Empirical Essays on Health and Aging

Maja Weemes Grøtting

Avhandling for graden philosophiae doctor (ph.d.)
Universitetet i Bergen
2019

UNIVERSITETET I BERGEN

Empirical Essays on Health and Aging

Maja Weemes Grøtting

Avhandling for graden philosophiae doctor (ph.d.)
ved Universitetet i Bergen

Disputasdato: 29.05.2019

År: 2019

Tittel: Empirical Essays on Health and Aging

Navn: Maja Weemes Grøtting

Trykk: Skipnes Kommunikasjon / Universitetet i Bergen

Empirical Essays on Health and Aging

Maja Weemes Grøtting

Dissertation for the degree of philosophiae doctor (PhD)
at the University of Bergen

2019

Acknowledgements

This thesis was written while I was employed as a doctoral research fellow at Norwegian Social Research at Oslo Metropolitan University. The dissertation was completed and submitted to the Department of Economics at the University of Bergen (UiB). I would like to thank both institutions for hosting me. I would also like to thank the Norwegian Research Council and NordForsk for financing the research in this thesis.

I owe my deepest gratitude to my supervisors for all their guidance and support. I am especially grateful to Astrid L. Grasdal, my principal supervisor, for always keeping the paramount objective of my research in focus, and for sharing valuable knowledge of health economics and the Norwegian health care sector. I value the inspiring conversations we have had. I am also indebted to Arild Aakvik, my co-supervisor, for providing invaluable insights into econometric techniques and for his attention to details. I would also like to express my sincere gratitude to my co-supervisor and project manager, Marijke Veenstra, for introducing me to social gerontology and inviting me to partake in the ageing research community, facilitating my research stay at Columbia University, and for all the guidance and mentorship during my work with this thesis.

I am grateful to my colleagues at Norwegian Social Research for contributing to an inclusive, supportive, and fun work environment, and for providing constructive feedback on my work. I am especially grateful to my office companion and partner in crime, Gøril Løset, for endless conversations – because procrastination is key in a creative process, and because life becomes more fun that way. I also wish to thank the rest of my lunch crew, Thomas, Jardar, and Hanna, for all your advise and for helping me clear my head at least once every day. This thesis owes a substantial debt to my two co-authors, Signe and Otto. I have learned so much from our collaborations, on so many levels. Thank you for your patience, drive, and for sharing your knowledge. I am also thankful to the rest of the staff at the Department of Economics at UiB for providing a particularly welcoming atmosphere.

I spent January 2015 to May 2016 at Columbia University, New York. I wish to thank Jonas Hjort for making this possible by introducing me to Matthew Neidell. I am grateful to Matt for hosting me and for the constructive feedback on my work. The output I got from my research stay at Columbia has been invaluable.

Finally, I wish to thank my friends and family for their endless support and well-needed distractions along the way. A special thanks to mum and Gina for the daily phone calls while walking to and from work, to dad, my mother in-law, Sissel, and Ingrid for caring for Hedda so I could return to work after my parental leave, to Ida for your prudence, and to Hege for moral support and constructive comments on my work. The process of writing this thesis has been all-consuming at times. Erling, your unconditional support has been beyond words. I am amazed by your limitless capacity. After everything you do at work and in our home, you still have the energy to lovingly care for (and sleep train) our daughter, before sitting down with me to discuss and comment on my work. This thesis is indebted to you, and I am forever grateful.

Oslo, April 2019
Maja W. Grøtting

Abstract

This thesis is composed of four chapters: One introductory chapter and three chapters that each include one of the three essays that make up the main body of the thesis. Common features of the essays are empirical assessments of health in later life and the econometrical challenges caused by selection–either in the form of people self-selecting into "treatment" or in the form of sample-selection in the data applied.

The first essay, co-authored with Signe A. Abrahamsen, is concerned with how increased access to publicly provided eldercare affects the immediate and long-term health of informal caregivers. The evidence of the causal effects of care responsibilities on health is limited, especially for long-term outcomes. In this essay, we estimate long-term effects of a formal care expansion for the elderly on the health of their middle-aged daughters. We exploit a reform in the federal funding of formal care for Norwegian municipalities that caused a greater expansion of home care provision in municipalities that initially had lower coverage rates. We find that expanding formal care for the elderly reduced sickness absence for single-child daughters of lone parents in the short run, and that the decrease seems to be driven by absences due to musculoskeletal disorders and psychological disorders. When assessing long-term health outcomes for the same group of daughters, we find that, overall, long-term health is not affected.

The second essay, co-authored with Otto S. Lillebø, is concerned with the health effects of retirement and the potential heterogeneity in these effects according to socioeconomic status. Using a local randomized experiment that arises from the statutory retirement age in Norway, we estimate causal effects of retirement on health across socioeconomic status. We apply health information from full population administrative registers and from survey data to investigate the effects of retirement on acute hospital admissions, mortality, and a composite physical health score. Our results show that retirement has a positive effect on physical health, especially for individuals with low socioeconomic status. We find no retirement effects on acute hospitalizations or mortality in general. However, our results suggest that retirement leads to reduced likelihood of hospitalizations for individuals with low socioeconomic status. Finally, we show that the positive health effects are driven by reduced pain and reduced health limitations in conducting daily activities. Our findings highlight heterogeneity in the health effects across socioeconomic status and across subjective and objective measures of health.

The third essay analyzes the extent and consequences of health-related non-response in survey data, defined as either initial non-response or attrition across waves. The first

part of the paper provides a thorough description of the initial non-response and attrition between waves in the Norwegian Study on Life Course, Aging, and generation (NorLAG), a typical survey sample. In the second part, the potential bias from sample selection is examined using econometric analyses of educational differences in health across age as a case study. This is an empirical application that is believed to be particularly prone to bias from health selection, as non-response correlates with both health, education, and age. I make use of a sample of Norwegian older adults and register data that cover the full population to make various assumptions about the selection process and find that using population weights, inverse probability weights, the Heckman selection model, and a Copula selection model does not produce substantially different estimates. Across all models, educational differences in health converge in old age, as is commonly found in analyses that use survey data. This convergence is also found using the register data. Taken together, these results suggest that survey data are well-suited to describe patterns of health inequalities as they manifest in the population. Finally, the register data reveal that healthy survivor bias among the low educated is an important factor driving the convergence in health inequalities in old age.

Contents

Chapter 1:

Introduction

Introduction

Population aging is a global phenomenon caused by increased longevity and decreased fertility rates (Bloom et al., 2015). The fraction of the population aged 60 and above in the world has more than doubled since 1980 and is expected to double again by 2050. The fraction of the population aged 80 and above is even expected to triple from 2017 to 2050 (United Nations and Social Affairs, 2017). Furthermore, the global old-age dependency ratio, defined as the number of people aged 65 and above for every 100 persons of working age (20-64), has increased from 20 in 1975 to 28 in 2015, and is expected to almost double by 2050 (OECD, 2017).

Population aging has lead to an increased focus on sustaining good health and well-being in later life. This is because maintaining good health and well-being in later life is a key societal aspiration in itself (Steptoe et al., 2015), but also because population aging causes challenges for the sustainability of public finances. Older people have low employment rates, more health problems and greater long term care needs than younger people (Bloom et al., 2015).

The pressure on public finances has lead to an increased focus on extending working lives. A widespread policy tool in this regard is pension reforms, as has been or are on the verge of being implemented in several developed countries (OECD, 2017). Despite health being of critical importance for staying in work, the key policy tool of these reforms has been improving financial incentives, rather than measures to preserve health. If postponed retirement is detrimental to health, unintended side effects of pension reforms may include increased health care expenditures or reduced efforts in the informal care sector.

As the share of the population and number of individuals aged 80 and above increases, so does the need for long-term care. Norway has among the highest spending levels on publicly provided elder care in Europe (Huseby and Paulsen, 2009). Still, the amount of informal care provided is estimated to be of the same magnitude as the amount of publicly provided elder care (for those aged 67 and above) (Ministry of Health and Care Services, 2006). Hence, another crucial resource for the sustainability of public finances is informal caregivers. Informal caregivers are typically peers or family members of frail individuals, who provide care without being financially compensated for the effort provided (Jakobsson et al., 2012).

An important group of informal caregivers are those caring for frail, older parents. These caregivers are often daughters who are in the last stage of their working lives. As younger cohorts of women have higher rates of labor force participation and as the aim of the pension reforms is to increase working lives, the health implications of caregiving

are important, not only for individual well-being and health care expenditures, but also for female labor supply.

Determinants of health in later life have received relatively little attention in economics. However, individuals in this phase of life represent a significant resource for the formal and informal labor markets. It is not uncommon to have at least 15 potentially active years from retirement to institutionalization or death. Health is crucial for the residual working capacity, as well as for individual well-being and public and personal finances. Thus, more knowledge about later life health and the health effects of common events, such as retirement and becoming an informal caregiver, in this phase of life is needed.

The health of people in early senescence and old age is the focus of this thesis, which comprise three self-contained essays on different aspects of health and aging. Common features of the essays are empirical assessments of health in later life and the econometrical challenges caused by selection–either in the form of people self-selecting into "treatment" or in the form of sample selection in the data applied. Specifically, the thesis is concerned with estimating causal health effects of retirement and informal caregiving in later life, and with the problems associated with health-related non-response in analyses of health and ageing based on survey data.

A common critique in the applied health economics literature is that limited sample sizes and subjective health outcomes pose problems for empirical analyses of health, typically due to the reliance on survey data. This is also a limitation in the literature on health effects of retirement and informal caregiving, and the literature on sample-selection. This thesis contributes to the literature by providing causal and descriptive evidence based on full population individual level register data.

Careful attention to the identification of causal effects is needed to provide policy relevant empirical evidence. As individuals to some degree self-select into retirement and informal caregiving, not accounting for this selection is a potential source of bias for two reasons. First, the age in which a person retires, or who becomes a caregiver, is likely to be correlated with unobservable factors that also correlate with the person's health, which can cause omitted variable bias. Second, poor health in itself is a likely cause of early retirement and poor connection to the labor market, which again increases the likelihood of providing care to a frail family member, which can cause bias due to dual causality.

The literature on causal studies of the health effects of retirement is growing. However, these studies are often based on lower retirement age thresholds and early retirement opportunities. The causal evidence of caregiving on health is altogether limited, and several authors highlight the importance of, as well as lack of evidence on, long-term health

effects of providing informal care (Leigh, 2010; Schmitz and Westphal, 2017). In this thesis, I employ quasi-experimental methods in which eligibility rules and institutional changes provide sources of exogenous variation in the retirement eligibility and caregiver burden and, thus, allow identification of the causal effects on health.

Socioeconomic status (SES), typically proxied by education, income, or occupation, is often overlooked in analyses of the implications of retirement and caregiving. However, SES can be important for the mechanisms behind the effects in question, and the heterogeneity in the effects across SES groups can be large. For example, having a physically demanding job or a job with high demands, but low control, can cause physical and mental strain. Retiring from such jobs is likely to have positive effects on health. These types of jobs are often concentrated among the lower SES groups (Case and Deaton, 2005). If, in addition, retiring from the types of jobs concentrated among the high SES groups is detrimental to health, then, on average, retirement can be found to have no effect on health, whereas, in fact, it significantly affects health. Ultimately, this heterogeneity can imply that retirement reforms are socially distortive, especially if early retirement implies a fiscal penalization. Those with low SES are faced with a trade-off between a monetary cost and a health cost, whereas individuals with high SES may not face such a trade-off. In this thesis, I provide evidence based on formal tests of effect heterogeneity by SES groups in the health effects of retirement and informal caregiving.

A large literature in the health sciences, demography, and social gerontology is devoted to the question of whether social inequalities in health are converging or diverging across age in later life. On the one hand, later life is a phase when determinants of inequalities in health, such as type of job and income, are of secondary importance as most individuals have retired and pensions are more evenly distributed than wages. On the other hand, after a life-time of exposures to risks and opportunities, several of which are socially determined, inequalities can be large and increasing due to the accumulation of these exposures. The majority of analyses on this topic are based on survey data.

A recurring issue in this thesis is that survey data, especially on older individuals, are hampered by health-related non-response. The third essay of the thesis is devoted to investigating how health-related non-response in survey data affects analyses of social inequalities in health in later life. While selective attrition across waves in survey data has been extensively studied, there is limited evidence on the impact and magnitude of health-related initial selection due to the limited information available for the initial non-responders. In this essay, access to full population registers for the population that the survey data is based on, is used to create an arguably generic measure of health, namely reimbursements from primary care health services. This health measure makes

it possible to study the health-related non-response in the survey data. This is a novel approach to assessing the impact of sample selection. In addition, it enables assessing initial non-response.

Health and Aging: What Can We Learn from Economic Theory?

The essays in this thesis are mainly empirical and methodological. However, the empirical questions raised have a theoretical foundation. A useful benchmark of how health evolves across the adult life-course is the widely cited human capital model of the demand for health, known as the Grossman model (Grossman, 1972). The Grossman model derives the structural relationship between costs of health care, earnings, education, prices and time preference, and the demand for health and health care. The model is originally a dynamic optimization problem where lifetime utility is maximized subject to a budget contraint, but it has also been common to report simplified two-period versions of the model (Zweifel et al., 2009). Here, I discuss a version of the Grossman model accommodated more specifically to investigate social inequalities in health and health deterioration in later life, applied by Case and Deaton (2005).

In the Grossman model, health is defined in terms of a stock variable that depreciates over age (time). The rate of depreciation is assumed to increase with age. Individuals can invest in own health by purchasing medical services and healthy food (or restrain from unhealthy consumption), or they can spend time doing physical exercises or other health promoting activities. Individuals demand health both for direct utility (consumption good) and as a means of production in the labor market for generating earnings or household production (investment good). Additionally, the individual faces a trade-off between health and consumption of all other goods.

In addition to health capital, the individual possesses human and financial capital. The three sources of capital are imperfect substitutes in generating earnings and utility. The rate of health decline across age is thus determined by two parts: a biological process of aging over which the individual has little control, and the usage of health capital in consumption and in generating earnings. Since the three sources of capital are substitutes in generating earnings and utility, individuals with low education or low levels of financial capital will rely more heavily on their health capital. Consequently, individuals with low education or wealth will experience a higher rate of depreciation of health as they age. In addition, education affects the effectiveness of health investments so that higher educated individuals are more efficient health producers.

The Grossman model provides theoretical insights into the health effects of retirement and informal caregiving. Individuals use health capital as an input in the formal and informal labor market. Thus, being relieved of caregiving responsibilities or retiring from

work is likely to affect health positively, all else constant. The Grossman model can also be used to study how retirement and informal caregiving affect health across SES groups differently. More specifically, the model predicts that being relieved of work and informal caregving tasks is likely to have a larger health effect for low SES groups compared to higher SES groups. There are mainly two mechanisms for this relationship in the model: First, individuals with low levels of human or financial capital use more of their health capital as inputs to generate earnings. Retirement will therefore have a larger health impact for this group. Second, individuals with higher levels of financial capital can purchase formal care, making them use less health capital in providing informal care. Being relieved of informal caregving tasks is thus likely to have a larger effect on health for the low SES groups.

In addition, the Grossman model predicts that individuals with lower levels of human and financial capital, will experience a higher rate of health depreciation as they age. This has implications for health- and SES-related non-response in survey data and how the selective non-response increases with age. Assuming that health above a certain level is needed in order to be able to respond to a survey, those with lower SES will have lower likelihood of responding, for a given age, due to poor health. This difference becomes stronger as age increases, causing the responders from the low SES group to become increasingly less representative of that group as a whole. This leads to sample selection problems in analyses of health, SES and aging, which is the topic of the third essay.

Measures of Health in Empirical Analyses: Survey Data vs. Register Data

How to measure health in a certain context is a recurring issue in this thesis. This section briefly discusses common measures of health as well as limitations and possibilities with different types of health data, especially within the context of later life health and SES.

A common definition of health is the one applied by the World Health Organization (WHO) in the first paragraph of its constitution: "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (World Health Organization, 1946). However, in order to conduct quantitative analyses, a measure that is quantifiable, or at least has an ordinal nature, is needed. The most common operalization of such a general health measure, is the "self-rated health" (SRH), where individuals are asked to rate their health on a scale from one to four or five (Idler and Benyamini, 1997). Jylhä (2009) describes SRH as a "cross-road between the social world and psychological experiences on the one hand, and the biological world, on the other" (Jylhä, 2009, p. 308). In addition to being a generic measure of health, the benefits of using SRH are that it is easy to implement and widely applied, thus facilitating comparisons across analyses.

Other commonly applied measures are the Short-Form 36 (SF36) (Ware Jr and Sherbourne, 1992) and the shorter Short-Form 12 (SF12) (Ware Jr et al., 1996). The SF36 includes a multi-item scale that assesses eight health concepts,[1] and was constructed for use in clinical practice and in population surveys (Ware Jr and Sherbourne, 1992). The SF12 is based on the SF36 and was developed to substantially shorten the survey time.[2] SRH is one of the components in the SF36 and SF12.

Although extensively used, measures of self-reported health are criticized for being contextual and to suffer from justification bias (Bound et al., 1991; McGarry, 2004). One can easily imagine how the response to a question about own health depends on several factors besides own health, such as the health of the respondent's peers or how current health is compared to past health or compared to what the respondent's expectations of health for a given age is.

The criticism of self-reported outcomes often lead the researcher to look for more objective measures, such as data from administrative registers (register data). Register data is reported by a third party and typically includes measures of health related to records of health care utilizations, such as hospital admissions or primary care consultations, diagnoses, or certified leaves of absence. In addition, register data often contain health records for the full population, and hence is not hampered by health-related sample selection like survey data often is, as discussed in the third essay in this thesis.

Register data do however have its own limitations. The health measures found in register data are often related to specific dimensions of health, such as a diagnosis or measures of health care utilization, rather than generic measures of health. For instance, in my third essay, I study how the social gradient in health [3] changes across age in later life. Without a generic health measure, using a health measure with a steep SES gradient (such as diabetes or cardiovascular disease) or a measure that does not have a clear gradient (such as certain types of cancer) could lead to very different conclusions.

Furthermore, although measures based on register data are objective in the sense that health is not self-assessed or self-reported, many are contingent on the patient actually seeking medical help. In addition, the diagnosis or treatments provided by the physician depend, to some degree, on the explanation of symptoms given by the patient. In this

[1] These are: 1) limitations in physical activities because of health problems; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions. (Ware Jr and Sherbourne, 1992)

[2] The items that constitute the SF12 are those from the SF36 that best predict the results using the SF36 on a US reference population as measured by the R^2 (Ware Jr et al., 1996).

[3] As education or income increase gradually, health increases gradually, i.e., those with higher income have better health than those with lower income, also within different SES groups (Lleras-Muney, 2005; Cutler et al., 2008)

sense, there is an element of self-assessment also in several of the register data health measures. In fact, it has been shown that SES matters for the amount of care provided by the secondary health care services (Elstad, 2018).

The challenges of endogeneity are further exemplified in the thesis' second essay, where we assess the health effects of retirement. In this setting, applying a health outcome that correlates with the alternative cost of time may lead to biased results. As retirement reduces the opportunity cost of time, the likelihood of seeking medical help, and hence, the likelihood of having a diagnose or record of health care utilization, increases post retirement. A change in health measured by various objective measures can therefore erroneously be ascribed to retirement, when in fact it was caused by the reduced opportunity cost of time. On the other hand, once an individual is retired, the need for sickness absence certifications is no longer present, thus reducing the need for health care services for that specific reason. In both instances, actual health can remain unchanged for the individual, but the changed circumstances that follows retirement cause the health record to change.

The gold standard of health measures among economists is mortality (Case and Deaton, 2005).[4] Mortality is indisputable in that it measures exactly the same across all subjects and contexts. It is an outcome that is readily comparable across data sources and circumstances. However, as mortality is an extreme and final outcome, the scope research question that can be assessed by this measure of health is limited. For instance, a central policy issue in the context of population aging is whether postponed mortality also implies compression of morbidity or rather a prolonged phase of morbidity prior to death. Obviously, mortality data alone cannot be used to answer this question.

The discussion above has followed the conventional discourse within the field of applied health economics, namely the statistical properties of different health measures.[5] Objective measures of health, often based on register data, are generally preferred over more subjective measures.

A dimension to this debate, which I often find overlooked, is that health by nature is to a large degree subjective. An individual's utility and demand for health is not independent of the circumstances in which the individual finds itself. Thus, having certain health limitations does not entail the same for every individual. The ability to work depends on the type of work, but also on how the individual experiences health. Provided the same objective health limitation or disease and the same type of job, some people will work, whereas others will remain at home or obtain a certificate of sickness leave. This idea is

[4] A benefit of self-reported health outcomes that is often emphasized, is their ability to predict mortality (Idler and Benyamini, 1997; Lacson et al., 2010; Jylhä, 2009).

[5] This is disregarding the large literature on the cost and demand and supply of health care.

also present in the Grossman model, where the amount of health that is demanded as an input in the formal and informal labor market depends on the amount of other sources of capital. As economists, we care about individual's utility, not health itself. With this in mind, self-reported measures, despite their poor reputation, can provide important insights and highlight dimensions of health that mortality or data from health registers cannot.

In this thesis, I use measures of health that, as far as possible, fulfil the objectivity and sample selection criteria within the particular contexts studied. In light of the above paragraph, however, I complement these objective measures with self-reported, generic health measures where possible.[6] The next sections summarize the three papers of this thesis and provides a brief conclusion.

