were used in this thesis.

The Register of the Total Population (RTB) is register that includes fundamental demographic data and is mainly used to produce statistics relating to population size and composition. It serves as the basis of statistics on topics such as migration, births, deaths, marriages, and divorces. As the register
covers all individuals registered as resident in Sweden, the RTB is used in population studies, either to retrieve a sample of individuals or to obtain supplemental data (Statistics Sweden).

The Cause of Death Register (CDR) covers all deceased persons registered as residents of Sweden at the time of death, regardless of whether the death occurred in Sweden or abroad. The CDR contains data from 1961 onward and is updated every year. The register is of high quality. There is little missing information on actual deaths; those problems that exist mainly concern the cause of death. (NBHW, 2015a).

The National Patient Register (NPR) gathers information from all 21 county councils in Sweden. It covers all inpatient and outpatient care, including information about date of admission/discharge, main diagnosis, and secondary diagnosis. Although the quality of the register is considered high, there is some drop-out in reporting (less than 1 percent) and some information on variables is missing (less than 3 percent). From 1987 onward, the NPR includes information on all inpatient care. From 2001 onward, it includes data on outpatient visits, including day surgery and psychiatric care provided by both private and public caregivers. Primary care is not yet covered by the NPR. The NPR is updated once a year (NBHW, 2015b).

4.4 Measurements

Demographics

Information about sex and age/birth year was used in all four studies; the variables were used as covariates in studies I and IV and as stratifying variables in studies II and IV. In Study III, the proportion men and women, mean birth year, and the proportion people in the sample who were married/cohabitating were used as dependent variables (outcomes). Information about civil status in Study III was self-reported for those interviewed at baseline. Information about civil status was also available on non-responders because the LNU 1968 sample was drawn from the Low Income Investigation from 1967, which included data on civil status.
Socioeconomic indicators

Two socioeconomic indicators were used in this thesis, social class and education. Social class is a socioeconomic indicator that positions an individual in a stratified social system on the basis of occupation. The social class measure used was the Swedish socioeconomic classification schema (SEI), which classifies occupations based on whether a person is self-employed or an employee and on the educational requirements that are typical for the occupation (Andersson, Erikson, & Wärneryd, 1981). The SEI schema is closely related to the commonly used Erikson, Goldthorpe, and Portocarero (EGP) class schema (Bihagen, Nermo, & Erikson, 2010; Erikson & Goldthorpe, 1992).

Education is a key stratifying variable. In older cohorts, educational achievement is rather stable across the adult life course. It has been suggested that analyses that use education as a socioeconomic indicator are thus less prone to problems of reverse causality than analyses conducted with other socioeconomic indicators, such as income (Herd, 2006).

LNU was the primary source of information on occupation and education in this thesis. Information on the occupation and education of LNU non-responders was available from the previously mentioned Low Income Investigation of 1967. This included data on occupation and education.

Total years of education were used as a covariate in Study I. Level of education (dichotomized into education beyond compulsory school/no education beyond compulsory school) was used in two of the studies. The proportion of people in the sample with low education was used as the dependent variable in Study III. In Study IV, low education was used as the independent variable. In Study III, social class, measured as percentage of manual workers, was used as the dependent variable.

Health and mortality

In Study I, three measures of health problems were used as dependent variables: diseases/symptoms, impaired mobility, and disability.

Diseases/symptoms were measured as the magnitude of self-reported diseases/symptoms assessed by a battery of 22 indicators of common diseases and symptoms in the questionnaire. All indicators were based on the question Have you had any of the following illnesses or ailments during the past 12 months? Response alternative were “no” (0 points), “yes, mild” (1 point),
and “yes, severe” (3 points). To capture both the number and the degree of poor health, we summed the total score for each respondent.

Impaired mobility was measured as self-reported difficulty with walking 100 m fairly briskly, walking up and down stairs, and getting up from a chair without using the armrests. Each respondent was given 1 point for each reported difficulty. Impaired mobility was used both as an index and as a dichotomous measure (have two or more difficulties or does not have two or more difficulties).