Paper 1: Formal care of the Elderly and Health Outcomes Among Adult Daughters

With an aging population, health-care expenditures and the demand for caregiving are increasing concerns for policy makers. Informal caregivers are a crucial national resource in this regard. In Norway, the amount of informal care is estimated to be of the same magnitude as the amount of the publicly provided formal care received by the elderly population (Berge et al., 2014). However, providing care to a frail family member can be detrimental for health.

In this essay, co-authored with Signe A. Abrahamsen, we assess the immediate and long-term health effects on informal caregivers of increasing the publicly provided formal care. The relationship between care responsibilities and health has been well assessed, and the majority of the empirical evidence documents negative correlations between health and caregiving.

Despite a well assessed link between caregiving and health, much of the empirical literature suffers from a lack of causality. Several studies have poor or even lack reliable identification strategies. In addition, a large share of the available studies are based on cross-sectional data with small sample sizes, and often rely on subjective measures of health. There is little evidence using large-scale register data that rely on objective health outcomes. Thus, the evidence for causal effects of caregiving remains limited, especially for long-term outcomes.

Estimating causal effects of reduced care responsibilities on health is difficult for several reasons. First, formal care uptake is endogenous. Elderly individuals who receive formal care typically have a higher need for care. They may be older and less healthy than

[6] Unfortunately, this was not possible in the first essay, as the survey data used could not be linked to municipal data. It would, however, have been interesting to assess the effects of increased formal care on self-reported and generic measures of health for the caregivers.

those not receiving formal care. Health has a genetic component and a strong connection to lifestyle which may also be inherited. Therefore, simply comparing children with different levels of parental care responsibilities may lead to biased results. Second, caregiving may have endogenous uptake, too. Individuals with poor relations to the labor market and, thus, lower opportunity costs are more likely to provide care (Carmichael et al., 2010). Poor relations to the labor market may be related to poor health in itself. Thus, poor health is a likely cause of being a caregiver which may further bias results obtained by comparing health outcomes for individuals with different levels of care responsibilities.

We address the endogeneity problem by exploiting a reform introduced in 1998, which aimed at equalizing the availability of care services across Norwegian municipalities. This reform led to an arguably exogenous regional variation in the expansion of formal care services, which enables us to estimate causal effects of increased formal care by comparing outcomes across municipalities with different levels of formal care expansions. We assess the sub-sample most likely to be affected by the reform: single-child daughters who have only one remaining parent who is at least 80 years old, and estimate causal effects of expanding publicly provided eldercare on short and long-term health outcomes for these daughters.

First, we replicate the main finding from Løken et al. (2017) and show that increased formal home-based care leads to reduced insured sickness absence from work. We supplement their finding by exploiting detailed full population register data on underlying diagnoses and assess whether certain diagnoses drive the observed reduction in leaves of absence. Our results indicate that musculoskeletal and psychological disorders are the main drivers.

We then assess long-term health outcomes for the same sub-sample of daughters. Here, we employ rich and detailed register data on the utilization of primary health-care services and on diagnoses and symptoms. With one exception, we estimate no effects on health care utilization or on the occurrence of a number of diagnoses identified in the literature as potentially affected by caregiving. The exception is an estimated 11 percent decrease in the risk of hypertension in the longrun, which may be explained by lower stress levels caused by reduced care burden following the reform. However, this finding does not remain statistically significant after correcting for multiple hypothesis testing, and thus, must be interpreted with caution.

In addition, we assess whether there are heterogeneous effects by SES group. This dimension is not much assessed in the literature, and it is not straightforward to hypothesize whether individuals with high or low SES will be more affected by the care expansion. In general, we find that there is not much difference in the effects of the care expansion across educational groups.

Paper 2: Health Effects of Retirement: Evidence from Survey and Register Data

With increasing life expectancy, the number of retired individuals as a share of the total population is rising in most OECD countries. This has led to concerns about the fiscal sustainability of public pension systems, and to policy initiatives that aim to extend working lives and increase retirement age.

In the second essay, co-authored with Otto S. Lillebø, we investigate the health effects of retirement across SES groups in Norway. This study is related to a growing body of economic research about the effect of retirement on health. Given the important aspect of this issue and the growing amount of literature on the topic, there is a surprising lack of consensus across studies. In this study, we seek to contribute to four limitations in the current literature. The first is that people self-select into retirement, making causal evidence difficult to identify due to the endogeneity of retirement in the health equation. The second is that most studies assess retirement effects in the early 60s, an age threshold that is substantially lower than proposed policies to postpone retirement toward age 70. The third is the reliance on self-reported outcomes from smaller samples of survey-data. Finally, heterogeneity in the retirement effects across SES is a dimension that is under-explored.

The endogeneity problem arises because people self-selection into retirement. Poor health is a likely cause of retirement, and simply comparing the health of the retired to the non-retired part of the population is therefore likely to produce biased results. Moreover, individuals might have unobservable knowledge about own longevity or other factors that influence retirement behaviour and correlate with health. To control for the endogeneity of retirement, we exploit a statutory retirement age that caused a discontinuous change in the likelihood of retiring at the exact timing of eligibility. This implies a local randomization around the retirement eligibility age threshold, and makes a regression discontinuity (RD) framework suitable. RD entails comparing the health outcomes for those right above the statutory retirement age threshold (i.e. the treatment group) to those right below (i.e. the control group). This allows for identification of the causal short-term effects of retirement on health. The age threshold that we exploit, is retirement at 67 years, which was the statutory retirement age in Norway in 2007.

We apply measures of health from full population register data and survey data to get a comprehensive approach to health. From the registers, health is measured by acute hospitalizations and mortality. To capture more moderate health effects and to enable comparison with the previous literature, we include the SF12 from the NorLAG survey. We assess both the overall physical score and the specific components that goes into the SF12.

Finally, although some studies highlight the importance of SES in the health effects (Coe and Zamarro, 2011; Eibich, 2015), there is limited evidence from formal of effect heterogeneity. SES is important in the analysis of health effects of retirement because it determines the kind of work situation an individual retires from. Manual labor jobs, associated with low education, are considered "wear and tear' types of jobs, in which health deteriorates at a more rapid pace than for individuals in non-manual professions Case and Deaton (2005). Thus, retirement can be seen as a mechanism that levels health inequalities between SES groups. In this analysis, we provide a formal test of effect heterogeneity according to SES group or gender.

Our results confirm what has been found in several studies, namely that retirement has a positive effect on health for subjective health outcomes. In addition, this study contributes to the literature by generalizing the positive physical health effect of retirement across a larger age span. Further, we find that the positive health effect is especially strong for the low SES group, whereas we find no effects for the high SES group. How retirement affects objective health is less clear as there exist little evidence using objective health measures, especially for the full population. In general, we find no effects on the objective measures, besides suggestive evidence of a retirement effect on reduced likelihood of acute hospitalizations for the low SES group. We thus conclude that retirement mainly affects subjective health.

The objective health measures assessed in this study can be seen as extreme in the sense that they cannot capture moderate health effects. Assessing the factors that go into the SF12, indicates that retirement affects health in a more fundamental way than subjective health. Future research should thus seek to assess objective health outcomes that are generic or that enables identification of moderate changes in health. In doing so, it is key to recognize that retirement necessarily coincides with the reduced opportunity cost of time.

Finally, this study accentuates the importance of assessing heterogeneity in the retirement effects due to individuals' different circumstances. Our findings indicate that the retirement reforms aimed at prolonging working life by financial penalties for early retirement can be socially distortive due to the differential health effects based on SES.

Paper 3: Old Problem, New Evidence: Health-related Sample Selection in Analyses of Health and Aging

Statistical inference about health is typically based on survey data. This is due to the feasibility of obtaining generic measures of health from surveys (compared to, e.g., administrative registers) and due to the limited availability of population-level data containing health information. Health might be a particularly important predictor of non-response

in survey data, and health-related non-response might become even more pronounced in samples of older individuals. This causes, studies of health and aging based on survey data to be particularly prone to bias due to selective non-response.

The impact of sample selection in analyses using survey data is an old question that has been extensively assessed. The majority of the empirical literature has focused on attrition across waves in longitudinal surveys. Although evidence of health-related attrition across waves has been documented in several well-known household surveys, it is found to cause little bias (Becketti et al., 1988; Fitzgerald et al., 1998; Contoyannis et al., 2004; Jones et al., 2006). Empirical evidence assessing the impact of initial non-response is, on the other hand, limited.

In this essay, I assess the bias from health-related sample selection, defined as either initial selection or selective attrition across waves, in a health survey of older individuals. The potential bias from selective non-response in the second wave of the NorLAG survey is assessed both compared to the survey's reference population using full population register data and population weights (initial selection), and compared to all those who responded to the first wave (selective attrition) using conventional sample selection methods, such as inverse probability weights (Wooldridge, 2002) and the Heckman selection model (Heckman, 1979), in addition to the more recently developed Copula selection model (Smith, 2003).

A novelty of this analysis is access to sufficient individual level health information collected from register data to compose a generic health measure for each individual in the population. The register data are not hampered by non-response, and thus, the registers can serve as a base for evaluation of health-related non-response in the survey data. This approach hinges on the comparability of the health measures applied from the two data sources. I apply the SF12 from the NorLAG data. From the registers, I propose the following measure of health: the annual reimbursement for a patient claimed by the primary care physician (PCP). I show that the annual PCP reimbursements (PCPR) have the same pattern of distribution as the SF12 across dimensions that are central to this analysis, and thus, can arguably be used as a benchmark for how the SF12 would be distributed in the population across the variables that are of interest in this study: gender, age, and education.

As a case study, I apply an empirical problem that is especially prone to bias from health-related sample selection: social inequalities in health across age. This is an application where the problem of health-related sample selection is believed to be especially salient, as the outcome of interest, health, and the explanatory variables, age and SES, are likely to correlate with non-response. Unless properly accounted for, the consequence of this sort of sample selection is empirical evidence of converging social inequalities in

health in later life, which is a common empirical result in the social gerontological literature, where the health-SES-age relation has been most widely studied (Mirowsky and Ross, 2008).

I find that the health trajectories estimated by the different weighting and selection models do not substantially differ from a baseline of no selection correction. This result corresponds to the literature on sample selection, where, despite evidence of selective non-response in the data, the bias is found to be small or negligible. Across all specifications, I find that the social inequalities in health are falling across age in later life. Converging inequalities in health in old age are also found in the analysis using the full population register data. Thus, taken together, these results suggest that selective non-response in the survey data is not causing the empirical evidence of converging health inequalities in later life, which suggests that survey data are well-suited to describe patterns of health inequalities as they manifest in the population. However, using a large panel of full population register data, I show that selective mortality leads to stronger healthy survivor bias in the lower SES groups compared to higher order SES groups.

The aim of this study was to shed new light on the potential bias from sample selection in analyses of health and aging, and the study makes two contributions to the existing literature. First, a novel feature is the presentation of a register-based generic measure of health, available for the the full population (PCPR), that serves as a base for evaluation of health-related non-response in the survey data. Second, although IPWs has commonly been used in similar settings, few examples use selection models to analyse the bias from health-related sample selection. This is especially true for the Copula selection model.

Finally, this study highlights the importance of selective mortality in analyses of health and aging. The population alive at older ages provide a poor representation of the population alive at younger ages. Mortality selection, present even in the most perfect dataset, is a concern in analyses attempting to measure the relationship between groups with different mortality rates across time or age, or in analyses of later life outcomes in general, provided mortality is correlated with the outcome of interest.

The magnitude and impact of health-related initial non-response in survey data remain largely unresolved. This essay highlights the importance of allowing survey data to be combined with measures of health from registers that are available for the full population. This would not only allow assessing the bias from health-related non-response more explicitly, as the exact health distribution in the population compared to in the survey data would be known. This would also allow assessing whether those in low SES groups or high age groups have lower response rates because these groups also have worse health outcomes, or if these characteristics are associated with lower response rates in themselves.

Concluding Remarks

Population aging warrants more knowledge about how circumstances and events in later life determine health. Despite its importance for labor market participation, health care costs, and general well-being, later life health has received relatively little attention in economics. This thesis comprise three self-contained essays assessing different aspects of aging and health.

The contributions from this thesis lies in providing causal evidence of the health effects of providing informal care and retirement, and providing new evidence on the impact of health-related non-response in analyses of health and ageing based on survey data. In addition, all essays in this thesis answers to the common critique of limited sample sizes and reliance on subjective health outcomes in the applied health economics discourse. The thesis provide evidence based on a combination of full population register data and survey data, and can, thus, be applied in a larger discussion about how survey results might differ from what we would have found if using full population data and more objective health outcomes.

A limitation of the findings from the two essays on the health effects of retirement and informal caregiving, is that the results are local in the sense that they cannot easily be generalized to other contexts. This is a common critique in the applied work that classify as what has been labelled *The credibility revolution in economics* (Angrist and Pischke, 2010). The focus within this field of research is on internal validity, which is fundamental for the identification of causal effects, whereas external validity is of secondary importance.

However, there is a cumulative force of studies on one particular topic if it is studied across a wide variety of settings. The magnitude of evidence on the same phenomena across various sub-populations or contexts is what constitutes the the evidence base and the epistemology of applied microeconomics. In light of this perspective, more knowledge about how an individual's circumstances in later life determine health is needed in order to provide reliable policy advice regarding how to best protect individuals against poor health in later life.

Future research should acknowledge the extent of heterogeneity in health among older individuals. Health in later life is a function of a life-time of health behaviours and differential exposure to risks. On the other hand, as age increases, the impact of selective mortality makes people more similar with respect to health as only those who are biologically more robust, i.e., the healthy survivors, are left in the population.

References

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30.

Becketti, S., Gould, W., Lillard, L., and Welch, F. (1988). The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics*, 6(4):472–492.

Berge, T., Øien, H., and Jakobsson, N. (2014). Formell og uformell omsorg. Samspillet mellom familien og velferdsstaten. *Velferdsforskningsinstituttet, NOVA-notat*, 2014(3).

Bloom, D. E., Chatterji, S., Kowal, P., Lloyd-Sherlock, P., McKee, M., Rechel, B., Rosenberg, L., and Smith, J. P. (2015). Macroeconomic implications of population ageing and selected policy responses. *The Lancet*, 385(9968):649–657.

Bound, J. et al. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources*, 26(1):106–138.

Carmichael, F., Charles, S., and Hulme, C. (2010). Who will care? Employment participation and willingness to supply informal care. *Journal of Health Economics*, 29(1):182–190.

Case, A. and Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging*, pages 185–212. University of Chicago Press.

Coe, N. B. and Zamarro, G. (2011). Retirement effects on health in europe. *Journal of Health Economics*, 30(1):77 – 86.

Contoyannis, P., Jones, A. M., and Rice, N. (2004). The dynamics of health in the british household panel survey. *Journal of Applied Econometrics*, 19(4):473–503.

Cutler, D. M., Lleras-Muney, A., and Vogl, T. (2008). Socioeconomic status and health: Dimensions and mechanisms. Technical report, National Bureau of Economic Research.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of Health Economics*, 43:1–12.

Elstad, J. I. (2018). Educational inequalities in hospital care for mortally ill patients in norway. *Scandinavian Journal of Public Health*, 46(1):74–82.

Fitzgerald, J., Gottschalk, P., Moffitt, R., et al. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *Journal of Human Resources*, 33(2):251–299.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Huseby, B. M. and Paulsen, B. (2009). Eldreomsorgen i Norge: Helt utilstrekkelig-eller best i verden. *Sintef Helsetjenesteforskning*.

Idler, E. L. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, pages 21–37.

Jakobsson, N., Hansen, T., and Kotsadam, A. (2012). Er det en sammenheng mellom formell og uformell omsorg i Norge? *Tidsskrift for Velferdsforskning*, 15(3):168–175.

Jones, A. M., Koolman, X., and Rice, N. (2006). Health-related non-response in the british household panel survey and european community household panel: Using inverse-probability-weighted estimators in non-linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3):543–569.

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, 69(3):307–316.

Lacson, E., Xu, J., Lin, S.-F., Dean, S. G., Lazarus, J. M., and Hakim, R. M. (2010). A comparison of sf-36 and sf-12 composite scores and subsequent hospitalization and mortality risks in long-term dialysis patients. *Clinical Journal of the American Society of Nephrology*, 5(2):252–260.

Leigh, A. (2010). Informal care and labor market participation. *Labour Economics*, 17(1):140–149.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the united states. *The Review of Economic Studies*, 72(1):189–221.

Løken, K. V., Lundberg, S., and Riise, J. (2017). Lifting the burden: Formal care of the elderly and labor supply of adult children. *Journal of Human Resources*, 52(1):247–271.

McGarry, K. (2004). Health and retirement do changes in health affect retirement expectations? *Journal of Human Resources*, 39(3):624–648.

Ministry of Health and Care Services (2005-2006). *Mestring, muligheter og mening. St. meld. 25.* Ministry of Health and Care Services, Oslo.

Mirowsky, J. and Ross, C. E. (2008). Education and self-rated health: Cumulative advantage and its rising importance. *Research on Aging*, 30(1):93–122.

OECD (2017). *Pensions at a Glance 2017: OECD and G20 Indicators*. OECD Publishing, Paris.

Schmitz, H. and Westphal, M. (2017). Informal care and long-term labor market outcomes. *Journal of Health Economics*, 56:1–18.

Smith, M. D. (2003). Modelling sample selection using archimedean copulas. *The Econometrics Journal*, 6(1):99–123.

Steptoe, A., Deaton, A., and Stone, A. A. (2015). Subjective wellbeing, health, and ageing. *The Lancet*, 385(9968):640–648.

United Nations, D. o. E. and Social Affairs, P. D. (2017). *World Population Ageing 2017*. United Nations Publications.

Ware Jr, J. E., Kosinski, M., and Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34(3):220–233.

Ware Jr, J. E. and Sherbourne, C. D. (1992). The mos 36-item short-form health survey (sf-36): I. conceptual framework and item selection. *Medical Care*, pages 473–483.

Wooldridge, J. M. (2002). Inverse probability weighted m-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, 1(2):117–139.

World Health Organization (1946). Preamble to the Constitution of WHO as adopted by the International Health Conference, New York, 19 June - 22 July 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of WHO, no. 2, p. 100) and entered into force on 7 April 1948.

Zweifel, P., Breyer, F., and Kifmann, M. (2009). *Health Economics*. Springer Science & Business Media.

Chapter 2:

Formal Care of the Elderly and Health Outcomes Among Adult Daughters

Formal Care of the Elderly and Health Outcomes Among Adult Daughters

Signe A. Abrahamsen * Maja Weemes Grøtting †

January 8, 2019

Abstract

Health-care expenditures and the demand for caregiving are increasing concerns for policy makers. Although informal care to a certain extent may substitute for costly formal care, providing informal care may come at a cost to caregivers in terms of their own health. However, evidence of causal effects of care responsibilities on health is limited, especially for long-term outcomes. In this paper, we estimate long-term effects of a formal care expansion for the elderly on the health of their middle-aged daughters. We exploit a reform in the federal funding of formal care for Norwegian municipalities that caused a greater expansion of home care provision in municipalities that initially had lower coverage rates. We find that expanding formal care reduced sickness absence in the short run, primarily due to reduced absences related to musculoskeletal and psychological disorders. In general, we find no effects on long-term health outcomes.

JEL Classification: I10, J14, J22, J38
Keywords: Formal and informal eldercare, sickness absence, health.

*Department of Economics, University of Bergen; email signe.abrahamsen@uib.no.
†Norwegian Social Research, Oslo Metropolitan University; email: maja.grotting@oslomet.no.

1 Introduction

With an aging population, the demand for care along with increasing health-care expenditures places an increasing burden on public finances. Informal care may both substitute for and complement publicly provided formal care; see e.g. Bonsang (2009), Bolin et al. (2008), and Van Houtven and Norton (2004). However, studies have shown that informal care may come at a cost to caregivers. For instance, care responsibilities can be a stressor as they often come in addition to other obligations, such as work and household chores, thus leading to reduced work hours or less leisure time (Vaage, 2000; Gautun and Hagen, 2010). Much of the care burden falls on close relatives, especially on the spouse or children of the person in need of care (Vaage, 2000; Jakobsson et al., 2012).

In a recent study, Løken et al. (2017) show that lifting the care burden for adult single-child daughters with one elderly parent led to a decrease in insured sickness absence from work. In this paper, we build on that result and assess which underlying diagnoses drive the reduction in sickness absence. Moreover, we assess whether the formal care expansion effected long-term health outcomes for these daughters.

The relationship between care responsibilities and health has been well assessed; see Bauer and Sousa-Poza (2015) and Pinquart and Sörensen (2003, 2007) for reviews. However, the link is complex. On one hand, taking care of a frail parent can be rewarding (Toljamo et al., 2012). On the other, care responsibilities often involve physically demanding tasks, as well as mental strains over substantial periods of time, which may have negative effects on health both in the short and long run. Finally, time spent on caregiving may also supersede health-enhancing activities, such as physical exercise and other recreational activities.

The majority of the empirical evidence in this area documents negative correlations between health and caregiving. Moreover, a substantial share of the empirical evidence focuses on psychological aspects, such as stress and depression (Bauer and Sousa-Poza, 2015; Gautun and Hagen, 2010; Pinquart and Sörensen, 2003, 2007). Living with elevated stress levels or poor mental health over time has, in turn, been linked to reduced physical health, such as cardiovascular disease, hypertension, lung ailments, accidents, and suicide (Von Känel et al., 2008; Capistrant et al., 2012). Care responsibilities may also have direct and immediate impacts on physical health through physically demanding tasks (Pinquart and Sörensen, 2006). Finally, it may take time for the burden of caregiving to manifest as health adversities. The importance of assessing delayed health effects of caretaking has been highlighted by Coe and Van Houtven (2009), Leigh (2010), and Schmitz and Westphal (2017).

Despite a well assessed link between caregiving and health, much of the empirical

literature suffers from a lack of causality. For instance, a large share of the available studies are based on cross-sectional data with small sample sizes, and often rely on subjective measures of health. In addition, several studies have poor or even lack reliable identification strategies. Also, there is little evidence using large-scale register data that rely on objective health outcomes. Therefore, the evidence for causal effects of caregiving remains limited, especially for long-term outcomes.

Estimating causal effects of reduced care responsibilities on health is difficult for several reasons. First, formal care uptake is endogenous. Elderly individuals who receive formal care typically have a higher need for care. They may be older and less healthy than those not receiving formal care. Health has a genetic component, as well as a strong connection to lifestyle which may also be inherited. Therefore, simply comparing children with different care responsibilities for their elderly parents may lead to biased results. Second, caregiving may have endogenous uptake, too. Individuals with poor relations to the labor market and lower opportunity costs are more likely to provide care. For instance, Mentzakis et al. (2009) find a negative correlation between health and the likelihood of providing care, and further, being employed is found to reduce the willingness to provide care (Carmichael et al., 2010). Therefore, poor health is a likely cause of being a caregiver, and comparing health outcomes for caregivers to non-caregivers likely provides biased results. Finally, poor relations to the labor market may also be related to poor health in itself, which may further bias results obtained by comparing individuals with high and low care responsibilities.

In this paper, we address the endogeneity problem and estimate causal effects of expanding publicly provided care of the elderly on later life health outcomes for their middle-aged children. We exploit a reform introduced in 1998, which aimed at equalizing the availability of care services across Norwegian municipalities. This reform led to an arguably exogenous regional variation in the expansion of formal care services, which enables us to estimate causal effects of increased formal care by comparing outcomes across municipalities with different levels of care expansion.

As in Løken et al. (2017) we focus on a sub-sample that is particularly likely to be affected by the reform, namely, single-child daughters who have only one remaining parent who is at least 80 years old. We replicate the main finding from Løken et al. (2017) and show that increased formal home-based care leads to reduced insured sickness absence from work in the short-run. We supplement this finding by exploiting detailed data on underlying diagnoses and assess whether certain diagnoses drive the observed reduction in leaves of absence. Our results indicate that musculoskeletal and psychological disorders are the main drivers. This is an interesting finding because musculoskeletal and psychological disorders are both associated, in previous studies, with caregiving. Moreover,

these diagnoses are particularly difficult to verify and thus, vulnerable to misdiagnosis.

We then assess long-term health outcomes for the same sub-sample of individuals. Here, we employ rich and detailed register data on the utilization of primary health-care services and diagnoses related to caregiving in the literature.[1] Overall, our results show that long-term health is not much affected by increased formal care. More specifically, with one exception, we estimate no effects on the occurrence of a number of diagnoses identified in the literature as potentially affected by care responsibilities. The exception is an estimated 11 percent decrease in the risk of hypertension in the long run, which may be explained by lower stress levels in the sample of daughters who experienced a reduction in their care burden following the reform. However, this finding does not remain statistically significant after correcting for multiple hypothesis testing, and thus, needs to be interpreted with caution.

Finally, there might be heterogeneous effects in dimensions such as socioeconomic status (SES). However, it is not straightforward to hypothesize whether individuals with high or low SES are more affected by the care expansion assessed in this paper. On one hand, we might expect individuals with low SES to be more affected by the policy change, as low SES individuals are more likely to take on a caregiver role (Schulz and Sherwood, 2008). On the other, we may also expect a strong reform response from high SES individuals as they may be more able to gain their parents access to the increased formal care capacity.[2] In general, being a caregiver can be especially stressful for low educated individuals as they might have less job flexibility with regard to e.g. working hours. At the same time, this group also has lower opportunity costs and may be more likely to reduce work hours due to caregiving responsibilities, making the higher educated group more likely to be in a situation that combines work and caregiving. However, in a meta-study Pinquart and Sörensen (2007) find no differences in the associations between caregiving and health across education. We add to the analysis by assessing differences in the effects for daughters with high and low education. In general, we find that there is not much difference in the effects of the care expansion across educational groups. The exceptions are a larger reduction in the probability of a sickness absence leave related to a psychological disorder for the group with low education, and that the estimated increase in the risk of hypertension in the long run seems to be driven by the group with high education.

The paper proceeds as follows. In Section 2, we review the related literature. In

[1] Importantly, as, for instance, the use of primary health-care services varies across the employed and the non-employed, we show that the expansion had no long-term effect on employment, or on death, or on the likelihood of receiving disability insurance.

[2] Fiva et al. (2014) argue that highly educated individuals (high SES) may be more able to navigate the public care provision bureaucracy to get proper care.

Section 3, we provide details on the institutional background and the reform. Section 4 describes the different sources of data, and Section 5 explains the empirical strategy. In Section 6, we discuss the results, and Section 7 concludes.

2 Related Literature

2.1 Informal caregivers

Several studies have documented that informal care of the elderly may to a certain extent substitute for formal care; see e.g Cutler and Sheiner (1994); Bonsang (2009); Stabile et al. (2006). However, the substitution rate seems to be weakening as the elderly individual in need of care becomes older or his or her health becomes more frail (Bonsang, 2009; Daatland et al., 2009; Bolin et al., 2008; Van Houtven and Norton, 2004), and some studies argue that when it comes to more severe care needs, such as nursing or inpatient hospital care, informal care is a complement to formal care; see Bonsang (2009) or Bolin et al. (2008). Nevertheless, with an aging population and an increasing demand for care, informal caregivers are a critical national resource. Even in Norway–a country with a large public sector and one of the highest spending levels on publicly provided eldercare in Europe (Huseby and Paulsen, 2009)–the amount of informal care is estimated to be of the same magnitude as the amount of the publicly provided formal care received by the elderly population (Berge et al., 2014).

Most informal care received by a frail elderly individual is provided by family members, usually the spouse or adult children (Vaage, 2000; Jakobsson et al., 2012). Informal care is, however, highly gendered. On the receiving end, mothers receive more care than fathers. However, this is likely because women tend to outlive their spouse, and thus, as men often die in a two-person household, women more often spend their last years widowed (Daatland et al., 2009). More notable is the gender difference in the provision of informal care: Female spouses provide more care than do male spouses, and daughters provide more help to their elderly parents than sons (Stark, 2005; Jakobsson et al., 2016).[3] Several potential explanations for these gender differences are given in the literature. For instance, women may have lower opportunity costs and/or a weaker connection to the labor market. A second explanation relates to the way formal care services are allocated. Stark (2005) and Jakobsson et al. (2016) note that public service managers often have gender-differentiated expectations for the amount of informal care that will be provided. In particular, informal care from daughters is seen as a closer substitute for formal care than informal care from sons, and no, or a lower level, of formal care is assigned when

[3]Interestingly, in a survey conducted by Jakobsson et al. (2012), men were more likely than women to report that they think family should provide care.

daughters are present.

Several studies have focused on the effects of care responsibilities on various labor market outcomes, and findings include reduced work hours (Schmitz and Westphal, 2017; Bauer and Sousa-Poza, 2015; Kotsadam, 2012; Gautun and Hagen, 2010), lower earnings (Løken et al., 2017; Schmitz and Westphal, 2017; Heitmueller and Inglis, 2007), and a weaker connection to the labor market in general (Kotsadam, 2012; Leigh, 2010; Lilly et al., 2007).[4] Most notable for the present study, Løken et al. (2017) find that expanding formal care in Norway led to reduced sickness absence for daughters of single parents aged 80 and older. These results were especially strong for the subsample of daughters with no siblings, which is the group most likely to have a higher burden of care. There are two possible mechanisms for why caregiving affects sickness absence. First, as further elaborated in the next section, being a caregiver can have direct and indirect health implications. Second, a sickness absence can be used as a means of being granted paid leave from work to gain flexibility to provide care for a frail parent. Although all sickness absences longer than three days must be certified by a physician to prevent fraudulent use of leave, several studies confirm that this program has been used for other reasons other than an individual's sickness. Markussen et al. (2011) conclude that the sickness insurance system in Norway is a more general absence system, where physicians help individuals cope with demanding life situations by certifying sickness absences. This conclusion has been confirmed by interviews with physicians (Carlsen, 2008) and by survey data (Gautun, 2008).

2.2 Health Effects of Caregiving

A vast amount of empirical evidence documents negative correlations between caregiving and various health outcomes. Bauer and Sousa-Poza (2015) and Pinquart and Sörensen (2007) provide reviews of this literature and note that the focus is often on various psychological aspects, or mental health, such as stress, depression, anxiety, and dementia, of which depression is the most common outcome. Bauer and Sousa-Poza (2015) argue that caregiving could have mental health consequences because caregiving is time-consuming and often difficult to combine with work and family duties, and because caring for close relatives may induce negative emotions due to compassion and fear of loss. Moreover, Schulz and Sherwood (2008) note that caregiving is seen as having all the features of a chronic stress experience, as caregiving is usually accompanied by high levels of unpredictability. Having caregiving responsibilities has even been used as a model for studying chronic stress (Schulz and Sherwood, 2008).

[4] Allthough some studies find that the labor market effects hold only for the more intensive caregivers (Lilly et al., 2007; Kotsadam, 2012), this is not true for all studies.

In addition to the documented effects on psychological aspects, caregiving can affect physical health. This link could be explained through the mental health channel, as living with depression, or poor mental health in general, over substantial periods of time may manifest as physical health problems. Elevated stress over a longer time span has, for instance, been linked to hypertension and cardiovascular disease in the long run (Kivimäki and Steptoe, 2018; Pinquart and Sörensen, 2006; Capistrant et al., 2012; McEwen, 1998). In addition, care responsibilities may have direct effects on physical health. Some studies link care responsibilities directly to an increased risk of hypertension and cardiovascular disease (Von Känel et al., 2008; Capistrant et al., 2012). Bauer and Sousa-Poza (2015) argue that caregiving often includes physically demanding tasks over longer periods of time that in turn, may cause musculoskeletal injuries and aggravation of arthritis and other chronic conditions. Having care responsibilities can also require physical effort, especially in the case where the care recipient develops behavioral problems (Pinquart and Sörensen, 2006). Finally, caregivers tend to neglect physical exercise, eating healthy, or other factors associated with a healthy lifestyle (Bauer and Sousa-Poza, 2015; Capistrant et al., 2012).

Despite the well-assessed link between caregiving and health, there are several drawbacks with the existing literature. The evidence on physical health effects is still very limited, especially for more objective health outcomes. The majority of studies apply survey data, where the representativeness of the sample and the fact that the health outcomes can be highly contextual are potential issues. Importantly, as discussed in the introduction, the endogeneity problem due to selection into caregiving is an issue seldom addressed (Leigh, 2010).

Among the evidence that accounts for endogeneity, Coe and Van Houtven (2009) find that long-term caregiving reduces self-rated health and increases depressive symptoms. The increase in depressive symptoms is persistent, and the long-term consequences are larger than the immediate consequences. The negative effect of caregiving on depression is supported, for instance, by Bauer and Sousa-Poza (2015). Schmitz and Westphal (2015) find short-term effects of care-giving on mental health but no effect on physical health. Finally, Do et al. (2015) apply parent-in-law's health limitations as instrumental variables and find negative effects on pain, self-rated health, and out-of-pocket health-care spending for daughter-in-law caregivers.

Gender differences are found in the health effects of caregiving. Female caregivers report higher rates of depression compared to men (Bauer and Sousa-Poza, 2015). However, this is partly due to women being far more exposed through longer and more intense care (Pinquart and Sörensen, 2006).

Finally, little is known about persistence or the long-term health effects of caregiving.

Coe and Van Houtven (2009) argue that it takes some time for the effects of caregiving on health to manifest, and several studies stress the importance of assessing long-term outcomes (Leigh, 2010; Schmitz and Westphal, 2017).

In sum, theoretical and empirical studies have provided links between caregiving and health outcomes. This study contributes to the literature by assessing long-term effects on a number of these outcomes, in particular, musculoskeletal diagnoses, cardiovascular disease, psychological disorders in general, depression, hypertension, and lifestyle disease. Our identification strategy is based on arguably exogenous variation in caregiving following a formal care expansion. We rely on unique, high-quality individual-level register data that cover the entire Norwegian population. Further, in contrast to a large share of this literature, we apply objective health outcomes, measured as the diagnosis provided by a physician.

3 Formal Care in Norway and the 1998 Reform

With the emergence of the modern Norwegian welfare state in the 1960s, public responsibility for care of the young, frail, and elderly was expanded. The legal responsibility for elder care shifted from the family to the public sector in 1964, and during the 1970s, public expenditures on elder care more than doubled. Most of this expansion was in the form of support for home-care services, either in private homes or in adapted facilities.

Responsibility for elder care services gradually shifted from the central authorities to the municipalities during the 1980s. This was a period of decentralization, and municipalities were granted more autonomy in the provision of services to their elderly population. This implied that earmarked grants for elder care were replaced by transfers to municipal budgets based on demographics, income, and estimated needs in the municipality. Consequently, municipalities were free to allocate their budgets with close to full discretion, resulting in large differences in the coverage rates of home-care services across municipalities.[5]

During the mid 1990s, the need for care expanded rapidly due to a growing elderly population resulting in declining coverage rates of home-based and institutionalized care for the elderly aged 80 and older. At the same time, the large differences in home-based care coverage rates across municipalities were seen as inequitable. This resulted in an action plan for elder care services.[6] The plan was implemented in January 1998, with a four-year implementation period. The main purpose of the action plan was to increase capacity in buildings and personnel, and to create a more equitable level of care services

[5] Although municipalities have nearly full discretionary control over their budgets, certain amounts of public services have to be provided as all Norwegian citizens have a statutory right to basic welfare services regardless of where in the country they live.

[6] In Norwegian: "Handlingsplan for Eldreomsorgen".

across municipalities (Brevik, 2001).[7] Explicit goals of the action plan included that all municipalities should be able to provide 24/7 assistance to at least 25 percent of their population aged 80 and older. Specifically, the plan was to increase the number of spaces in adapted apartments and institutions, and to increase the number of labor input in the home-based care sector nationwide by 6000 work years (Borge and Haraldsvik, 2006).

There was a strong desire to preserve the autonomy of the care recipients, and most of the expansion in care services took the form of home-care in adapted facilities, rather than in nursing homes. This also had cost advantages compared to institutionalized services, as in the latter there is 24/7 access to highly qualified personnel (Løken et al., 2017). The plan was implemented by the municipalities through federal grants. Although all municipalities in principle could apply for the grants, there is evidence suggesting that the municipalities with the lowest home-based care coverage rates were more likely to apply (Borge and Haraldsvik, 2006). We confirm this in a regression of growth in home-based care coverage rates on municipality characteristics in Table 1. We show that municipalities with the lowest pre-reform home-based coverage rates experienced the largest post-reform increase in home-based care coverage. This regression also confirms that there was no relation between the increase in coverage rates and pre-reform institutionalized care coverage, the share of the population aged 80 or older or aged 67 or older, respectively, and the municipal budget.

In Figure 1, we show that the levels of home-based care coverage rates are converging and almost levelling off in the post-reform period. Moreover, we show that the level of pre-reform coverage rates of home-based care is a strong predictor of absolute change in coverage rates from the pre-reform period to the post reform period. In Figure A3, the top graph shows that the coverage level for institutionalized care remained fairly constant across the time period. The same holds for the fraction of the population aged 80 and older in the bottom graph. This is important because we might worry that the change in home-based care coverage rates was offset by changes in institutionalized care, or that individuals who are in need of care might move to municipalities that apply for grants to improve care provision. Additionally, we compare pre-reform characteristics across municipalities with high and low pre-reform coverage rates in Table A1, and conclude that the municipalities are similar on most measures. One notable difference is the tendency that municipalities with lower pre-reform coverage have larger populations.

The descriptive results discussed above confirm the government's stated strategy of emphasizing home-based care compared to nursing homes in combating coverage discrepancies across municipalities (Daatland and Veenstra, 2012). However, Figure 1 captures only the aspect of the reform associated with increasing the proportion of elderly re-

[7] The plan also included goals for quality improvements in the home-based care and existing properties.

ceiving care. The reform may have affected other aspects, such as improved quality of the care provided or more hours of care for those already receiving formal care.[8] The reduced-form effects of the care expansion, thus, are likely to work through all of these channels.

4 Data

We define the baseline sample as adult daughters with no siblings and only one surviving parent aged 80 years or older in each year of the time period 1995-2003.[9] This particular sample of daughters is a group expected to have a greater care burden for an elderly parent, and thus, the group most likely to experience a reduction in care responsibilities caused by the reform. The sample must be seen as repeated cross sections as new daughters enter every year as the remaining parent turns 80 years old, or as a parent becomes a widow(er), provided he or she is 80 or older. Similarly, some daughters drop out of the sample due to the death of the remaining parent.

4.1 Municipality-Level Data

To obtain information on the type and extent of formal care on the municipality level, we use annual data on elder care from the regional database provided by the Norwegian Social Science Data (NSD). These data provide information on the number of users of different types of elder care in each municipality per annum in the time period 1993-2014. Our treatment variable is the fraction of individuals aged 80 and older who receive formal home-based care, i.e., home-based care provided by the municipality.[10] From the regional database, we also assess the fraction of individuals aged 80 or older living in care institutions in each municipality each year. We link individuals to municipality data based on their mother's (or the living parent's) municipality of residence in 1993.

4.2 Individual-Level Data

We apply rich, individual-level register data provided by Statistics Norway (SSB), which covers the entire resident population over the period 1993-2014. These registers include demographic information, such as year of birth, gender, immigration status, municipality of residence, and socioeconomic data, such as education and earnings. In addition, we have information on every individual's income and use of different welfare benefits, including disability insurance benefits. Using unique identifiers, we are able to link sib-

[8] No data are available on quality or hours of care. Thus, we are not able to separate between differing quality across municipalities.

[9] We define daughters with no siblings as daughters of mothers who have only one child.

[10] The data report only the number of individuals who receive care, not the amount of care each individual receives.

lings and parents to their children, and to link individuals to data from relevant health registers, explained in more detail below.

4.2.1 Sickness Absence and Related Diagnoses

Following the National Insurance Act (Folketrygdloven, 1997), all employees are entitled to sickness benefits to compensate for the loss of labor income due to illness or injury. To assess the impact of caregiving on sickness absence and the underlying diagnoses, we apply records of leaves of absence reported to the Norwegian Labor and Welfare Service (NAV). These data contain every sickness absence spell certified by a physician and their related diagnoses from 1995 to 2014. The absence spell is recorded from the first day of a certified absence. A physician's certification is required for all sickness absence spells lasting longer than the allowed number of self-reported days.[11] As a general rule, workers are entitled to at least three self-reported absence days per leave, but in some workplaces workers are entitled to up to eight days.[12] Each sickness absence spell can last for a maximum of 12 months with full wage compensation up to a certain ceiling. Our measure of sickness absence is a binary indicator equal to one if a person has at least one spell of certified absence during the year. We condition sickness absence on employment, which means we drop individuals who are not eligible for reimbursed sickness absence.[13] Being employed is defined as having a labor income exceeding the basic amount (G). The G levels are administratively set and adjusted each quarter of the year, and used in the national social security system to determine old-age pensions and eligibility for disability and unemployment benefits.[14]

All absence certifications should be followed by at least one diagnosis reported by the physician who issued the certificate.[15] In addition to the indicator for sickness absence at all, we construct separate indicators for absences related to the two largest diagnoses categories following a sickness absence: namely musculoskeletal disorders (MSD), and psychological disorders. We group all other diagnoses into a rest category, and finally, we keep as a separate category all leaves of absence with a missing diagnosis.

4.2.2 Long-Term Health Outcomes

For the long-term health outcomes, we use detailed register data from the Control and Distribution of Health Reimbursement database (KUHR), available for the years 2006-2014. These registers cover all Norwegian citizens' utilization of primary health-care services

[11] There is no available national record of self-reported sickness absence days.

[12] Workers who are frequently absent may need certification from day one.

[13] Although this choice could potentially affect the results, we expect that it is of little importance as Løken et al. (2017) estimate no effects of the care expansion on employment at the extensive margin.

[14] In 2006, 1 G represented approximately USD $10,000.