Disability was measured as self-reported problems with Activities of Daily Living (ADL) (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963). The index included the ability to eat, use the toilet, dress/undress, get up/go to bed, cut one’s toenails, and wash one’s hair. The response alternatives were: “Yes, completely by myself” (0 points), Yes, with help” (1 point), and “No, not at all” (2 points). We created an index by calculating the total score for each respondent. Additionally, we created a dichotomized ADL measure: having difficulty with (needing help) with at least three ADL/not needing help with at least three ADL. The reason for including cutting toenails, which is not commonly used in ADL indices, was to capture a wide range of disabilities. Difficulty cutting one’s toenails is considered an early stage of disability, whereas losing other abilities, such as the ability to eat without help, comes later (Kingston et al., 2012). The cut-off point of at least three ADLs was chosen to capture severe disability; i.e., having problem in several ADL domains.

The fourth dependent variable in Study I was mortality risk. Participants’ mortality risk was followed up for approximately 2.5 years, from the date of their interview until 18 April 2005. For the non-respondents, the median interview date was chosen as the start date.

In Study II, we were obligated to rely on register data to compare morbidity and mortality in the target population and the responders in the SPHS. Three different indicators of morbidity were used in the study: hospital admission, mean number of days in hospital, and mean number of diagnoses. Mortality was measured as the proportion that died in the year after the data collection period. The follow-up period for mortality was limited to 13 months because we did not have access to mortality data for a longer period.

In Study IV, disability and mortality were used as dependent variables. Disability was defined as a dichotomous measure of self-reported problems with ADL. In contrast to Study I, cutting toenails was not included in the measure. Use of a cut-off point of at least two ADL limitations rendered the ADL measure approximately comparable with the dichotomous disability
measure in Study I. Mortality was followed for one and two years from the date of the interview.

Subgroup division

In studies I & IV, the respondents were stratified into three subgroups: Restricted response group, proxy interviewed group, and institutionalized group. The restricted response group consisted of people who lived in the community and who had been interviewed without the help of a proxy. The proxy interviewed group consisted of those for whom a proxy answered some questions (mixed interview) or all questions (indirect interview). We decided to include mixed interviews in the proxy interviewed group because non-response would likely have been the alternative to either a mixed or an indirect interview. The institutionalized group included those who lived in some form of residential situation intended specifically for older people, such as nursing homes. Studies I and IV use slightly different definitions of “institution.” In Study I, the institutions comprised both blocks of service flats and needs-assessed forms of housing such as nursing homes, whereas in Study IV, they comprised only needs-assessed forms of housing. This is because by 2011, the year the data for Study IV were collected, service flats had become rare in Sweden. Instead a new form of housing had emerged, special apartments for people 55 years and older, and these types of apartments cannot be considered institutions.

4.5 Statistical analyses

The analyses were performed with SPSS 17 (Study I), SAS 9.3 (Study II), and STATA 13 (Studies III & IV).

Study I

Subgroup comparisons of characteristics and health were performed with $\chi^2$-test for proportions and t-test (independent samples) for means. Differences in mortality risk were analyzed by Cox proportional hazard regression. Both crude and adjusted models were used. The assumption of proportionality was tested by the “log minus log” plot in SPSS. The lines of the different groups were parallel and did not cross in any of the models. Hence, the proportion-
ality assumption was considered fulfilled. Finally, to test whether weighting adjustment would correct for the bias caused by subgroup exclusion, we post-stratified the restricted response group by age and sex to correspond to the full sample. That is, individuals in underrepresented strata (stratified by sex and 1-year age groups) were given the weight required to make the sample correspond to the original sample regarding sex and age.

Study II
The indicators of disability and mortality were calculated as proportions and means in the target population (by sex divided into 5-year age groups). The same calculations were made for the respondents; 95 percent confidence intervals were used. This provided the opportunity to evaluate whether the respondent estimate differed statistically significantly from the target population; that is, to ascertain whether or not the confidence interval covered the population value. Analyses were performed with and without a weighting variable that had been constructed by Statistics Sweden, the government agency that administered the questionnaire. The weighting variable was constructed on the basis of available auxiliary variables from different national registers, such as sex, age, country of birth, civil status, income, educational level, sickness allowance, area of residence, and the co-variation of these variables with the survey data.