[15] Sickness absence diagnoses follow the International Classification of Primary Care system, 2nd Edition (ICPC-02).

and entail administrative records of all reimbursements claimed by primary physicians. All Norwegian citizens belong to a specific primary physician's list, and each physician is responsible for providing primary health care services to patients that belong to his or her list. As a requirement in the physician's payment scheme, the physician is obliged to report to the national claims database (KUHR) on all services provided and actions taken after every consultation, including the main symptom or diagnosis, referrals and certification of sick leaves. Physicians are required to include at least one diagnosis code per consultation to be reimbursed for the services provided (Sørensen et al., 2016). Classification of main symptoms and diagnoses follow The International Classification of Primary Care (ICPC-02). To assess long-term effects of caregiving on health, we lag the health outcomes by 11 years. As the health data are available for nine years (2006-2014), lagging the data by 11 years ensures that we have enough observations in the pre- and post-reform periods.[16] As illustrated in Figure A2 in the Appendix, this means that, for instance, for individuals in the sample in year 1995 we measure long-term health outcomes in 2006. For individuals in the sample in year 1996, long-term health outcomes are measured in 2007, and so on.

To measure utilization of primary health-care services, we record the yearly likelihood of consulting a primary physician and the number of consultations per year. As primary physician consultations, we include visits to the emergency room and the causality clinics. As discussed in the literature review, caregiving has been linked to a range of health outcomes, as well as to poor lifestyle choices, which can manifest in poor health. We assess the likelihood of having a record of any of the following broad diagnosis categories: MSD, psychological, and cardiovascular disorders. Further, we look at the more specific diagnoses of hypertension and depression. Finally, we measure lifestyle related disorders as having any of the following diagnoses: Non-insulin-dependent diabetes (type 2), overweight or obesity, or substance abuse, that being either alcohol, tobacco, medication or drug abuse. The exact construction by ICPC-02 codes that go into each outcome is listed in Table A2 in the Appendix.

5 Empirical Strategy

To estimate the effects of the expansion in formal care for the elderly on health outcomes for their adult children, we use a reduced form model that exploits the differential increase in the allocation of federal funds across municipalities caused by the 1998 reform.[17] We compare individual outcomes across municipalities that experienced different levels of expansion in formal care. As a predetermined indicator of the intensity of the municipal

[16] As a robustness check, we include an alternative where we lag the health outcomes by 10 years.

[17] Our empirical strategy follows the strategy applied in Løken et al. (2017).

response to the reform, we use the average level of home care coverage measured before the reform, more specifically, in the years 1993-1996. As noted in Section 3, municipalities with lower pre-reform coverage rates experienced a larger expansion in home-based care coverage from the pre- to the post-reform period. Importantly, we rely on informal care responsibilities being related to the amount of formal care provided in the municipalities.[18]

We split the post-1998 period into a phase-in period, defined as the period 1998-2000, and a post-reform period, defined as 2001-2003. In the phase-in period, we expect smaller effects due to delays in the implementation of the reform, whereas in the post-reform period, funding levels had increased, and home-based care coverage rates had almost converged between treatment and control municipalities. We estimate the following regression:

$$\begin{aligned} Y_{it} = \alpha_0 + \alpha_1 PreCoverage_i + \alpha_2 Phasein_t + \alpha_3 Post_t + \alpha_4 \Big(PreCoverage_i \\ \times Phasein_t \Big) + \alpha_5 \Big(PreCoverage_i \times Post_t \Big) + \mathbf{X}'_{it}\delta + \mu_{it}\,, \end{aligned} \tag{1}$$

where Y is the outcome of interest at time t for individual i. $Phasein$ indicates the transition period, defined as 1998-2000, while $Post$ indicates the post-reform period, 2001 to 2003. $PreCoverage$ represents the treatment intensity of the municipality in which the elderly parent of individual i lived in 1993. Finally, X is a set of control variables including municipality fixed effects, child age and education, parent age and gender, and immigrant status for both parent and child.[19]

The coefficients of interest, α_4 and α_5, are interpreted as the intention to treat effect of the policy change in the phase-in period and the post-reform period, respectively. In addition to care expansion at the extensive margin, it is likely that the reform led to care expansion at the intensive margin and to quality improvements, which we are not able to observe. The key identifying assumption is that the change in outcomes from the pre- to post-period would have been the same across municipalities with different levels of pre-reform coverage rates in the absence of the reform. Our approach assumes a linear relationship between pre-reform care coverage rates and the outcomes of interest. As the linearity assumption is not supported for extreme values of pre-reform coverage levels, we drop municipalities with extreme pre-reform coverage levels, defined as lower than the 10th percentile and higher than the 90th percentile.[20]

[18]This relationship, i.e., that individuals in municipalities with higher levels of formal care are less inclined to provide informal care, is confirmed by Jakobsson et al. (2012).

[19]For some specifications in the robustness checks (where we estimate the effects for all daughters, with and without siblings), we include the child's birth order and the number of siblings as controls.

[20]The exclusion of these outliers is supported by Løken et al. (2017).

5.1 Multiple Hypothesis Testing

When testing a number of outcomes that are potentially correlated, and are estimated using the same source of variation, we may increase the risk of accepting an incorrect hypothesis. Therefore, we test whether the estimated effects remain statistically significant after we correct for multiple hypothesis testing. For each outcome in our main tables, we include the p-value obtained using the stepwise procedure described in Romano and Wolf (2005a,b, 2016). As outcomes occurring in the same table are conceptually similar, we test the outcomes in the same table simultaneously.

5.2 Differences by Socioeconomic Status

When estimating the average effect on the entire group of daughters affected by the expansion, difference in responses across important subgroups may be concealed. Of particular interest in this setting is the differences in response across groups with different SES. We use education level as a proxy for SES, and investigate differences in the response to the expansion of elder care across daughters with high and low education by interacting both $(PreCoverage_i \times Phasein_t)$ and $(PreCoverage_i \times Post_t)$ in Equation 1 with an indicator equal to 1 if individual i had less than 10 years of education pre-reform (measured in 1987).

5.3 Alternative Specifications and Robustness

To shed light on the robustness of our results, we perform a number of checks. As our first robustness check, we assess the sensitivity of the estimates to various specification checks. Specifically, we start by excluding large cities and the most rural municipalities that might be very different from the rest of the sample. Further, we include alternative ways of treating the year 1997 (the year before the reform), first by excluding 1997 altogether and second by including 1997 in the phase-in period. Moreover, we include a specification where we split the municipalities into treatment and control groups based on whether the municipalities fall above or below the median in the pre-reform coverage distribution.

Additionally, we re-estimate the effects using an alternative approach where we exclude the phase-in period and treat all years after the reform as post-years. This means estimating the following equation:

$$Y_{it} = \alpha_0 + \alpha_1 PreCoverage_i + \alpha_3 Post_t + \alpha_4 \Big(PreCoverage_i \times Post_t \Big) + \mathbf{X}'_{it}\delta + \mu_{it}\,, \quad (2)$$

where post is defined as 1998-2003. Excluding the phase-in period gives us more post-reform years; however, it also means that the treatment exposure is smaller, on average,

as we include individuals from the phase-in years. Thus, we expect the estimates of α_4 from Equation 2 to be smaller than the estimated post effects, α_5, using Equation 1.

We further estimate the reform effects on alternative samples and run two placebo tests. Specifically, for the alternative samples we consider all daughters and sons with no siblings. These samples are assumed to have lower care burdens and thus, likely to be less affected by the care expansion. For placebo tests, we estimate the effects on samples we assume not to be affected by the care expansion. Here, we consider daughters with no siblings where both parents are deceased and those with younger parents (age 60-72). We exploit these various degrees of treatment exposure as a means to further ensure that the estimated effects stem from the care expansion.

Finally, our choice for the number of lags for the long-term health outcomes is restricted by the fact that only nine years of data are available, and that we need at least some years both before and after the reform to estimate the effects. However, it is not obvious by exactly how many years we should lag the outcomes, and this choice may potentially affect the results. For the final robustness check, we therefore provide results where the health outcomes have an alternative lag, namely, 10 years instead of 11 years.

6 Results

In the following sections, we present the results for the baseline sample of daughters with no siblings and only one living parent aged 80 or older. All results are presented graphically and in tables with the intention-to-treat (ITT) estimation results. For the graphical presentation, we include figures where municipalities are split at the median of the pre-reform coverage rate distribution.[21] We start by presenting results for the short-term effects of the coverage expansion on sickness absence, and assess whether the effect on sickness absence is driven by a specific diagnosis group. We then assess the long-run health outcomes for the same sample. Next, we discuss heterogeneity by education, and finally, we discuss the robustness of the results.

6.1 Sickness Absence and Related Diagnoses

The results for sickness absence and the diagnoses underlying the absence are presented graphically in Figure 2. The first graph shows the likelihood of having a sickness absence at all. We see almost overlapping trends prior to the reform for municipalities with

[21] An alternative version of the figures is to graph the absolute change from the pre- to the post-reform period in the outcome considered against the pre-reform coverage level. Although we do show this continuous version for the first stage figure, to avoid including too many graphs, we show only the binary version of the graphs, i.e., where we split municipalities into treatment and control groups, for the rest of the outcomes. The binary figures have the advantage of a more intuitive representation of the development of the outcomes, and the differences across municipalities with higher and lower pre-reform coverage levels. Additionally, the figures give us a way of visually inspecting the pre-trends.

pre-reform coverage rates above (control) and below (treatment) the median. The graphs show a clear tendency toward divergence in the post-reform period. Figure 2 further shows the probability of absences related to specific diagnosis groups, namely, the two largest diagnosis groups: MSD, and psychological disorders, a rest category containing all other diagnoses, as well as the group in which no diagnosis is specified. For these specific groups of absence spells, we see a clear tendency toward a reduction in absence related to MSD, and, although less clear, a tendency towards a reduction in spells related to psychological disorders in the treatment municipalities compared to the control municipalities. For the rest category and the spells with no specified diagnosis, there seems to be no effect of the care expansion.

Table 2 provides the corresponding ITT estimates for the sickness absence outcomes. The regression results confirm the tendencies observed in the graphs in Figure 2. The phase-in ITT indicates the reform effect in the phase-in period, defined as 1998-2000, while Post-ITT indicates the effect in the post period, defined as 2001-2003, i.e., where we expect stronger effects as the expansion of formal home-based care has had time to take effect. The estimated ITT effect on the likelihood of a sickness absence at all (the first column) is statistically significant at the 5 percent level for both the phase-in period and the post-reform period. Specifically, we estimate that a 10 percentage point lower pre-reform coverage rate is related to a 14 percent reduction in the likelihood of a sickness absence among adult single-child daughters (a 3.0 percentage points reduction from a mean of around 21 percent).[22] As expected, the estimates are larger for the Post-ITT than for the Phase-in ITT effect.

As previously discussed, there are reasons to believe that a sickness absence is granted both due to own health and to gain flexibility to provide care for a frail parent. In this setting, it is therefore interesting to assess whether the reduction in sick leave is driven by a specific underlying diagnosis. We investigate this question in the following four columns of Table 2. Here, we find statistically significant reductions for the post-ITT in the probability of absences related to both MSD and psychological disorders, but no effect on absences related to other diagnoses (the rest category) or the groups of unspecified diagnoses.[23]

[22] Although we use a slightly different sample and use sickness absence data from a different register, these results are in line with the results on sickness absence found in Løken et al. (2017). We replicate their result in Figure A5 and Table A14 applying the same source of sickness absence and the same period of analysis, 1993-2005, as they did. However, to obtain data on the diagnoses connected to each sickness absence spell, we need to use a different register for our analysis (the NAV Register). These data are available only from 1995. Moreover, as our long-term health outcomes are available for only nine years, we need to limit the period under analysis further compared to Løken et al. (2017), and are left with the years 1995-2003.

[23] Testing for statistical differences in these estimates confirms that the estimate for MSD is not equal to the unspecified diagnoses group and the rest category. For absences related to a psychological condition,

As discussed in the literature section, MSD and psychological disorders are health outcomes that are likely to be affected by caregiving. For the probability of having a sickness absence related to MSD, the long-term ITT effects are estimated to be -2.5 percentage points from a mean of 9.2 percent, which translates into an effect of about 27 percent reduction in the likelihood of a sickness absence. Similarly, for the probability of an absence related to a psychological condition, our estimates translate into a 37 percent reduction (a 1.1 percentage points reduction from a mean of 2.95 percent). Finally, as shown by the Romano-Wolf p-values provided in the table, the estimated effects on the probability of an absence at all, as well as the probability of leaves related to MSD and psychological disorders, are still statistically significant at conventional levels when the p-values are corrected for multiple hypothesis testing.

We also investigated more specific diagnoses by splitting up the broad MSD and psychological disorders categories into smaller groups but found no clear pattern of the sickness absence reduction being driven by any of these smaller groups. These results are shown in the Appendix Table A3.

6.2 Long-Term Health Outcomes

The graphical presentations of results for the long-term health outcomes are presented in Figure 3. As previously explained, for this part of the analysis the health outcomes are measured 11 years ahead.[24] In general, the graphical presentation provides no clear evidence of effects in any direction on the outcomes considered. However, we see some tendencies toward a reduction in the risk of hypertension in the treatment municipalities compared to the control municipalities.

The estimation results for the long-term outcomes are presented in Table 3. The Phase-in and Post-ITT are defined as above. The first column presents the reform effects on the probability of a primary care consultation. The point estimate shows an increase of 2.4 percentage points from a mean of 79.7 percent, which translates into an effect of 3 percent.[25] In the remaining columns of Table 3, we investigate the likelihood of the occurrence of specific diagnoses identified in the literature as potentially affected by caregiving responsibilities. Overall, we find little evidence of long-term health being affected by the care expansion. Specifically, we find no effects on the likelihood of the

however, we are not able to reject that the estimates are statistically different for the two categories.

[24]We retain the same baseline sample of adult daughters. However, the number of observations differs slightly as we do not condition on employment, and we drop individuals who were dead in the years in which health outcomes are measured in this part of the analysis.

[25]In Table A4, we show that we find no effects on employment, disability, and mortality. These long-term outcomes are interesting in themselves, and important for this analysis as they are states that may affect the likelihood and regularity of primary care consultations, e.g., as employed individuals need a physician to certify sickness absences.

broad groups of MSD, psychological disorders and cardiovascular diseases, or on the more specific categories of depression and lifestyle related diseases. However, we find a statistically significant negative impact of the care expansion on the likelihood of having a hypertension diagnosis. The estimated effect is statistically significant at the 10 percent level for the Phase-in ITT and at the 5 percent level for the Post-ITT. The effect is also larger for the latter estimate. The Post-ITT shows a reduction in the likelihood of hypertension of almost 11 percent, a 2 percentage point decrease from a mean of 19 percent. However, when correcting for multiple hypothesis testing the calculated Romano-Wolf p-value becomes 0.235, and thus, the effect is not statistically significant at any conventional level.

6.3 Heterogeneity by Education

We then assess heterogeneity in the outcomes among daughters with different levels of education. In Tables 4 and 5, we present the effects estimated using Equation 1, where we include interaction terms between the treatment variables and an indicator for low education. The results for sickness absence and related diagnoses remains fairly similar to those in the main analysis for both educational groups. The effect for the group with high education is represented by the $PostITT$ and the effect for the low educated group is represented by the sum of the $PostITT$ and the $post \times loweduc$ (the p-value represents the joint statistical significance of the Post-ITT and the interaction term).

For the results for sickness absence the estimated reduction in the overall probability of an absence and in the probability of leaves of absence related to MSD and psychological disorders remains and is statistically significant for both educational groups (the p-value for the joint statistical significance of the Post-ITT and the interaction term is 0.095). The only statistically significant difference is that we estimate a slightly larger decrease in leaves of absence related to psychological disorders for the group with low education. While the estimated effect of 1 percentage point is still negative and statistically significant at the 5 percent level for the group with high education, we estimate a 1.3 percentage point decrease for the group with low education.

The heterogeneity results for the long-term health outcomes are presented in Table 5. We estimate a slight decrease in the probability of a primary care consultation for the group with low education, in contrast to the small increase for this outcome for the full sample. However, this effect is very close to zero (-0.2 percentage points).[26] The heterogeneity analysis further shows that the care expansion increased the risk of developing a depression diagnosis, which is not revealed in the main analysis. Specifically,

[26]The table also shows a slight increase in the number of primary care consultations. Though this increase of 0.02 consultations is statistically significant, it is also fairly close to zero considering the average number of consultations is about 3.6.

we estimate a 0.2 percentage point increase for the group with low education.[27] Finally, the heterogeneity analysis shows that the estimated effect on hypertension seems to be driven by the group with higher education as we find an effect of 2 percentage points for this group (the same as in the full sample), whereas the p-value for the joint significance of the Post-ITT and the interaction term is 0.11 yielding a not statistically significant effect for the low education group.

6.4 Robustness

The results from the robustness checks are presented in the Appendix in Tables A5 and A6. In these tables, we include the baseline specification for comparison in the first column, and we show only the results for the post-ITT. We first exclude large cities and the most rural municipalities (in columns 2 and 3, respectively), and we then include alternative ways of treating the year 1997 (the year before the reform), first by excluding that year (column 4) and second by including 1997 as a phase-in year (column 5). Finally, in column 6, we include a specification where we split the municipalities into treatment and control groups based on whether the municipalities fall above or below the median in the pre-reform coverage distribution.

The specification checks for sickness absence spells are shown in Table A5, and for the long-term health outcomes, results are found in Table A6. The estimated decrease in the probability of sickness absences remains relatively robust, except in column 2 where we exclude large cities. Here, the estimate is smaller, but the direction of the effect remains the same. The results are similar to the baseline specification for the other specification checks. The diagnoses that seem to be driving the results on sickness absence, MSD and psychological disorders remain robust throughout all specifications, while estimates for the other categories remain close to zero and are not statistically significant. For the long-term health outcomes, most specification checks show, as in the baseline specification, non-statistically significant effects. The estimated small increase in the likelihood of a primary health-care consultation is robust to most specifications, except when we split the municipalities into treatment and control groups based on median pre-reform coverage in column 6. Last, the estimated decrease in the likelihood of a hypertension diagnosis remains statistically significant and about the same size throughout all the specification checks.

Next, we estimate the reform effects on alternative samples: all daughters and sons with no siblings. These samples are assumed to have lower or no care burdens, and thus, are less likely to be affected by the care expansion. The results for these alternative

[27] Albeit rather small, this finding contrasts with the existing empirical evidence in the literature on the health effects of having care responsibilities.

samples are shown in Tables A7 and A8. We maintain the baseline results in the first column, while the results for all daughters and for sons with no siblings are presented in the following columns. Table A7 shows no effects on either sickness absence or on absence related to specific diagnoses for any of the groups assumed to have fewer care responsibilities. The results for the long-term health outcomes for the alternative samples, presented in Table A8, provide the same picture. Here, too, we find no effects on the sample of sons with no siblings. For all daughters, we estimate statistically significant, yet smaller, effects of the care expansions on the likelihood of a primary health-care consultation and on the likelihood of a hypertension diagnosis 11 years later.[28] On average, this group of daughters is likely to have a smaller care burden than the daughters with no siblings, i.e., lower treatment intensity. Thus, finding results that vary with the potential treatment intensity reassures us that the observed effects stem from reduced caregiving.

We further perform placebo tests by estimating the effects on daughters with no siblings where both parents are deceased and on those with younger parents (aged 60-72). The results for the placebo tests are presented in Tables A9 and A10. As seen in the first table, we find no statistically significant effects of the care expansion on the probability of sickness absence. For the long-term outcomes, we estimate a slightly significant increase in the risk of a cardiovascular disease and of a lifestyle-related disease for the group of daughters with no living parents. As noted by Løken et al. (2017), the group of daughters with no living parents may have care responsibilities for parents-in-law, and thus be affected by the care expansion. However, the direction on the estimated effect in the placebo test contradicts the direction we would expect given the findings in the literature.

Finally, we re-estimate the main effects using two alternative approaches. First, we exclude the intermediate period and treat all years after the reform as post-reform years. Second, we lag the health outcomes by 10 years instead of 11. Results for the specification excluding the intermediate period are presented in Tables A11 and A12. We see that excluding the intermediate period does not alter the conclusions from the main analysis. Though a little smaller, the point estimates for sickness absence, as well as for leaves of absence related to underlying diagnoses, are still statistically significant for absence at all and absences related to MSD and psychological disorders, and small and not statistically significant for the groups of unspecified diagnoses and the rest category. For the long-term health outcomes in Table A12, we no longer estimate the increase in the likelihood of a primary care consultation, as in the main results, but we still find a reduction in the

[28] An estimated increase in the risk of MSD also turns up as statistically significant at the 10 percent level for daughters with siblings. After correcting for multiple hypothesis testing, the effect is no longer statistically significant.

risk of a hypertension diagnosis.

Results from the specification where we apply different lags to the health outcomes are shown in Table A13. Again, we find similar results as in the baseline specification. Overall, there is not much evidence pointing toward long-term health effects of the care expansion, except from a small increase in the probability of a primary care consultation and a decrease in the likelihood of a hypertension diagnosis (here 10 years after).

In sum, the robustness analysis shows that the results found in the main analysis are robust to various specifications and reassuringly, that when assessing the same outcomes for samples less likely to be affected by the reform, apart from a few exceptions, we find essentially no effects.

7 Discussion and Conclusion

Summing up the main findings of this paper, we show that expanding formal care of the elderly led to a decrease in the probability of sickness absence for single-child daughters with only one surviving parent aged 80 or older. This finding is in line with the previous findings in Løken et al. (2017). We expand this analysis and show that the decrease in absences seems to be driven by leaves of absence related to MSD and psychological disorders. These findings remain statistically significant at conventional levels when correcting for multiple hypothesis testing, and they are robust to a range of specification checks. In addition, results for the placebo tests as well as when estimated on alternative samples increase our confidence in these results.

Although the results show that the reduction in leaves of absence is driven by MSD and psychological disorders, this does not necessarily imply reduced own illness alone. An early study of physicians' certifying practice in Norway suggests that the diagnosis stated on the absence certificate is a good reflection of a person's health (Tellnes et al., 1989), and the ICPC-2 coding of absence certifications conducted by Norwegian physicians has been judged by Brage et al. (1996) to be of satisfactory quality. However, several studies have documented that the sickness insurance system in Norway is used partly to insure employees against a broad range of circumstances, other than own illness (Markussen et al., 2011; Carlsen, 2008; Gautun, 2008). Moreover, in recent years there has been a tendency toward increased sickness absence certifications based on more subjective complaints and diffuse symptoms (Brage et al., 2013). The explanations provided include assertions that may imply that the diagnoses related to a sickness spell do not necessarily reflect individual health perfectly.[29]

[29]The explanations discussed by Brage et al. (2013) include that a diagnosis given by a physician may have consequences for an individual, e.g., when purchasing health insurance, and that this makes physicians more careful in their diagnostic practice. A further explanation is that at the initial consultation the diagnosis given often relies primarily on the patients' own complaints, and that some physicians do

In the second part of the analysis, we assess the effects of expanding publicly provided elder care on long-term health outcomes among the sample of daughters. Overall, we find little evidence of long-term health outcomes being affected by reducing the care burden. We investigate long-term effects on primary care consultations and the likelihood of having specific diagnoses detected in the literature as potentially affected by care responsibilities. These diagnoses are MSD, psychological, or cardiovascular disorders, depression, hypertension, and, finally, lifestyle related disorders. We find no statistically significant effects of the care expansion on the long-term health outcomes, except hypertension, where we estimate an 11 percent reduction in the risk of a hypertension diagnosis. To the extent that we can rely on this finding, it can be explained by lower stress levels following a reduced care burden.[30] Despite the well-established link between caregiving and health, we document no causal effects of caregiving on some of the most common health outcomes in this literature, such as depression, psychological disorders, and MSD in the longer term. As a closing comment, given that we estimate short-term effects on sickness absence, an explanation for why we find essentially no effects on long-term health outcomes could be that sickness absences have been used as a preventive measure to avoid severe illness in the future.

not update the diagnoses later when the absence is expanded.

[30]The result on hypertension is robust to a range of specifications, but not to correcting for multiple hypothesis testing, which means that this result needs to be interpreted with caution.

References

Bauer, J. M. and Sousa-Poza, A. (2015). Impacts of informal caregiving on caregiver employment, health, and family. *Journal of Population Ageing*, 8(3):113–145.

Berge, T., Øien, H., and Jakobsson, N. (2014). Formell og uformell omsorg. Samspillet mellom familien og velferdsstaten. *Velferdsforskningsinstituttet, NOVA-notat*, 2014(3).

Bolin, K., Lindgren, B., and Lundborg, P. (2008). Informal and formal care among single-living elderly in europe. *Health Economics*, 17(3):393–409.

Bonsang, E. (2009). Does informal care from children to their elderly parents substitute for formal care in europe? *Journal of Health Economics*, 28(1):143–154.

Borge, L.-E. and Haraldsvik, M. (2006). Empirisk analyse av handlingsplanen for eldreomsorgen. *Senter for økonomisk forskning, SØF-rapport*, 6:2006.

Brage, S., Bentsen, B. G., Bjerkedal, T., Nygård, J. F., and Tellnes, G. (1996). ICPC as a standard classification in Norway. *Family Practice*, 13(4):391–396.

Brage, S., Nossen, J. P., and Thune, O. (2013). Legemeldt sykefravær 2000-2012: Flere blir sykemeldt for symptomer of plager. *Arbeid og Velferd*, (2-2013).

Brevik, I. (2001). Handlingsplan for eldreomsorgen. *Regional stats vurdering av investeringstilskuddene. Noen erfaringer midtveis i byggeperioden. NIBR-notat*, 107.

Capistrant, B. D., Moon, J. R., Berkman, L. F., and Glymour, M. M. (2012). Current and long-term spousal caregiving and onset of cardiovascular disease. *Journal of Epidemiol Community Health*, 66(10):951–956.

Carlsen, B. (2008). Dobbeltmoralens voktere? Intervjuer med fastleger om sykemelding. *Tidsskrift for Velferdsforskning*, 11(4):259–275.

Carmichael, F., Charles, S., and Hulme, C. (2010). Who will care? Employment participation and willingness to supply informal care. *Journal of Health Economics*, 29(1):182–190.

Coe, N. B. and Van Houtven, C. H. (2009). Caring for mom and neglecting yourself? The health effects of caring for an elderly parent. *Health Economics*, 18(9):991–1010.

Cutler, D. M. and Sheiner, L. (1994). Policy options for long-term care. In *Studies in the Economics of Aging*, pages 395–442. University of Chicago Press.

Daatland, S. O. and Veenstra, M. (2012). Bærekraftig omsorg? Familien, velferdsstaten og aldringen av befolkningen. *Velferdsforskningsinstituttet, NOVA Rapport*, 2012(2).

Daatland, S. O., Veenstra, M., and Lima, I. A. (2009). Helse, familie og omsorg over livsløpet. *Velferdsforskningsinstituttet, NOVA Rapport*, 2009(4).

Do, Y. K., Norton, E. C., Stearns, S. C., and Van Houtven, C. H. (2015). Informal care and caregiver's health. *Health Economics*, 24(2):224–237.

Fiva, J. H., Hægeland, T., Rønning, M., and Syse, A. (2014). Access to treatment and educational inequalities in cancer survival. *Journal of Health Economics*, 36:98–111.

Folketrygdloven (1997). Law of February 28th 1997 Number 19: Lov om folketrygd. Online; accessed 1 May 2018 https://www.nav.no/rettskildene/lov/L19970228-19_P8-23#L19970228-19_P8-23.

Gautun, H. (2008). Arbeidstakere og omsorg for gamle foreldre – den nye tidsklemma. *Fafo Rapport*, 40(2008):393–409.

Gautun, H. and Hagen, K. (2010). How do middle-aged employees combine work with caring for elderly parents? *Community, Work & Family*, 13(4):393–409.

Heitmueller, A. and Inglis, K. (2007). The earnings of informal carers: Wage differentials and opportunity costs. *Journal of Health Economics*, 26(4):821–841.

Huseby, B. M. and Paulsen, B. (2009). Eldreomsorgen i Norge: Helt utilstrekkelig-eller best i verden. *Sintef Helsetjenesteforskning*.

Jakobsson, N., Hansen, T., and Kotsadam, A. (2012). Er det en sammenheng mellom formell og uformell omsorg i Norge? *Tidsskrift for Velferdsforskning*, 15(3):168–175.

Jakobsson, N., Kotsadam, A., Syse, A., and Øien, H. (2016). Gender bias in public long-term care? A survey experiment among care managers. *Journal of Economic Behavior & Organization*, 131:126–138.

Kivimäki, M. and Steptoe, A. (2018). Effects of stress on the development and progression of cardiovascular disease. *Nature Reviews Cardiology*, 15(4):215.

Kotsadam, A. (2012). The employment costs of caregiving in Norway. *International Journal of Health Care Finance and Economics*, 12(4):269–283.

Leigh, A. (2010). Informal care and labor market participation. *Labour Economics*, 17(1):140–149.

Lilly, M. B., Laporte, A., and Coyte, P. C. (2007). Labor market work and home care's unpaid caregivers: A systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work. *The Milbank Quarterly*, 85(4):641–690.

Løken, K. V., Lundberg, S., and Riise, J. (2017). Lifting the burden: Formal care of the elderly and labor supply of adult children. *Journal of Human Resources*, 52(1):247–271.

Markussen, S., Røed, K., Røgeberg, O. J., and Gaure, S. (2011). The anatomy of absenteeism. *Journal of Health Economics*, 30(2):277–292.

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, 338(3):171–179.

Mentzakis, E., McNamee, P., and Ryan, M. (2009). Who cares and how much: Exploring the determinants of co-residential informal care. *Review of Economics of the Household*, 7(3):283–303.

Pinquart, M. and Sörensen, S. (2003). Differences between caregivers and noncaregivers in psychological health and physical health: A meta-analysis. *Psychology and Aging*, 18(2):250.

Pinquart, M. and Sörensen, S. (2006). Gender differences in caregiver stressors, social resources, and health: An updated meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 61(1):P33–P45.

Pinquart, M. and Sörensen, S. (2007). Correlates of physical health of informal caregivers: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 62(2):P126–P137.

Romano, J. P. and Wolf, M. (2005a). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Romano, J. P. and Wolf, M. (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Romano, J. P. and Wolf, M. (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters*, 113:38–40.

Schmitz, H. and Westphal, M. (2015). Short-and medium-term effects of informal care provision on female caregivers' health. *Journal of Health Economics*, 42:174–185.

Schmitz, H. and Westphal, M. (2017). Informal care and long-term labor market outcomes. *Journal of Health Economics*, 56:1–18.

Schulz, R. and Sherwood, P. R. (2008). Physical and mental health effects of family caregiving. *Journal of Social Work Education*, 44(sup3):105–113.

Sørensen, M., Arneberg, F., Line, T., and Berg, T. (2016). Cost of diabetes in Norway 2011. *Diabetes Research and Clinical Practice*, 122:124–132.

Stabile, M., Laporte, A., and Coyte, P. C. (2006). Household responses to public home care programs. *Journal of Health Economics*, 25(4):674–701.

Stark, A. (2005). Warm hands in cold age – on the need of a new world order of care. *Feminist Economics*, 11(2):7–36.

Tellnes, G., Svendsen, K.-O. B., Bruusgaard, D., and Bjerkedal, T. (1989). Incidence of sickness certification: Proposal for use as a health status indicator. *Scandinavian Journal of Primary Health Care*, 7(2):111–117.

Toljamo, M., Perälä, M.-L., and Laukkala, H. (2012). Impact of caregiving on finnish family caregivers. *Scandinavian Journal of Caring Sciences*, 26(2):211–218.

Vaage, O. F. (2000). Til alle døgnets tider. *Tidsbruk*, 2000:1971–2000.

Van Houtven, C. H. and Norton, E. C. (2004). Informal care and health care use of older adults. *Journal of Health Economics*, 23(6):1159–1180.

Von Känel, R., Mausbach, B. T., Patterson, T. L., Dimsdale, J. E., Aschbacher, K., Mills, P. J., Ziegler, M. G., Ancoli-Israel, S., and Grant, I. (2008). Increased framingham coronary heart disease risk score in dementia caregivers relative to non-caregiving controls. *Gerontology*, 54(3):131–137.

Figures

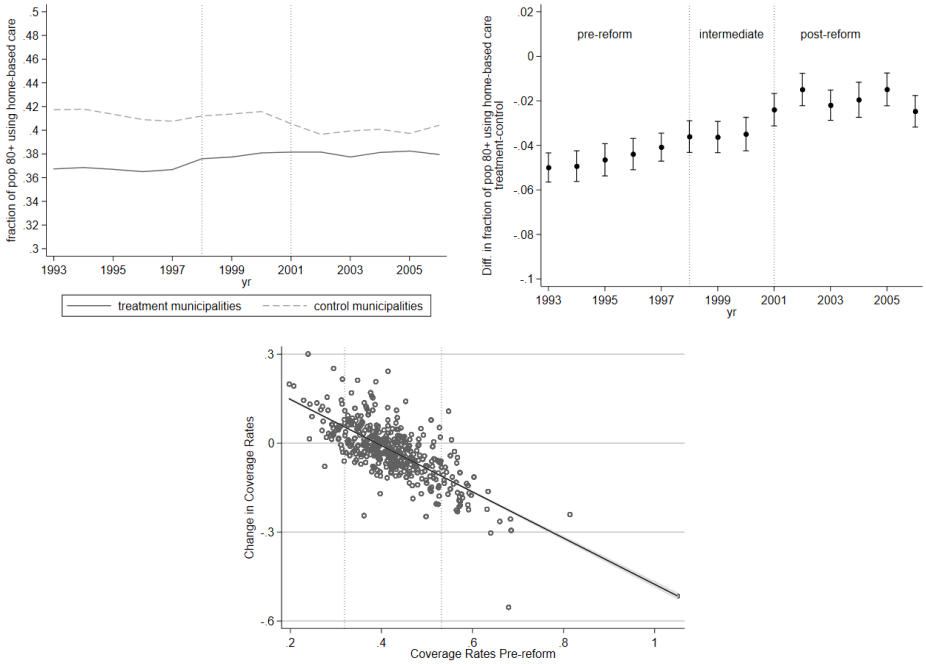


Figure 1: Home Based Care – Coverage Rates

Note: Home based care is defined as care at home or in adapted facilities. In **the first two figures**, treatment is defined as falling below the median home based care coverage rate in the pre-reform period, defined as 1993-1996. The dotted vertical lines represent the reform year 1998, and the year 2001. The period 1998-2000 represents the phase-in period. Municipalities with pre-reform home care coverage rates falling below the 10th and above the 90th percentiles are excluded. **The top left figure** graphs the fraction of the 80+ population using home based care in treated municipalities (solid line) and control municipalities (dashed line) over the period 1993-2005. **The top right figure** shows the differences per year in the fraction of the 80+ population using home based care in treated vs. control municipalities. **The lower figure** graphs the absolute change in coverage rates against the pre-reform coverage rates. In this figure, the dotted vertical lines indicate the 10th and 90th percentiles. Municipality list from the year 2000 (N=435).

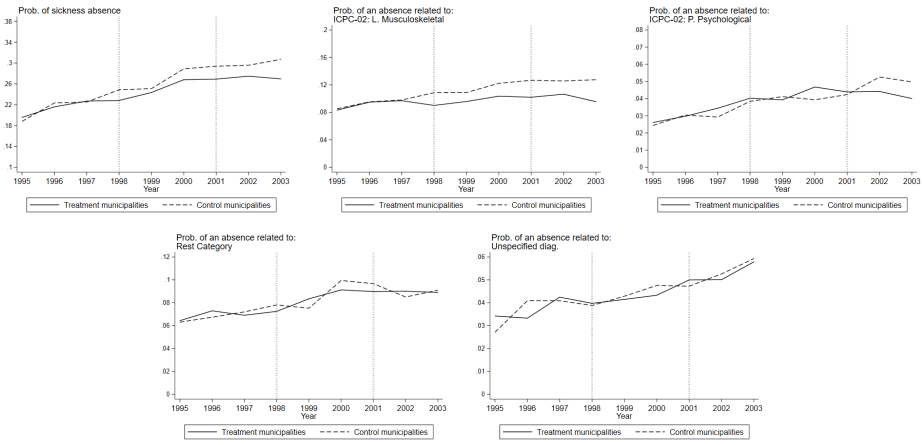


Figure 2: Probability of a Sickness Absence, 1995-2003

Note: The graphs show the probability of a sickness absence spell (cond. on employment), and spells realted to specific diagnoses for daughters with no siblings and with only one living parent at least 80 years old. Municipalities are split into treatment and control groups based on whether they fall below or above the median home based care coverage rate in the pre-reform period, defined as 1993-1996. The vertical lines represent the reform year 1998, and the year 2001. The period 1998-2000 represents the phase-in period. Municipalities with pre-reform home care coverage rates falling below the 10th and above the 90th percentiles are excluded. Municipality list from the year 2000 (N=435).

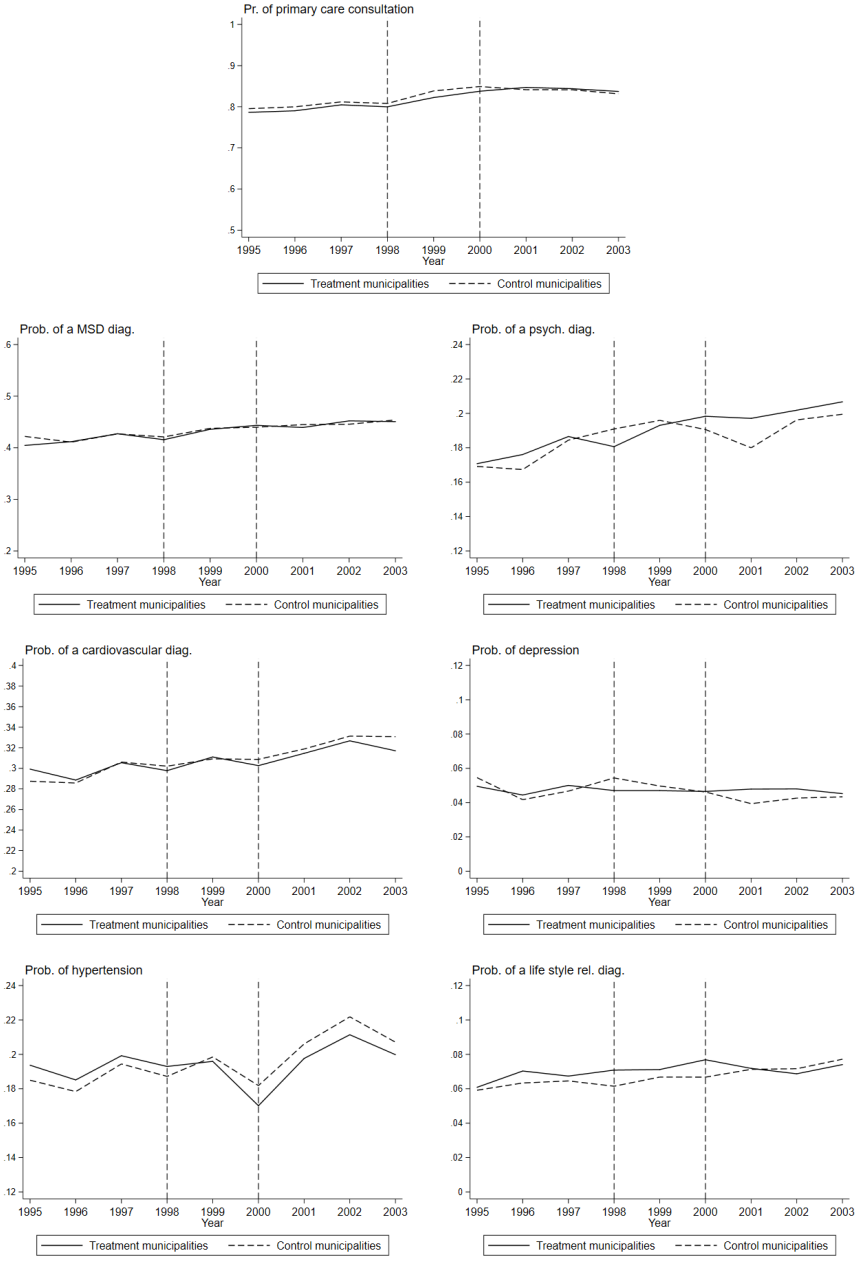


Figure 3: Long-Term Health Outcomes:
Utilization of Primary Health Care Services

Note: The graphs show the probability of a primary care consultation and consultations related to a specific diagnosis (lagged by 11 years) for daughters with no siblings and with only one living parent at least 80 years old. Municipalities are split into treatment and control groups based on whether they fall below or above the median home based care coverage rate in the pre-reform period, defined as 1993-1996. The vertical lines represent the reform year 1998, and the year 2001. The period 1998-2000 represents the phase-in period. Municipalities with pre-reform home care coverage rates falling below the 10th and above the 90th percentile are excluded. Municipality list from the year 2000 (N=435).

Tables

Table 1: Post-Reform Growth in Home Care Coverage Rates

	Abs change in home care cov. rates (post - pre)	
	(1)	(2)
(-) Homecare coverage rate, pop. 80+ (1993-1997) (scaled by 10)	0.068***	0.078***
	(0.0070)	(0.0042)
(-) Inst. based care coverage rate, pop. 80+ (1993-1997) (scaled by 10)	0.001	0.003
	(0.0075)	(0.0069)
Share of pop 67yrs+	0.070	0.032
	(0.1999)	(0.2013)
Share of pop 80yrs+	-0.056	0.040
	(0.5544)	(0.5572)
Munic. unrestricted budget per capita	0.000	-0.000
	(0.0000)	(0.0000)
Exclude extremes	Yes	No
Observations	347	435

Note: The (-) in front of homebased and institution based care means that these controls enter with negative values. Hence, the coefficients for these controls are interpreted as, e.g. in col. 2, that a 10 percentage points *lower* home care coverage rate in the pre-reform period is related to a 7.8 percent increase in the coverage rate from the pre- to the post-reform period. The pre-reform period is defined as 1993-1996, while the post-reform period is defined as 2001-2003 (we define the period 1998-2000 as a phase-in period). Extremes are defined as municipalities with pre-reform home care coverage rates falling below the 10th or above the 90th percentiles in the pre-reform homecare coverage distribution. Robust standard errors in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table 2: Main Results: Sickness Absence – Specific Diagnosis Categories

		Absence with specific diagnosis:			
	Sickness absence at all	Musculoskeletal diag.	Psychological diag.	All other diag.	Unspecified diag.
Phasein ITT	-0.023**	-0.016*	-0.004	-0.000	-0.004
	(0.0116)	(0.0085)	(0.0040)	(0.0069)	(0.0050)
Post ITT	-0.030**	-0.025***	-0.011***	-0.004	-0.001
	(0.0146)	(0.0090)	(0.0040)	(0.0073)	(0.0048)
Mean	[.214]	[.0924]	[.0295]	[.0685]	[.0368]
Romano-Wolf p-value	0.096	0.028	0.036	0.857	0.920
Obs.	57753	57753	57753	57753	57753

Note: Indicators for the Phase-in (1998-2000) and the Post (2001-2003) period are interacted with the negative pre-reform coverage rate, scaled by 10. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01. The Romano-Wolf p-value relates to the Post ITT coefficient, and is obtained using the Romano-Wolf stepwise procedure to correct for multiple hypothesis testing.