Study III
A graphical approach was used in the third study to show how a longitudinal sample changes over time because of selective mortality and dropout. We recalculated the proportions of a variety of baseline characteristics at each follow-up i) after mortality and ii) after dropout.

Study IV
Absolute and relative differences in prevalence of disability and mortality between those with low and those with high levels of education were calculated by subgroups (year of interview, age, and sex). Differences were then tested by crude and adjusted logistic regressions. Because it can be problematic to compare odds ratios of similar models across groups (Mood, 2010), results were also calculated with margins, presented as relative proportions. The same calculations were performed i) with the restricted response group; ii) including both the restrict response group and those interviewed by proxy; and iii), including the restricted response group, those interviewed by proxy, and those living in institutions. Sample weights were used to correct for the
different inclusion probabilities. Cluster-correlated robust estimate of variance was used in the analyses because some individuals were included twice (Hardin & Hilbe, 2012).
5. Overview of the studies

In the first two studies, we explored who is likely to be missing in surveys of older people, the consequences that the exclusion of these people has on prevalence rates, and whether it is possible to adjust for a non-representative survey sample by non-response weighting adjustment. In the third study, we investigated how a longitudinal sample changes over time because of non-participation and selective mortality. In the fourth study, we explored whether associations might also be affected by the systematic non-participation that is likely to occur in surveys of older people.
5.1 Study 1

In the first study, we used the 2002 wave of SWEOLD to investigate a) who would be missing from a survey that excluded those living in institutions and that did not use indirect interviews, b) how the prevalence rates would be affected by these exclusions, and c) whether post-stratifying the data by sex and age (weighting adjustment) would correct for any systematic measurement error.

Results showed that those living in an institution and those interviewed by proxy were significantly older, had lower levels of education, and were more likely to be women than those who lived at home and were interviewed directly. They also had significantly more diseases and symptoms, mobility problems, and ADL limitations, and their mortality risk was four times higher than that of the other group.

Exclusion of people living in institutions and non-use of proxy interviews led to underestimated prevalence rates for all the studied health outcomes. Weighting the data by sex and age did not correct for any of these underestimations.

Figure 5.1. Estimated proportion of individuals 77 years and older with ADL limitations, by the use or non-use of proxy interviews and the inclusion or exclusion of institutionalized individuals.
5.2 Study 2

In the second study, we compared mortality and hospitalization rates in those who responded to a postal questionnaire with mortality and hospitalization rates in the target population. We also tested whether a weighting variable created with a number of auxiliary variables (produced by the three steps of weighting: base weight, non-response weight, and post-stratification) could correct for the differences.

The results showed that those who took part in the survey represented the target population fairly well regarding morbidity in the age groups between 65 and 80 and regarding mortality among men. After age 80, morbidity in participants was slightly lower than in the target population. Among women in all age groups, mortality was significantly lower in participants than in the target population. Weighting the data only marginally improved the underestimated results.

Figure 5.2. Relative mortality risk in participating women compared with women in the population (13 months of follow-up) by age group.
5.3 Study 3

In the third study, we followed a longitudinal cohort sample for 43 years, from LNU 1968 to SWEOLD 2011. By recalculating baseline characteristics at each follow-up we investigated how the sample changed after a) selective mortality and b) survey non-participation. Because numerous studies are based on complete data, we created a restricted response group that only included those respondents who participated in all survey waves up to and including each wave.

The results showed that over time, selective mortality changed the sample by successively increasing the proportion of women, people with non-manual occupations, younger birth cohorts, people with more than a compulsory education, and people who were married at baseline. Sample selection related to sociodemographic variables (age, sex, and civil status) was primarily driven by selective mortality, whereas sample selection related to socioeconomic variables (social class and education) was substantially compounded by selective survey participation.

![Figure 5.3. Accumulated attrition in the 1914-1934 cohorts in the LNU/SWEOLD sample between 1968 and 2011.](image)
5.4 Study 4

In the fourth study, we used SWEOLD 2004 and 2011 to investigate if estimates of health inequality in old age are affected by the systematic non-participation that is likely to occur in surveys of old individuals. To achieve this aim, we compared absolute and relative differences in disability and mortality between individuals with compulsory education (low education) and above (high education), depending on the use or non-use of indirect interviews and the inclusion or exclusion of institutionalized individuals.