Table 3: Main Results: Long-Term Health Outcomes

	Prim. care consultations:		Specific diagnoses (groups):					
	Prob.	No.	MSD	Psych.	Cardio.	Depr	Hyper-tension	Life style
Phasein ITT	0.004	-0.037	0.007	-0.007	-0.012	-0.002	-0.015*	0.008
	(0.0098)	(0.0940)	(0.0099)	(0.0089)	(0.0108)	(0.0050)	(0.0082)	(0.0060)
Post ITT	0.024**	0.058	0.017	0.009	-0.014	0.005	-0.020**	0.001
	(0.0110)	(0.1019)	(0.0121)	(0.0106)	(0.0118)	(0.0055)	(0.0100)	(0.0066)
Mean	[.797]	[3.62]	[.417]	[.177]	[.296]	[.0478]	[.19]	[.0649]
Romano-Wolf p-value	0.235	0.924	0.745	0.892	0.745	0.892	0.235	0.924
Obs.	73963	73963	73963	73963	73963	73963	73963	73963

Note: Health outcomes are lagged by 11 years. Indicators for the Phase-in (1998-2000) and the Post (2001-2003) period are interacted with the negative pre-reform coverage rate, scaled by 10. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01. The Romano-Wolf p-value relates to the Post ITT coefficient, and is obtained using the Romano-Wolf stepwise procedure to correct for multiple hypothesis testing.

Table 4: Heterogeneity by Education: Sickness Absence – Main Categories

	Prob. of sickness absence				
	At all	MSD	Psych	Rest category	Unspecified diag.
Post ITT	-0.031**	-0.027***	-0.010**	-0.004	-0.001
	(0.0147)	(0.0090)	(0.0040)	(0.0074)	(0.0048)
post * low educ	-0.002	0.004	-0.003**	-0.000	-0.001
	(0.0039)	(0.0027)	(0.0015)	(0.0021)	(0.0019)
Joint sign. p-value	.0953	.00442	.00442	.831	.766
Obs.	57531	57531	57531	57531	57531

Note: In the first row an indicator for Post (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The second row shows the coefficients where we interact Post with the negative pre-reform coverage and an indicator for low education, defined as education below 10 years. An interaction with the phase-in period (1999-2000) is included, but not shown in the table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table 5: Heterogeneity by Education: Long-Term Health Outcomes

	Prim care consultations:		Specific diagnoses (groups):					
	Prob.	No.	MSD	Psych	Cardio.	Depr.	Hyper-tension	Life style
Post ITT	0.078	0.024**	0.018	0.009	-0.015	0.005	-0.020**	-0.000
	(0.1012)	(0.0110)	(0.0124)	(0.0106)	(0.0119)	(0.0056)	(0.0099)	(0.0067)
post * low educ	-0.080**	0.000	-0.004	-0.002	0.000	-0.003**	-0.002	0.002
	(0.0379)	(0.0023)	(0.0039)	(0.0024)	(0.0048)	(0.0014)	(0.0042)	(0.0021)
Joint sign. p-value	.105	.0784	.286	.433	.459	.0614	.112	.561
Obs.	73647	73647	73647	73647	73647	73647	73647	73647

Note: In the first row an indicator for Post (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The second row shows the coefficients where we interact Post with the negative pre-reform coverage and an indicator for low education, defined as education below 10 years. An interaction with the phase-in period (1999-2000) is included, but not shown in the table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Appendix

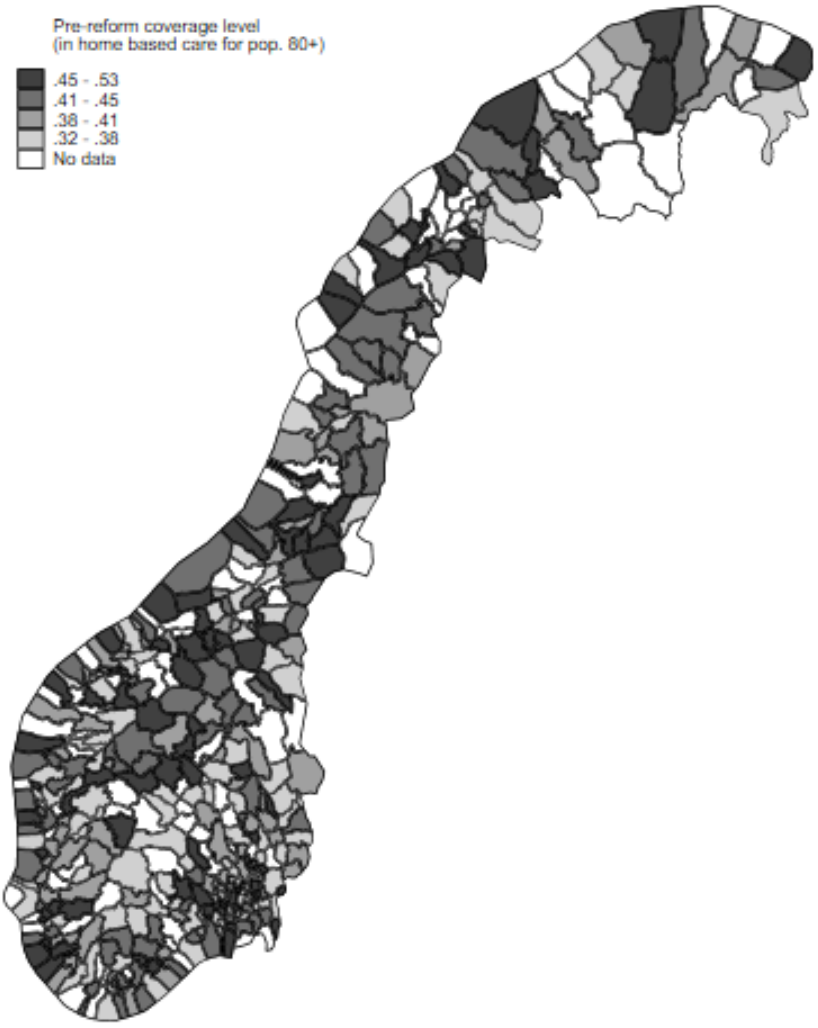


Figure A1: Regional Variation in Pre-Reform Coverage Rates

Note: The map shows variation in pre-reform home care coverage rates across municipalities. Municipalities not in the preferred sample are those with pre-reform home care coverage rates falling below the 10th or above the 90th percentiles of the pre-reform coverage distribution.

Table A1: Municipality Characteristics (1997 numbers)

	Municipalities with pre-reform home based care coverage:	
	Below median	Above median
Total population	12892	8849
	(39619)	(19397)
Share of pop 67yrs+	.159	.157
	(.0334)	(.0368)
Share of pop 80yrs+	.0483	.0486
	(.0146)	(.0156)
Share of males 16+ with compulsory schooling	.374	.386
	(.0764)	(.0789)
Share of females 16+ with compulsory schooling	.424	.435
	(.0732)	(.079)
Share of males 16+ with high school	.47	.473
	(.0507)	(.0531)
Share of females 16+ with high school	.413	.413
	(.0462)	(.0542)
Share of males 16+ with university degree	.11	.103
	(.0345)	(.0323)
Share of females 16+ with university degree	.14	.133
	(.0367)	(.0341)
Share of working age pop. employed	96.3	96.1
	(1.9)	(1.9)
Private income (100,000 NOK)	6967	5674
	(5631)	(5198)
Munic. unrestricted budget per capita	22641	23233
	(7266)	(6356)
Centrality index	3.94	3.58
	(2.51)	(2.54)
Population density	4.29	3.92
	(2.75)	(2.87)
Home based care coverage rate (pop. 80+)	.362	.411
	(.0661)	(.0892)
- Pre reform coverage rate (1993-1997)	.373	.459
	(.0257)	(.0334)
- Change from pre to post	.00893	-.0478
	(.0654)	(.0649)
Inst. care coverage rate (pop. 80+)	.193	.192
	(.0659)	(.0632)
- Pre reform coverage rate (1993-1997)	.214	.205
	(.0652)	(.0588)
- Change from pre to post	-.0422	-.0382
	(.0415)	(.0521)
Observations	173	174

Note: The table shows mean values and standard deviations of municipality characteristics measures in 1997 when no other year is specified. Municipalities are divided into groups based on whether they fall below or above the median home based care coverage rate in the pre-reform period, defined as 1993-1996. Municipalities with pre-reform home care coverage rates falling below the 10th and above the 90th percentiles are excluded.

Table A2: ICPC-02 Classification

Diagnosis	ICPC-2 Code	ICPC-02 Titles
Musculoskeletal (MSD)	Any code with prefix L	Chapter L: Musculoskeletal
Psychological	Any code with prefix P	Chapter P: Psychological
Cardiovascular	Any code with prefix K	Chapter K: Cardiovascular
Depression	P03	Feeling Depressed
	P73	Depressive Disorder
Hypertension	K85	Elevated blood pressure
	K86	Hypertension uncomplicated
	K87	Hypertension complicated
Life style	T07	Weight gain
related diseases	T82	Obesity
	T83	Overweight
	T90	Diabetes non-insulin dependent
	P15	Chronic alcohol abuse
	P16	Acute alcohol abuse
	P17	Tobacco abuse
	P19	Medication abuse
	P19	Drug abuse

Note: The table shows the ICD-02 codes used to classify sickness absence spells related to specific diagnoses, and to construct the long-term health outcomes.

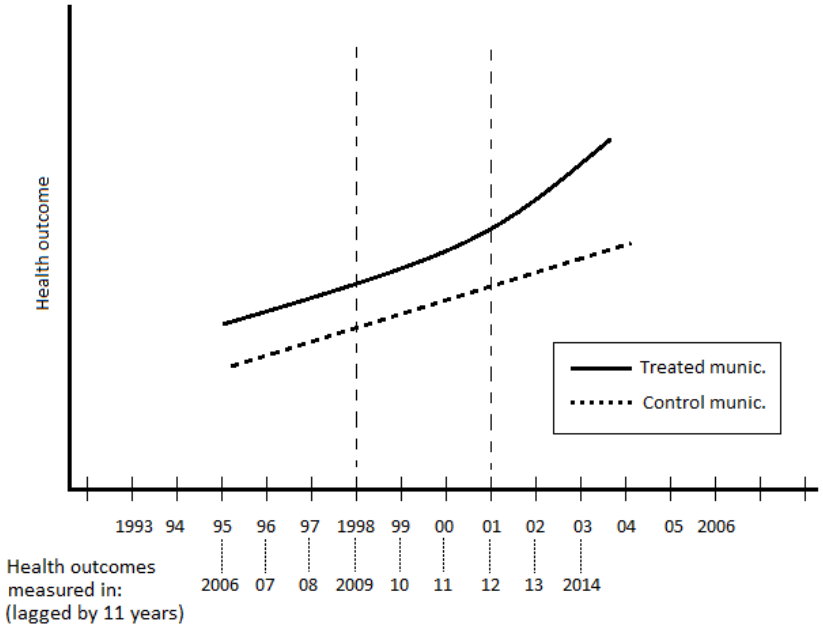


Figure A2: Illustration of the Long-Term Outcomes Design

Note: The figure is an illustration of how our long-term health outcomes are measured. In each year in the period 1995-2003 our sample entails daughters with no siblings who have only one living parent at least 80 years old in the particular year. Long-term health outcomes are measured 11 years later, e.g. for the 1995 sample, long-term outcomes are measured in 2006.

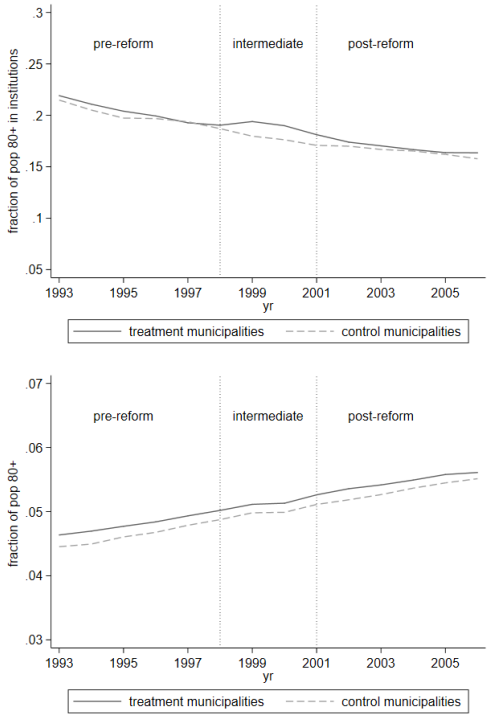


Figure A3: Other Municipality Characteristics

Note: The figures show the development in coverage rates in institution based care, and in the fraction of population aged 80 and above. Institution based care is defined as care in nursing homes. Treatment is defined as falling below the median home based care coverage rate in the pre-reform period, defined as 1993-1996. The dotted vertical lines represent the reform year 1998, and the year 2001. The period 1998-2000 represents the phase-in period. Municipalities with pre-reform home care coverage rates falling below the 10th and above the 90th percentiles are excluded. Municipality list from the year 2000 (N=435).

Table A3: Additional Results on Sickness Absence – Selected Diagnosis Categories

	Sickness absence (selected diagnoses):					
	P01-P29 Psych. symptoms	P70-P99 Psych. diag.	P03;P76 Depression	L01-L29. MSD symptoms	L83-L95;L98;L99 MSD diag.	L72-L81;L96 MSD injury
Phasein ITT	-0.000	-0.004	0.000	-0.006	-0.005	-0.007**
	(0.0029)	(0.0031)	(0.0038)	(0.0040)	(0.0069)	(0.0030)
Post ITT	-0.006**	-0.006**	-0.005*	-0.007*	-0.014*	-0.007**
	(0.0027)	(0.0030)	(0.0032)	(0.0037)	(0.0074)	(0.0033)
Mean	[.00847]	[.0215]	[.0223]	[.019]	[.0615]	[.0145]
Obs.	57753	57753	57753	57753	57753	57753

Note: The table show estimated effects of expanding formal eldercare on sickness absence spells related to diagnoses broken down to more specific categories than in the main results. Indicators for the Phase-in (1998-2000) and the Post (2001-2003) period are interacted with the negative pre-reform coverage rate, scaled by 10. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

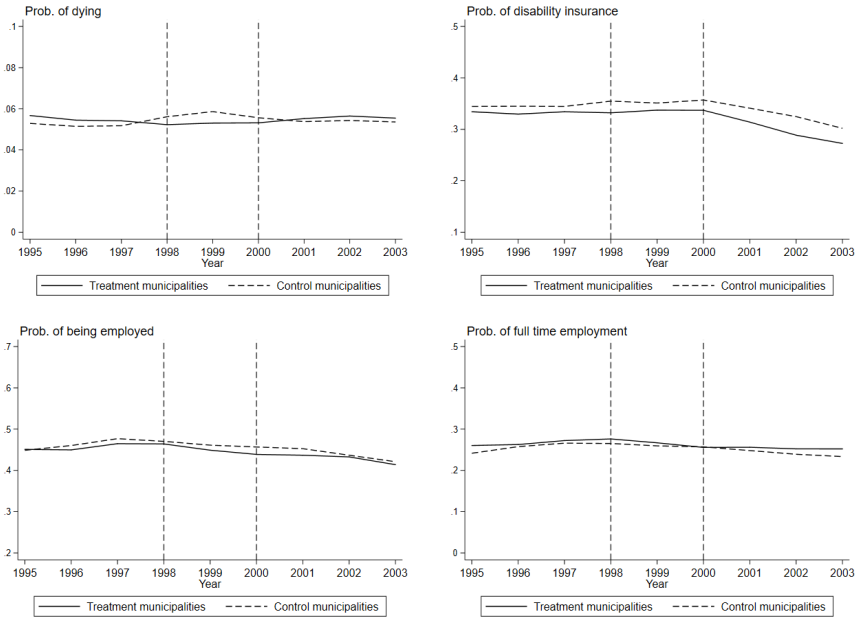


Figure A4: Additional Long-Term Outcomes
Work, disability, and mortality

Note: The graphs show results on long-term outcomes for daughters with no siblings and with only one living parent at least 80 years old. Treatment is defined as falling below the median home based care coverage rate in the pre-reform period, defined as 1993-1997. The vertical line represent the reform year 1998. Municipalities with pre-reform home care coverage rates falling below the 10th and above the 90th percentiles are excluded. Municipality list from year 2000 (N=435).

Table A4: Additional Long-Term Outcomes – Employment, Disability, and Mortality

	Probability of:			
	Dying	Disability insurance	Employment	Full time employment
Phasein ITT	-0.007	-0.004	0.004	0.001
	(0.0049)	(0.0104)	(0.0081)	(0.0091)
Post ITT	-0.002	-0.009	-0.003	-0.001
	(0.0067)	(0.0124)	(0.0104)	(0.0122)
Mean	[.0539]	[.337]	[.458]	[.262]
Obs.	78225	73963	73963	73963

Note: The table show estimated effects of expanding formal eldercare on long-term labor market outcomes. The outcomes are lagged by 11 years. Indicators for the Phase-in (1998-2000) and the Post (2001-2003) period are interacted with the negative pre-reform coverage rate, scaled by 10. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A5: Specification Checks: Sickness Absence - Main Diagnosis Categories

	Daughters, no siblings					
	Baseline	Excl. large cities	Excl. rural munic.	Drop 1997	Incl. 1997 as post	Split at median coverage
Sickness absence at all	-0.030**	-0.024	-0.036**	-0.033*	-0.030**	-0.031***
	(0.0146)	(0.0148)	(0.0155)	(0.0169)	(0.0146)	(0.0110)
Related Diagnoses:						
Musculoskeletal	-0.025***	-0.021**	-0.027***	-0.026**	-0.025***	-0.026***
	(0.0090)	(0.0092)	(0.0097)	(0.0106)	(0.0090)	(0.0064)
Psychological	-0.011***	-0.011***	-0.013***	-0.009**	-0.011***	-0.008**
	(0.0040)	(0.0044)	(0.0042)	(0.0043)	(0.0040)	(0.0037)
All other diagnoses	-0.004	-0.002	-0.007	-0.008	-0.004	-0.003
	(0.0073)	(0.0073)	(0.0077)	(0.0085)	(0.0073)	(0.0064)
No specified diagnosis	-0.001	0.000	-0.001	-0.000	-0.001	-0.001
	(0.0048)	(0.0053)	(0.0050)	(0.0053)	(0.0048)	(0.0038)
Obs.	57753	43292	52770	51878	57753	57753

Note: Indicator for the Post period (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates of the effect in the Phase-in period is excluded in this table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A6: Specification Checks: Long-Term Health Outcomes

	Daughters, no siblings					
	Baseline	Excl. large cities	Excl. rural munic.	Drop 1997	Incl. 1997 as post	Split at median coverage
Primary health care cons.:						
Prob. of cons.	0.024**	0.029**	0.023*	0.029**	0.024**	0.012
	(0.0110)	(0.0114)	(0.0119)	(0.0129)	(0.0110)	(0.0101)
No. of cons.	0.058	0.116	0.028	0.065	0.058	-0.040
	(0.1019)	(0.1032)	(0.1084)	(0.1116)	(0.1019)	(0.0973)
Prim. health care cons. related to specific diagnosis groups:						
MSD	0.017	0.023*	0.010	0.017	0.017	0.002
	(0.0121)	(0.0121)	(0.0125)	(0.0135)	(0.0121)	(0.0110)
Psych.	0.009	0.013	0.002	0.007	0.009	0.006
	(0.0106)	(0.0112)	(0.0106)	(0.0122)	(0.0106)	(0.0096)
Cardiovascular	-0.014	-0.014	-0.018	-0.020	-0.014	-0.013
	(0.0118)	(0.0128)	(0.0126)	(0.0131)	(0.0118)	(0.0105)
Depression	0.005	0.003	0.003	0.008	0.005	0.004
	(0.0055)	(0.0058)	(0.0058)	(0.0059)	(0.0055)	(0.0049)
Hypertension	-0.020**	-0.022**	-0.021**	-0.025**	-0.020**	-0.016*
	(0.0100)	(0.0108)	(0.0105)	(0.0106)	(0.0100)	(0.0084)
Life style related	0.001	0.002	0.003	0.002	0.001	-0.004
	(0.0066)	(0.0071)	(0.0072)	(0.0071)	(0.0066)	(0.0054)
Obs.	73963	56474	67372	66427	73963	73963

Note: Health outcomes are lagged by 11 years. Indicator for the Post period (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates of the effect in the Phase-in period is excluded in this table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A7: Alternative Samples: Sickness Absence – Main Categories