The results showed that both absolute and relative health inequality would have been underestimated in a survey that excluded those living in institutions and/or did not use indirect interviews. The same patterns were found among men and women and regardless of mode of data collection (telephone or face-to-face). The degree of underestimation was lower among those 85 years and older, compared with those 69-84 years old.

![Figure 5.4. Estimated relative risk for ADL limitations in those with a low level of education compared with those with a high level of education, by the use or non-use of proxy interviews and the inclusion or exclusion of institutionalized individuals.](image-url)
6. Discussion

In keeping with previous research, this thesis shows that people with low socioeconomic positions tend to be less represented in both cross-sectional and longitudinal surveys of older people than people with high socioeconomic positions. In addition, the oldest age groups, women, and those with poorest health tend to be less represented than younger age groups, men, and those with better health, a result found both in face-to-face surveys and surveys that use postal questionnaires. The results of this thesis add to previous knowledge by estimating the magnitude of bias that the selection into the samples might cause both in health-related prevalence rates and in associations between socioeconomic factors and health. The results also give new insight into the difficulties of compensating for a non-representative study sample by weighting adjustment. Finally, this thesis provides a better understanding of the selection processes that occur over time in a longitudinal sample: whereas the selection over time related to sociodemographic factors (sex, birth year, and civil status) is primarily driven by mortality, the sample selection caused by higher mortality in those with low socioeconomic positions tends to be compounded by the lower response rates in people from these groups.

The results in this thesis rely to great extent on the SWEOLD study. The advantages of SWEOLD are its use of probability sampling, inclusion of those in institution, use of fieldwork methods adjusted to older persons, and relatively high response rates. However, SWEOLD is a small-scale survey; one of its disadvantages is thus low statistical power. In addition, although SWEOLD is a study with good data, these data do not represent true population values. As with all survey data, there are non-response errors and measurement errors. Although the non-response rates in SWEOLD are relatively low, it is still likely that they affect the results to some extent. A higher response rate does not necessarily lead to a less selected response group, but we know that the health-related non-response rate in SWEOLD has been minimized by the use of proxy interviews and the extensive field work. However, if SWEOLD was too successful in recruiting people with poor health, the proportion of healthy people in the non-response group might be
higher than in the population. Several studies suggest that extra fieldwork efforts to include hard-to-recruit individuals do not lead to biased results (Fuchs, Bossert, & Stukowski, 2013; Hox, de Leeuw, & Chang, 2012; Olson, 2013), but those studies focused primarily on number of contact attempts rather than on specific fieldwork methods used to recruit older individuals with poor health. More research on this topic would be welcome.

In addition, it is difficult to evaluate measurement errors that are present because of the use of indirect interviews and mixed-mode data collection. We know that different modes of data collection tend to yield different results (Bowling, 2005) and that there is a risk of bias in proxy reports (Neumann et al., 2000). However, research has shown that for disability, one of the main outcomes in this thesis, all types of survey data (including self-reported, proxy reported, face-to-face, and telephone interview data) yield lower rates of disability than clinical assessment (Walsh & Khatutsky, 2007). Hence, if clinical assessment is considered the gold standard, the mixed-mode approach and use of indirect interviews should be considered less problematic than the bias that would have resulted from the non-response that would have been the alternative.

Sociology has a long tradition of using population-based surveys. Such surveys are powerful tools for describing living conditions in a given population and between populations. The results of population-based surveys are valuable to policymakers and researchers from a variety of disciplines. However, who these surveys represent is crucial, as it will affect the cross-sectional and longitudinal information gathered about individuals’ health and living conditions as well as comparisons between different socio-demographic groups.

Older individuals are a rapidly expanding part of the world’s population, both in absolute and relative numbers. This demographic shift will have economic consequences for society and will affect health and social care service planning. Only if all groups of older people, regardless of health status, living situation, education, and income are represented in surveys can the impact of current political initiatives and legislation be accurately evaluated. As long as we build our understanding of society with the help of survey data, it will make a difference whether or not the results reflect the society or just a privileged fraction of it.

Five principal conclusions can be drawn from this thesis: First, the people in the oldest age groups, women, those with low socioeconomic positions, and those with poorest health tend to be less represented in surveys of older
people than people in younger age groups, men, those with high socio-economic positions, and those with better health.

Second, systematic non-participation will lead to underestimation of health-related prevalence rates—underestimation that should not be ignored. The underestimation of factors related to the reason for non-participation, such as disability, might be severe. If we are aware of the problem of underestimation of prevalence rates, we might be able to adjust for it by assuming that health problems among the oldest old are systematically underestimated in most studies. The level of underestimation might be appraised with information about the groups that are absent from the survey. For instance, information about the current thresholds for institutional care and the number of individuals in such care might help indicate the degree of underestimation.

Third, systematic non-participation in studies of older people might also lead to biased associations—biases for which it is probably difficult to compensate. The results of this thesis demonstrate that estimates of social health inequality in old age might be underestimated, or undetected, because of the systematic non-response that is likely to occur in surveys of older people. Thus, selected survey samples might not only be problematic for cameral sociology, but also for research categorized as cognitive sociology. One theoretical implication of selective samples might be results that lead to the impression that health inequality diminishes with age.

Fourth, the results of this thesis do not support the assumption that weighting adjustment compensates for non-response or non-coverage in studies of older people. None of the tested weighting methods improved the population estimates in any substantial way. On the other hand, they did not bias the estimates in any substantial way either, except for an increased variance that resulted in broader confidence intervals. This result does not mean that it is impossible to correct for selections into the sample by weighting adjustment, but it clearly demonstrates that it is a challenge to solve the issue of health-related non-response in aging studies with weighting adjustment. The assumption behind non-response weighting of cross-sectional data is that the proportion of the outcome is the same among respondents and non-respondents in each stratum used in the weighting procedure. A problem thus arises when there is a difference in each stratum that correlates with the probability of participating in the survey. Assuming that poorly done field-work can be compensated by retrospectively weighting data is not the way forward. Instead, we must learn more about the specific biases that occur in
surveys of very old people and use this information to develop better methods to improve the precision of our estimates.

Finally, this thesis shows that selective mortality and selective participation affect the composition of a longitudinal sample over time differently for different factors. As previously noted, selection related to sociodemographic factors is mainly driven by selective mortality, whereas selection related socioeconomic position is driven both by selective mortality and low response rates among people with lower socioeconomic positions. Thus, in longitudinal studies in which socioeconomic position is an independent variable, selective mortality and selective participation can work together to compound the problem of underestimated associations.

6.1 Concluding remarks

There is no doubt that falling response rates and higher data-collection costs pose real challenges to the future of sample surveys (Groves, 2011). It is also clear that new technology has opened up new opportunities for data collection (Savage & Burrows, 2007). However, studies’ research questions must guide how surveys are conducted and the extent to which other data sources can replace or complement survey data. As Groves put it:

“Survey research is not dying; it is changing. The self-report sample survey provides insights into the thoughts, aspirations, and behaviors of large populations in ways that data tracking naturally occurring behaviors are unlikely ever to capture. The survey method has strengths and deficits that are reflections of the society that it measures; the very act of speaking candidly to a stranger is governed by norms that can and do change. Survey research has always and must always adapt to those changes.” (Groves, 2011, p. 870).

The debate about representativeness is important and must take into account the research questions we would like to answer. Although not all studies need to include a sample that is representative of a specific population, but if there is selection into the sample, there is always a risk that the results of the study will underestimate or even miss associations. We must bear this in mind when we conduct surveys and when we interpret their results. Many surveys are multipurpose studies, and as a consequence, are not always tailored to the specific analyses conducted using the data. The results of this thesis emphasize the importance of being aware of the potential biases that might have occurred during data collection. Those who conduct surveys
must thus be very explicit about their methods, and researchers who use survey data must publish clear and detailed explanations of how the data were gathered.