	Daughters		Sons
	No siblings	All	No siblings
Sickness absence at all	-0.030**	-0.004	-0.004
	(0.0146)	(0.0070)	(0.0090)
	[.214]	[.213]	[.154]
Related Diagnoses:			
Musculoskeletal	-0.025***	-0.002	-0.003
	(0.0090)	(0.0046)	(0.0058)
	[.0924]	[.0935]	[.0643]
Psychological	-0.011***	-0.002	0.001
	(0.0040)	(0.0022)	(0.0031)
	[.0295]	[.0287]	[.0162]
All other diagnoses	-0.004	-0.002	-0.003
	(0.0073)	(0.0030)	(0.0054)
	[.0685]	[.0684]	[.0538]
No specified diagnosis	-0.001	0.001	-0.001
	(0.0048)	(0.0022)	(0.0034)
	[.0368]	[.0356]	[.027]
Obs.	57753	290863	89576

Note: Indicator for the Post period (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates of the effect in the Phase-in period is excluded in this table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A8: Alternative Samples: Long-Term Health Outcomes

	Daughters		Sons
	No siblings	All	No siblings
Primary health care cons.:			
Prob. of cons.	0.024**	0.013*	-0.002
	(0.0110)	(0.0078)	(0.0104)
	[.797]	[.793]	[.715]
Number of cons.	0.058	0.032	-0.060
	(0.1019)	(0.0534)	(0.1103)
	[3.62]	[3.51]	[3.06]
Prim. health care cons. related to specific diagnosis groups:			
MSD	0.017	0.011*	-0.003
	(0.0121)	(0.0057)	(0.0095)
	[.417]	[.412]	[.297]
Psych.	0.009	0.001	0.000
	(0.0106)	(0.0050)	(0.0071)
	[.177]	[.169]	[.111]
Cardiovascular	-0.014	-0.002	-0.012
	(0.0118)	(0.0060)	(0.0112)
	[.296]	[.283]	[.343]
Depression	0.005	0.002	-0.000
	(0.0055)	(0.0028)	(0.0033)
	[.0478]	[.047]	[.0273]
Hypertension	-0.020**	-0.011**	-0.005
	(0.0100)	(0.0048)	(0.0089)
	[.19]	[.177]	[.187]
Life style related	0.001	-0.000	-0.000
	(0.0066)	(0.0027)	(0.0061)
	[.0649]	[.0586]	[.0877]
Obs.	73963	357934	97186

Note: Health outcomes are lagged by 11 years. Indicator for the Post period (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates of the effect in the Phase-in period is excluded in this table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A9: Placebos: Sickness Absence – Main Categories

	Daughters, no siblings	
	No living parents	Younger parents
Sickness absence at all	-0.017	-0.034
	(0.0122)	(0.0380)
	[.228]	[.244]
Related Diagnoses:		
Musculoskeletal	0.003	-0.020
	(0.0078)	(0.0275)
	[.0977]	[.108]
Psychological	0.002	-0.006
	(0.0040)	(0.0176)
	[.0285]	[.0378]
All other diagnoses	-0.010	0.014
	(0.0079)	(0.0210)
	[.0777]	[.0797]
No specified diagnosis	-0.007	-0.021
	(0.0048)	(0.0143)
	[.0385]	[.035]
Obs.	70753	6708

Note: Indicator for the Post period (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates of the effect in the Phase-in period is excluded in this table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A10: Placebos: Long-Term Health Outcomes

	Daughters, no siblings	
	No living parents	Younger parents
Primary health care cons.:		
Prob. of cons.	0.006	-0.033
	(0.0114)	(0.0346)
	[.809]	[.781]
Number of cons.	0.044	-0.122
	(0.1113)	(0.3896)
	[3.93]	[3.64]
Prim. health care cons. related to specific diagnosis groups:		
MSD	0.009	0.004
	(0.0120)	(0.0412)
	[.426]	[.413]
Psych.	0.003	0.027
	(0.0090)	(0.0380)
	[.178]	[.213]
Cardiovascular	0.018*	-0.010
	(0.0109)	(0.0374)
	[.356]	[.261]
Depression	0.001	0.016
	(0.0046)	(0.0245)
	[.0482]	[.0621]
Hypertension	0.009	0.001
	(0.0103)	(0.0303)
	[.228]	[.172]
Life style related	0.013**	-0.018
	(0.0051)	(0.0231)
	[.076]	[.0754]
Obs.	105503	8144

Note: Health outcomes are lagged by 11 years. Indicator for the Post period (2001-2003) is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates of the effect in the Phase-in period is excluded in this table. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A11: No Intermediate Period: Sickness Absence – Main Categories

		Absence with specific diagnosis:			
	Sickness absence at all	Musculoskeletal diag.	Psychological diag.	All other diag.	Unspecified diag.
Post ITT	-0.027**	-0.021***	-0.007**	-0.002	-0.002
	(0.0113)	(0.0071)	(0.0034)	(0.0063)	(0.0042)
Mean	[.214]	[.0924]	[.0295]	[.0685]	[.0368]
Obs.	57753	57753	57753	57753	57753

Note: Post covers the entire post-reform period, 1998-2003, and is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A12: No Intermediate Period: Long-Term Health Outcomes

	Prim. care consultations:		Specific diagnoses (groups):					
	Prob.	No.	MSD	Psych.	Cardio.	Depr	Hyper-tension	Life style
Post ITT	0.014	0.012	0.012	0.001	-0.013	0.001	-0.018**	0.005
	(0.0098)	(0.0921)	(0.0098)	(0.0090)	(0.0104)	(0.0048)	(0.0081)	(0.0057)
Mean	[.797]	[3.62]	[.417]	[.177]	[.296]	[.0478]	[.19]	[.0649]
Obs.	73963	73963	73963	73963	73963	73963	73963	73963

Note: Health outcomes are lagged by 11 years. Post covers the entire post-reform period, 1998-2003, and is interacted with the negative pre-reform coverage rate, scaled by 10. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Table A13: Alternative Lags: Long-Term Health Outcomes – Primary Health Care Services

	Prim. care consultations:		Specific diagnoses (groups):					
	Prob.	No.	MSD	Psych.	Cardio.	Depr	Hyper-tension	Life style
Phasein ITT	0.013	-0.033	0.002	-0.010	-0.015	0.001	-0.007	0.002
	(0.0100)	(0.0884)	(0.0105)	(0.0094)	(0.0121)	(0.0051)	(0.0094)	(0.0059)
Post ITT	0.027**	0.029	0.016	0.007	-0.019	0.008	-0.017*	-0.001
	(0.0120)	(0.1112)	(0.0121)	(0.0109)	(0.0124)	(0.0056)	(0.0101)	(0.0070)
Mean	[.791]	[3.56]	[.41]	[.174]	[.279]	[.0477]	[.179]	[.0644]
Obs.	77324	77324	77324	77324	77324	77324	77324	77324

Note: The table show estimated effects on long-term health outcomes when outcomes are lagged by 10 years (we lag by 11 years in the main specification). Indicators for the Phase-in (1998-2000) and the Post (2001-2003) period are interacted with the negative pre-reform coverage rate, scaled by 10. The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

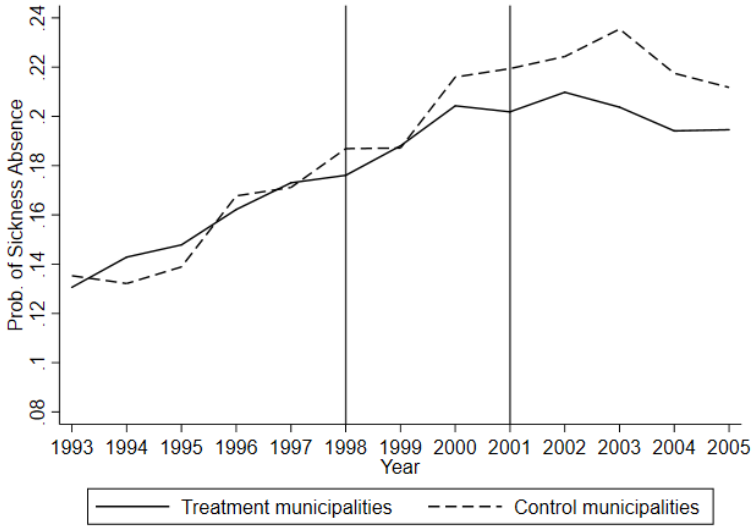


Figure A5: Replication Results on Sickness Absence (16 days+)

Note: The graph shows the replication of the sickness absence result for daughters with no siblings and with only one living parent at least 80 years old from Løken et al. (2017). Treatment is defined as falling below the median home based care coverage rate in the pre-reform period, defined as 1993-1996. The vertical lines represent the reform year 1998, and the year 2001. The period 1998-2001 represents the expansion period. Municipalities with pre-reform home care coverage rates falling below the 10th and above the 90th percentiles are excluded. Municipality list from year 2000 (N=435).

Table A14: Replication Results on Sickness Absence (16 days+)

	Probability of sickness absence	Probability of sickness absence (cond. on work)	Number of sick days (cond. on work)
Main sample			
Daughters, no siblings			
Phasein ITT	-0.011	-0.020	-2.609
	(0.0101)	(0.0128)	(1.8557)
Post ITT	-0.024**	-0.028**	-2.695
	(0.0105)	(0.0139)	(2.0629)
	[.146]	[.194]	[20.4]
Obs.	101785	70948	70948

Note: Sickness absence from the FD-trygd register. Indicators for the Phase-in (1998-2000) and the Post (2001-2003) period are interacted with the negative pre-reform coverage rate, scaled by 10. Note that the period indicators are called "short-term" and "long-term" in Løken et al. (2017). The estimates are interpreted as the effect of having a 10 percentage points lower pre-reform coverage rate, which corresponds to having a 7.8 percentage points increase in the coverage rate from the pre- to the post-reform period. Mean values (pre-reform) in brackets and standard errors clustered at the municipality level in parenthesis, * p<0.1, ** p<0.05, *** p<0.01.

Chapter 3:

Health Effects of Retirement: Evidence from Survey and Register Data

Health Effects of Retirement: Evidence from Survey and Register Data

Maja Weemes Grøtting * Otto Sevaldson Lillebø †

May 6, 2019

Abstract

Using a local randomized experiment that arises from the statutory retirement age in Norway, we study the effect of retirement on health across socioeconomic status. We apply data from administrative registers covering the entire population and from survey data of a random sample to investigate the effects of retirement on acute hospital admissions, mortality, and a composite physical health score. Our results show that retirement has a positive effect on physical health, especially for individuals with low socioeconomic status. We find no retirement effects on acute hospitalizations or mortality in general. However, our results suggest that retirement leads to reduced likelihood of hospitalizations for individuals with low socioeconomic status. Finally, we show that the positive health effects are driven by reduced pain and reduced health limitations in conducting daily activities. Our findings highlight heterogeneity in the health effects across socioeconomic status and across subjective and objective measures of health.

JEL Classification: H75; I14; I18; J26
Keywords: Retirement; Health; Socioeconomic Status; Gender; Regression Discontinuity Design

*NOVA, Oslo Metropolitan University; email: maja.grotting@oslomet.no
†Department of Economics, University of Bergen; email: otto.lillebo@uib.no

1 Introduction

With increasing life expectancy, the number of retired individuals as a share of the total population is rising in most OECD countries. This has led to concerns about the fiscal sustainability of public pension systems, and to policy initiatives that aim to prolong working lives and increase retirement age. An important issue that seems to be overlooked in policy debates over these reforms is the impact that prolonged working lives has on health, and especially if there are heterogeneous retirement effects by socioeconomic status (SES).

Findings in the empirical literature regarding the health effects of retirement are mixed. Some studies report positive effects (Coe and Zamarro, 2011; Eibich, 2015), whereas others report negative effects (Behncke, 2012; Mazzonna and Peracchi, 2017) or no effects (Hernæs et al., 2013; Heller-Sahlgren, 2017). Although some studies highlight the importance of (SES) in these health effects (Coe and Zamarro, 2011; Eibich, 2015), there is limited evidence from formal tests assessing whether effects differ by SES. Another limitation in the literature is that most studies assess retirement effects in the early 60s, an age threshold that is substantially lower than proposed policies to postpone retirement toward age 70. Finally, most studies rely on survey data or administrative records (of sub-samples of the population), which often imply subjectivity in the health outcomes or small sample issues.

In this paper, we investigate the health effects of retirement across socioeconomic status and gender in Norway by applying both survey and register data, where the latter cover the entire population. We assess the health effects of retirement at age 67, which was the statutory retirement age in 2007. This is a higher age threshold than what has previously been studied. To control for individuals self-selecting into retirement, we exploit that the statutory retirement age causes a discontinuous change in the likelihood of retiring at the exact timing of eligibility. This implies a local randomization around the retirement eligibility age threshold, and makes a regression discontinuity (RD) framework suitable. RD entails comparing the health outcomes for those right above the statutory retirement age threshold (i.e. the treatment group) to those right below (i.e. the control group). This allows for identification of the causal short-term effects of retirement on health.

Most studies in this field rely on survey data with the well-known limitations related to non-response and recall bias. Furthermore, while measures of subjective health provide important insights into how individuals experience and rate their own health, such measures have been criticized for being contextual, and can suffer from justification bias (see e.g. McGarry (2004) for a thorough discussion). Another possible concern is that

survey data of older adults is especially prone to health related selection, as poor health correlates with non-response.

The Norwegian register data is attractive with respect to overcoming some of these concerns. In particular, the register data covers the entire population and contain health records that can be considered objective. Still, measures of health from public registers are often extreme outcomes in the sense that they measure very poor health, such as mortality and acute hospital admissions, and hence unsuited for studying moderate health effects. In addition to records of mortality and acute hospital admissions from public registers, we include a composite measurement of self-assessed health from a representative sample of Norwegian older adults (The NorLAG Panel Survey (Slagsvold et al., 2012). This measure is the physical component of the short form-12 (SF-12) health survey (see (Ware Jr et al., 1996). We assess both the overall physical score and the specific components that goes into the SF-12.

We believe that our health measures, collectively, will provide important insight into the multidimensional effects of retirement on health. Moreover, both data sources (the register data and the NorLAG data) contain exact birth month and retirement date from public registers, ruling out recollection bias. Finally, having monthly records allows for a more precise estimation of the effects of retirement on health, as it enables a more local estimation around the timing of retirement compared to analyses using data on the year level.[1]

Socioeconomic status is important in the analysis of health effects of retirement because it determines the kind of work situation an individual retires from. Higher education and white collar jobs are often less physically demanding and associated with greater autonomy and control over the work situation, compared to low SES jobs (Case and Deaton, 2005; Mazzonna and Peracchi, 2012). Moreover, Case and Deaton (2005) argue that manual labor jobs, associated with low education and low income, are more "wear and tear" types of jobs, in which health deteriorates at a more rapid pace than for individuals in "non-manual" professions.

According to the (Grossman, 1972) model of health demand, individuals with low education or low financial capital (low SES) will have to rely more heavily on their health as an input in the labor market, compared to individuals with higher SES, as the different sources of capital are substitutes in the labor market. This is typically manifested through strenuous manual labor for the low SES groups. Moreover, individuals with higher education are assumed to be more efficient in promoting their own health. In sum, the two mechanisms make it more costly for low SES groups to continue working.

[1] See Dong (2015) and Lee and Card (2008) for a discussion of why age in years might yield inconsistent results unless properly accounted for.

Retirement can therefore be seen as a mechanism that levels health inequalities between SES groups. As SES can be an important factor in analyses of retirement and health, we systematically assess how the health effects differ by socioeconomic status.

The RD application in this study identifies the short term health effects of retirement. On the one hand, we can expect to see short term effects on health as the relief from strenuous physical work or the relief from working in a stressful environment is an instantaneous change of circumstance. On the other hand, retirement may lead to a reduced sense of purpose before new routines has been developed (Rohwedder and Willis, 2010).

Our results show that retirement yields a sizeable and positive effect on physical health. This effect is especially strong for the low SES group, whereas we find no effects for the high SES group. We find no effects on mortality or acute hospitalizations in general. However, for the low SES group, we find that retirement leads to a reduction in the likelihood of acute hospitalizations. Our results show that SES is important when studying the effect of retirement on health, but we find no gender differences. Moreover, we find that the reason why retirement leads to better physical health is due to reduced pain and a lower likelihood of reporting that physical health is a limitation in completing both "daily" tasks and "specific tasks profoundly". The results for physical health and mortality are robust to a wide range of robustness and specification checks, whereas the results for hospital admissions are less robust, and must therefore be interpreted with some caution.

The paper proceeds as follows. Section 2 provides a review of previous research and describes the institutional structure of the Norwegian pension system. Section 4 describes our empirical strategy. In Section 3, we present the data, outcome variables, and some basic summary statistics. Our main results are presented in Section 5, and Section 6 concludes.

2 Earlier Literature and Institutional Setting

Our paper is related to a growing body of economic research about the effect of retirement on health. Given the important aspect of this issue and the amount of literature on the topic, there is a surprising lack of consensus across studies. One reason for this is that a large fraction of the existing evidence reports correlations rather than well-identified causal effects. Lately, there has been an increasing amount of well-identified studies, most of which apply exogenous variation in the retirement eligibility as sources of identification. As the majority of these studies apply survey data or administrative records for subsamples of the population, we contribute to the literature by providing objective health outcomes for the entire Norwegian population.

One of the most cited related studies is Coe and Zamarro (2011). They study the ex-

tent to which retirement affects measures of self-reported health and a composite health index across several European countries using the Survey of Health, Ageing and Retirement in Europe (SHARE) data. They find that retirement reduces the likelihood of reporting bad self-rated health and leads to an improvement in a composite measure of subjective health.

From the US setting, Neuman (2008) uses age-specific retirement incentives as instruments for retirement. Applying data from the Health and Retirement Study (HRS), he provides evidence of retirement being both preserving and improving for self-rated health. He argues that since retirement removes the time constraint induced by labor market participation, more time can be devoted to activities that both preserve and enhance individuals' health. This is in line with Grossman's model of health demand, where it can be shown that especially time-intensive workouts may be more attractive after retirement, when the opportunity cost of participating in such activities drops.

Insler (2014) uses data from HRS, and apply workers' self- reported probabilities of working past ages 62 and 65 as instruments. He finds that retirees experience positive effects on a health index, which consists of both objective and subjective measures of health. Moreover, he finds that retirees tend to reduce smoking and participate more in health-enhancing activities.

However, not all studies have shown retirement to have such a positive impact. Using data from the English Longitudinal Study of Aging (ELSA), Behncke (2012) reports that retirement actually increases the risk of being diagnosed with a cardiovascular disease[2] and cancer. Also contradictory to the findings of the aforementioned studies, she finds that retirement increases the probability of reporting poor health, and the risk of being diagnosed with a chronic condition.

Bound and Waidmann (2007) apply measures of self-assessed and objective health from the ELSA study, and find that retirement leads to a small, but significant positive effect on physical health for men. Physical health entails self-assessed health, physical functioning and biomarkers. Moreover, they show that these results are highly sensitive to job characteristics and differences in socioeconomic status. As these differences arguably play an important role in determining the effect of retirement on health, there has recently been a growing interest in tackling these heterogeneity issues. To the best of our knowledge, only a small number of studies have investigated the presence of heterogeneity across SES or gender in the effects of retirement on health.

Mazzonna and Peracchi (2017) stress the importance of heterogeneity in the health effect of retirement, and argue that the previous literature have failed to detect the

[2]Retirement is also found to have an impact on increased obesity (Godard, 2016; Rohwedder and Willis, 2010).

potential heterogeneity. Using the SHARE data, they find that for people working in more physically demanding jobs, retirement has an immediate beneficial effect on both a health index of self-reported measures and cognition. For the rest of the workforce, however, retirement has negative long-term effects on health and cognition.

In the paper closest to our study, Eibich (2015) applies a regression discontinuity framework, to study the effect of retirement on several subjective measures of health in Germany. The empirical evidence suggests the presence of effect-heterogeneity by socioeconomic status. Whereas he uncovered no effect of retirement on health for individuals with higher education, individuals who retire from strenuous jobs seem to experience a large and positive change in physical health.

From the Norwegian setting, Hernæs et al. (2013) employ a stepwise introduction of early retirement ages in Norway in the 1990s as instruments to assess whether retirement age matters for mortality. They find no relationship between lowering early retirement age and mortality up through age 77.[3] Moreover, they question whether retirement has a causal impact on mortality.

Based on the relevant literature, it is unclear to what extent and in what direction retirement affects health. Previous findings are characterized by differences in methodology, be it an instrumental variable approach, regression discontinuity approach, or difference-in-difference approach. Another aspect of the literature is the different outcomes of health. While self-rated physical health often is positively associated with retirement, others document a decline in mental health and cognitive abilities.

2.1 Institutional Setting in Norway

This section provides background information on the institutional setting in Norway in 2007/2008.[4] We start with a brief description of the pension system, as this is the main focus of our study. An individual can start claiming retirement pension the first month after reaching the statutory retirement age of 67, and is, in our analysis, considered retired once this claim is made. The main provider of retirement pension is the mandatory public National Insurance System (NIS). This is a pay-as-you-go defined benefit system, and all individuals with a minimum number of years of residence are covered. Once retired, the pension consists of a mix between fixed earnings-independent basic pension and pension contributions based on previous labor market income. Replacement rates from annual earnings have been found to be around 72% on average (Røed and Haugen, 2003).

In theory, the statutory retirement age did not force individuals to retire. However,

[3]Early retirement in Norway was introduced at age 65, but later reduced in a stepwise matter to age 62. The authors exploits this stepwise reduction as a source of exogenous variation.