Finally, future aging research cannot exclusively rely on large-scale surveys in which response rates are subordinated to statistical power. The future must also include high-quality surveys in which sampling procedures ensure the inclusion of representative samples and in which enough time and effort are spent to minimize systematic non-response.
7. Acknowledgements

Six years ago (approx.) I accidently happened to be in the right place at the right time. Instead of finding a supervisor for a bachelor’s thesis in statistics, I met people from ARC who told me that aging research is the thing. That was the start of a six-year-long journey that has culminated in this thesis. I have had great companions and support along the way, and I am grateful to all of you.

My greatest thanks go to my main supervisor Carin Lennartsson for traveling this road with me from the start: first convincing me to apply to become a research assistant at ARC, then supervising my master’s thesis, and finally agreeing to be my supervisor. Your constant support has been tremendously important to me. I never made it easy for you. You had to struggle with my resistance to writing like a sociologist and with my inability to say no to new projects. You kept me on the right track, protected me from working part time, and forced me to take some time off now and then. You are a sjutusan till first-time supervisor! I also appreciate our travels together, always working in the air, most of the time sharing a room, and always appreciating a good beer.

My next thanks go to Mats Thorslund, my co-supervisor. Among numerous (!) applicants, you chose a mother of three (with some statistical skills). Your endless faith in my ability to manage everything has meant a huge amount to me. I will always be grateful for your honest criticism, your never-fading enthusiasm, and of course for introducing me to Koko’s party. This “byrå” would not have been the same without you.

I would also like to thank my great co-authors. It has been a privilege to work with you:

Stefan Fors, co-author, a great colleague and former officemate, for your intellectual capacity, great lunch company, all the interesting conversations at pubs, and for guiding me through Amsterdam.
Neda Agahi, co-author, a great colleague and travel companion, for your generosity, your fantastic time-optimism, and for introducing me to alcohol research. Looking forward to the marshmallows!

Karin Modig, co-author at KI, a pleasure to work with and always supportive, straightforward, and quick to respond. There is something exciting about email conversations late Saturday evenings or during Christmastime.

Pingo, my soon to be co-author and a very special colleague, for all methodological discussions and support, for the late nights at Stringfellows, and for just being you!

Kozma, my co-author outside this thesis, for all the lunches, the lively discussions, and frozen margaritas in San Diego.

I am also grateful to all my coworkers at ARC for providing a stimulating and inclusive research climate. Extra thanks go to the administrative staffs of ARC and Äldrecenterum for your invaluable work and your patient support. I would like to give a special thanks to Kim Kane, who with endless patience corrected my English, for never giving up on the idea that it still might improve.

I have had the privilege of being a member of the Social gerontology group at ARC. I am grateful to all the former and present members of this world-leading group, of which some deserve a special mention: Helen Long, who stood outside ARC and rang the doorbell with me on our first day. I miss sharing a room with you and I miss having you in the same country, but I still feel your friendship and support. Josephine Heap, for all the conferences we have successfully attended since our rookie days and for being a caring and good friend. A special thanks to you and Martin Heap (and some supportive social gerontologists) for the fantastic cover illustration. Finally, Jonas Wastesson, office-mate and friend, for your intellectual input, encouragement and support, and wise advice on questions from child care to statistics. You are the ultimate PhD journey companion.

I am also indebted to the Department of Sociology at Stockholm University for generously including me in their working environment and for taking care of me during this time. Special thanks to Maria Bagger-Sjöbäck for all her support and for giving me the opportunity to teach.

I am also grateful to Michael Gähler for scrutinizing this work not only once but twice.
My final thanks go to my family and friends. I have been privileged to have fantastic friends, some of you so close that the word friend is not enough. Thank you for making sure my life consists of more than work. To my family, who always supports and encourages me to do what makes me happy in life. To Gerth and Anita, for your endless help with building houses, cooking dinner, driving kids, and helping them doing their homework. This family would have trouble existing without you! To Kalle, for being my life companion, struggling together side by side. Together we have created a family and a life that I am grateful for every day. Finally, Simon, Jakob, and Saga, life would have no meaning without you. Jag älskar er till månen och tillbaka.

A last thanks to those who can’t be with me today.

And to the pomodoro.

Susanne Kelfve, October 2015
8. References