[4]The pension system was reformed in 2011, but none of the new rules was in place throughout our study-period.

most companies had contracted retirement upon reaching the statutory retirement, and the norm was that people retired once they hit this age threshold. Moreover, for most of the workforce there was little economic incentive to prolong working life once eligible for old age pension. There was a full earnings test in place for individuals aged between 67 and 69 for earnings above 2 basic amounts,[5] resulting in a 40% reduction of the old age pension for each dollar earned.[6]

Besides the statutory retirement age, there are two other commonly-used exit routes from the labor market: disability insurance (DI) and the Early Retirement Program (ER). These are early exits routes that are temporarily available until the statutory retirement age. Eligibility for DI is based on health status and must be certified by a physician based on a permanent reduced ability to work. DI can also be graded in a way that allows individuals to combine work and DI. ER was available for all public and about half of private sector workers from age 62.[7] At 67, recipients of DI and ER are automatically transferred to retirement pension.

Table 1 summarizes the labor market status for individuals aged 56-79 in 2007. This table shows the fraction of individuals who are either working, on ER, DI, or claiming retirement pensions. The shares do not summarize to unity because it is possible for the same individual to be in two states, e.g. by combining partial uptake of DI and working.

Table 1 shows two important preconditions for our empirical analysis: labor market participation rate remains relatively high for older workers in Norway, and most individuals start claiming pensions as soon as they reach the age of 67. Provided the strong link between retirement pension uptake and exit from employment, we argue that claiming retirement pension in practice means withdrawing from the labor market. Strictly speaking, in this analysis, we are estimating the intention-to-treat (ITT) effects of offering retirement pension at age 67. Because uptake of pension in practice means withdrawal from the workforce for the majority of the population, we assume that the health effects to a large degree will stem from the relief from work related tasks. We refer to claiming retirement pension as retirement in the remainder of this article.

3 Data

3.1 Data

We use data from two separate sources in our analysis. The first is a survey carried out on a representative sample of Norwegian older adults, and the second is comprised of administrative health and population registers covering the entire population. Unfortu-

[5] One basic amount is the lowest earnings required to accrue pension points.
[6] This was lifted in 2008 for 67 year-olds.
[7] See Hernæs et al. (2013) or Kudrna (2017) for more details about the ER system.

nately, individuals from the two sources cannot be connected, as the first data source has been anonymized.

The NorLAG Survey Data

The first datasource is a survey carried out on a representative sample of Norwegian older adults, the Norwegian Study on Life-Course, Aging and Generation (NorLAG) panel study. [8] The data was collected in 2002 and 2007. NorLAG contains individual data on a range of health outcomes, as well as information about socioeconomic status. Data collection was carried out by Statistics Norway with computer-assisted telephone interviews (CATI).

All respondents to the survey are merged with administrative registers for the period 2002-2012. The registers contain information on year and month of birth and of retirement. Furthermore, the registers contain various sociodemographic background information such as labor income, social insurance take-up, and educational attainment. We are thus able to construct detailed information for each individual regarding attachment to the labor market, retirement status and social security take-up, enabling identification of the exact timing of retirement, and whether the individual retired directly from the labor force or transitioned from disability insurance or other welfare programs.

Currently, the panel consists of two waves. For the main analyses, we use the second wave as this contains a larger sample than the first wave.[9] However, for some specifications in the sensitivity analysis, we rely on data from the first wave to obtain information about past labor market performance. This is outlined in more detail in Section 3.2.

Our measure of health from the NorLAG data is a composite measure of physical health, namely the physical component of the Short Form 12 (SF12) scale (Ware Jr et al., 1996). Self-rated health (SRH) is one of the components that go into the SF12. Other factors are measures of the degree to which an individual is able to perform tasks like vacuuming, moving a table or climbing stairs, whether there are certain tasks that could not be performed due to health limitations, or whether pain limits daily activities. The score is standardized on a scale from 0-100 with a mean of 50 and standard deviation of 10 using the US population as a reference. SF12 has been found to be a strong predictor of hospitalization, future health care expenditures, and mortality (see e.g. Ware Jr et al. (1996); DeSalvo et al. (2009); Lacson et al. (2010). Occupational status in the NorLAG data is coded in accordance with the ISCO-88 scale. This has been re-coded into two occupational groups: manual and professional workers, following the classical division

[8] See Slagsvold et al. (2012) for a thorough description.

[9] The first wave contains 5,559 observations (response rate 67%), whereas the second wave contains 15,149 observations (response rate 60%).

into blue and white collar workers of higher and lower skills.[10] Professional workers are defined as high skilled white collar workers, the term "manual workers" refers to three categories: high and low skilled blue collar workers and low skilled white collar workers. We apply this categorization of manual workers, because the latter three groups are more similar based on observable characteristics.

Register Data

Our second data source is comprised of administrative registers that covers the entire Norwegian population. All residents are assigned to a unique personal identification number, which enables them to link information from various administrative registers, such as health registers, income and social insurance registers, and population registers. These registers contain information on year and month of birth, death and retirement, as well as educational attainment, income, and social security uptake.

We apply two measures of health from the register data. The first is a binary indicator of whether a person has been acutely hospitalized in a particular month. This information comes from the national patient register (NPR), which contains records of all inpatient and outpatient stays at Norwegian hospitals from 2008 through 2014. Admissions are coded by whether the hospitalization is a result of a planned or unplanned admission. The latter can be thought of as acute in the sense that treatment has been deemed necessary, typically as a result of an accident, stroke, or severe heart condition.[11] The second health outcome is a binary indicator of whether a person passed away in a particular month. This information comes from the Norwegian cause-of-death registry, and contains all recorded deaths in Norway from 1992 through 2014. Both outcomes thus yield the likelihood of the particular outcome at a specific age-in-month.

Importantly, these measures of health are not correlated with the time cost to consult medical expertise. As individuals have more time at their disposal after retirement, the opportunity cost of seeking medical help is reduced once retired compared to when working. It is therefore likely that the prevalence of a diagnosis or a medical treatment that is not acute increases after retirement, when the opportunity cost of seeing a physician has fallen. Applying a health outcome that is correlated with the opportunity cost of medical consultations can therefore erroneously lead to the conclusion that retirement caused the increased prevalence of the health outcome.

[10] Coded according to NACE Rev.1.1.

[11] All admissions are coded in accordance with the International Statistical Classification of Diseases and Related Health Problems, ICD-10, (see WHO (1992)).

3.2 Sample Selection

We restrict our attention to individuals aged 56-79 in 2007 and 2008 in both data-sources. From the administrative records, we use data from 2008.[12] This leaves 4,619 individuals in the NorLAG sample and 892,908 individuals in the register sample. The register data in our analysis is a panel data set, with monthly records of hospitalization, mortality, retirement, and age in months. As such, month by month, the treatment variable is determined according to age in months exceeding the retirement age threshold. Including fixed effects is unnecessary for identification in an RD design. Moreover, as the source of identification is a comparison between those just below and just above the threshold, which can be carried out with a single cross-section, imposing a specific dynamic structure introduces more restrictions without any gain in identification (see Lee and Lemieux (2010)). We therefore treat the sample from the register panel data as repeated cross-sections and pool all months together, treating each observation as an individual. This also makes the register data more comparable to the NorLAG data.

In order to maintain the intention to treat in the RD design and to ensure that we have enough data for inference, we place no further restrictions on the sample for the main analysis. This means that our analytical sample will include individuals on DI or individuals who are not working for other reasons. Individuals on DI are automatically classified as retired once they hit the age threshold. In theory, we should expect no retirement effects for this group, as their work status remains unchanged when they retire. This would bias our results towards zero. However, the health outcomes in the survey data can suffer from justification bias. Being on disability insurance might make an individual, consciously or subconsciously, under-report their health in order to justify their status as disabled. The need for this justification is no longer present once they are transferred to retirement pension. In this case, the estimates would be biased upwards and we might worry that the positive effect on health was driven by these individuals. As a sensitivity analysis, we therefore run the whole analysis including only individuals who were gainfully employed or working until retirement.

Ideally, we want to compare individuals working up to retirement age to individuals who retired from working. In the NorLAG data, this is done by adjusting the sample by two rules. The first rule implies including only individuals who had income from labor the previous year in the analysis; the second rule implies including only individuals who have stated that they are working or were working before they became retired. Some caveats are worth mentioning: the first rule results in a substantial reduction in the sample size, as we need to use the balanced panel from both waves of the NorLAG study to identify

[12] This is the earliest year in which data on hospitalizations are available.

labor income in 2006. A potential concern with the second rule is that the formulation of the question to the working and retired part of the population differs slightly in the NorLAG data. To maintain continuity across the retirement threshold, it is crucial that we apply exactly the same selection rule on either side of the threshold when identifying the sub-samples for the sensitivity analyses. In the register data, we define individuals as working if they currently have positive income or if they had positive income before retirement. We find that these sensitivity analyses does not alter our conclusions.[13]

3.3 Descriptive Statistics

Table 2 displays summary statistics for the sample from both the NorLAG data and the register data. These are men and women aged 56-79 in 2007 and 2008 respectively.

The first two columns are summary statistics for the whole sample, whereas the next two columns show the summary statistics for those within the bandwidth of 10 months below and 10 months above the retirement threshold of 805 months (age 67). These are the observations within the bandwidth used for estimating the short-term retirement effects in the regression analysis. It is important that the two groups are balanced with respect to the covariates. T-tests (not shown) confirm that individuals on either side of the threshold are similar with respect to education, living arrangements and occupation.

4 Empirical Strategy

4.1 Regression Discontinuity Design

We investigate the impact of retirement along several dimensions of health. Ideally, we seek to investigate the following linear relationship between health and retirement:

$$Health_i = \beta_0 + \beta_1 Retirement_i + X_i'\beta_2 + \varepsilon_i, \tag{1}$$

where $Retirement_i$ is a dummy variable equal to one if the individual has retired and zero otherwise and X_i is a vector of relevant covariates. If retirement were to be considered a random event, Equation (1) would provide us with an unbiased estimate of the effect of retirement on health. However, people typically decide themselves when to retire. Moreover, unobservable factors such as knowledge about own longevity or other factors that correlate with both health and the retirement decision remain unaccounted for in Equation (1). This cause omitted variable bias in β_1. Importantly, own health is likely to affect retirement, causing bias in β_1 due to reverse causation. In order to circumvent these issues in the OLS specification, we apply regression discontinuity design (RD).

[13]The results from the sensitivity analysis are shown in the Appendix.

RD exploits institutional settings that determine access to a treatment. The idea is that treatment (retirement) is determined by a running variable (age), reaching a known threshold (the statutory retirement age). Units above the threshold receive the treatment and units below the threshold do not receive the treatment. This means that we use age as an allocation mechanism that determines retirement, rather than using actual retirement behaviour. The RD design relies on local identification by comparing individuals' right above and right below the retirement age cut-off. The discontinuity gap in health at this point identifies the treatment effect. Since the probability of retirement is discontinuous at the cut-off age 67, we assume that reaching this age limit is what causes individuals to retire. Importantly, this assumption only holds for individuals close to the cut-off on the age distribution.

As described in Section 2, the general rule was that individuals started claiming retirement pensions at the statutory retirement age of 67. However, about 16 percent of men and 13 percent of women within the eligible age groups chose to retire early through ER, and a small fraction retired later. This is a setting of imperfect compliance. The Fuzzy RD (FRD) design is therefore more appropriate. Unlike in the Sharp RD, where all treated units are compliers, i.e. the likelihood of treatment goes from zero to one at the threshold, the fuzzy RD allows for a smaller discontinuity in the probability of retirement at the threshold.[14]

4.2 Estimation

The FRD design resembles a setting with instrumental variables, with retirement coefficients consistently estimated by using two stage least squares (2SLS) (Imbens and Lemieux, 2008). The treatment effect is to be interpreted as a local average treatment effect (LATE), i.e. the estimated treatment effect of retirement on health, for individuals induced by the age threshold to retire (Hahn et al., 2001). In the setting of imperfect compliance with the treatment, the intention-to-treat (ITT) is as if randomized, which implies a causal interpretation of the estimated coefficients. The estimated effects are interpreted as the health effects of offering retirement pension at age 67.

Formally, we instrument for retirement using age equal to, or above the retirement threshold at 805 months, the month after which an individual turn 67 years of age. Specifically, we estimate the following two equations:

$$Retirement_i = \gamma_0 + \gamma_1 1[Age_i \geq c] + \gamma_2 Age_i^B + \gamma_3 Age_i^A + u_i, \tag{2}$$

[14]The difference between sharp and fuzzy RD is parallel to the difference between a randomized experiment with perfect compliance and a randomized experiment with imperfect compliance, when only the intention to treat is randomized.

where the endogenous regressor $Retirement_i$ is a binary variable equal to one if the individual is retired, i.e. is claiming retirement pension. $1[\bullet]$ is an indicator function taking the value one if the condition inside the brackets is true, and zero otherwise. c represents the retirement eligibility threshold at 805 months (age 67). Age is measured in months, and we include continuous age controls. These are allowed to have different slopes at either side of the threshold. Superscript B refers to ages below the retirement threshold at age 67, and superscript A refers to ages above the threshold.

The first stage in this 2SLS set-up is actual retirement predicted by age exceeding the threshold, controlled for the general effect of age on retirement. We apply retirement as predicted in the first stage, and the second stage is given by:

$$Health_i = \beta_0 + \tau \widehat{Retirement} + \beta_1 Age_i^B + \beta_2 Age_i^A + e_i, \tag{3}$$

here, $Health_i$ represents the different health measures for individual i. Our parameter of interest is τ, and its estimate is the jump in the outcome variable at the threshold, divided by the fraction induced to take up treatment at the threshold. This is the estimated treatment effect of retirement on health, for individuals induced by the age threshold to retire.

As the health effects in the RD design is only identified close to the retirement threshold, the estimations are done locally around the threshold. We choose the optimal bandwidth, i.e. how many months on either sides of the age cut-off to include in the estimation,[15] in a cross-validation procedure suggested by Imbens and Kalyanaraman (2012). This is designed to minimize the mean squared error, and provides a trade-off between bias and variance. Based on this bandwidth selector, we choose a bandwidth of 10 months.[16] This means that only individuals in the age range 795 months to 815 months (10 months before and 10 months after the retirement age threshold) are included in the estimations.[17] In the sensitivity analysis, we assess different bandwidths to check the sensitivity of the results with respect to choice of bandwidth. In addition to assessing different bandwidths, we perform a range of robustness checks. Here we follow the guide to practice by Imbens and Lemieux (2008) for robustness checks using the RD design.

[15] Dong (2015) show that using regression discontinuity design calls for careful consideration of the unit of measurement when age is the forcing variable, as age in years, as opposed to age in months, might lead to inconsistent results.

[16] The optimal bandwidth suggested by Imbens and Kalyanaraman (2012) varies by SES-group. The suggested bandwidth is in the range 8-12 months for all the groups. For simplicity, we apply a bandwidth of 10 months in all estimations. Choosing different bandwidths within this interval has little influence on the estimated effects. See the robustness checks in the appendix for more on sensitivity of bandwidths.

[17] Due to the small sample size left in the survey data when we apply the 10 months bandwidths, we also ran the entire analysis using a bandwidth of 20 months. This does not change the results from the survey data in any substantial way.

These results are presented in the appendix, but we discuss them briefly in Section 5 (Results).

Finally, in the cross-sectional survey data, we follow Lee and Card (2008) and cluster at the age group level. As noted by Lee and Card (2008), for RD applications where the running variable is discrete, estimating a parametric function away from the discontinuity point can be seen as a form of random specification error. This implies a common component of variance for all the observations at any given value of the running variable. Thus, they suggest clustering at the age group level to account for this imperfect fit, as clustering leads to wider confidence intervals. In the panel data from the administrative records, we cluster at the individual level to account for the within-person correlation in the error term. The structure of these data will be discussed in more detail in the next section.

5 Results

5.1 Graphical Results

To motivate the use of the FRD design, Figure 1 displays the share of retired individuals from age 55 until age 79. The two upper graphs are constructed using the survey data, whereas the two lower graphs are constructed using the register data. The age span in the four graphs are the same (55-79), but the x-axis on the two left graphs depicts age in years, whereas the x-axis on the two right graphs depicts age in months. The latter is to show that the discontinuity in retirement coincides with the first month after turning 67 (the first month of retirement eligibility).

In all of the four figures, the patterns are very similar.[18] There is a substantial discontinuity in the likelihood of being retired at age 67 (805 months). Since some workers chose to retire early, we also see a small discontinuity at age 62, the lowest eligible age for early retirement. Only a negligible share of individuals chose to retire later than age 67. The graphical evidence thus show a clear response in terms of retirement at the statutory retirement age. We build our empirical analysis on the discontinuity at age 67.

Figure 2 presents graphical evidence on the relationships between health and age for the three outcomes used in our study: physical health, acute hospital admissions, and mortality. The age range spans from 55 to 79 years, and the x-axes are depicted as age-in-months relative to the retirement age threshold at 805 months, normalized to zero. The lines are fitted on either side of the threshold using a second order polynomial global

[18] In the graphs, retirement refers only to those who have actually retired, either through the early pension program or at the retirement age of 67. This means that individuals on DI are not considered retired. If we remove all individuals that are currently on DI or who were on DI before they retired from our sample, the picture looks the same.

fit.

The upper graph (a) in Figure 2 shows the observed health pattern for physical health for all individuals aged 56-79 in the NorLAG sample. Physical health declines with increasing age, but there is a substantial jump at the retirement threshold. At this threshold, the trajectory shifts up to a level of someone 80 months younger, which amounts to 6.5 years.

For acute hospitalizations and mortality, the two lower graphs, (b) and (c) respectively, we see that the incidence rate increases across the age-span 56-79, but there does not seem to be any substantial discontinuities in the outcomes reflected in the graphs. For acute hospitalizations, we see a small, possible negligible, downward shift at the threshold.

There is an ongoing debate as to whether it is the cumulative or contemporaneous effects of retirement that are the largest (see Coe and Zamarro (2011); Mazzonna and Peracchi (2017)). As mentioned above, the effects estimated using RD are only identified close to the threshold, so any prolonged retirement effects becomes mere speculation in this setting. However, by visual inspection of the graph for physical health, (a) in Figure 2, there is suggestive evidence of a prolonged effect of retirement on physical health, as retirement shifts individuals to a higher health trajectory, where they seem to stay as age increases.

5.2 Regression Results

We present the 2SLS regression results for all three health dimensions in Table 3 - Table 6. The effects are estimated using a bandwidth of $+/-$ 10 months around the threshold, which is the optimal bandwidth using the selector suggested by Imbens and Kalyanaraman (2012). We estimate the effects for each gender and for the different SES-groups separately. In Table 7, we present results from a formal test of heterogeneous retirement effects in which the instrument is interacted with indicators of the different SES-groups.

In Table 3, we present the first stage of the 2SLS regression results. This is the estimated effect of crossing the statutory retirement age on the probability of retirement, i.e. τ from Equation (2). The results in Table 3 show that crossing the statutory retirement age significantly increases the probability of retirement, thus indicating a strong first stage. These results are in line with the graphical results presented in Figure 1.

5.2.1 The Effect on Physical Health

Table 4 displays the results of the short-term retirement effects on physical health. We find that retirement leads to a 5.7 points increase in physical health for the population as a whole. This is a substantial effect given that the mean and standard deviation for this health outcome is 47 and 10 points, respectively. We find a strong and positive effect

for men (8 points), and a positive (4 points), but not statistically significant, effect for women. Our findings are in line with evidence from Coe and Zamarro (2011) and Eibich (2015), who suggest that, in general, retirement leads to an increase in physical health in both the USA and Germany. Although our estimates are short-term effects, previous findings suggest that retirement also has a cumulative effect on physical health through increased physical activity (e.g., Eibich (2015)).

Based on the discussion in the introduction, we can expect different health effects of retirement depending on education and occupation. The four latter columns in Table 4 show the effects for the different SES-groups. For the manual workers and low educated groups, the effects are large (13.2 and 8.4 points respectively) at about one standard deviation, and significant at the 1 percent level. For the high SES groups, we find no statistically significant effects, and the coefficients are closer to zero.

These results are in line with the findings of Eibich (2015). He shows that highly educated individuals benefit less from retirement in terms of self-reported health, compared to individuals with low SES. Moreover, Insler (2014) suggests that wealthy people have more time to invest in their health while working.

Power calculations show that a sample of at least 90 is needed to ensure a power of 0.8. Allthough well above this threshold, the sub-group samples are fairly small. It could be argued that this should lead to the application of wider bandwidths. However, wider bandwidths also imply more bias (Lee and Lemieux, 2010). We did, however, run the whole analysis using a bandwidth of 20 months. This about doubles the observations in each sub-group, but the effects sizes and significance levels remains fairly the same.

To sum up, the results are clear in that retirement leads to better physical health for men, and for the low SES groups. For women, the results are similar in effects size, yet statistically not significant. We find no health effects of retirement for the high SES group. Based on this analysis, there does not seem to be substantial differences by gender, but both the gender difference and the differences by SES will be formally assessed in Section 5.2.5.

5.2.2 The Effect on Acute Hospitalization

We now turn to our estimates from the register data. Acute hospitalization is based on a dummy for inpatient care in which treatment is deemed necessary. The results are presented in Table 5.

First, we explore how retirement affects acute unscheduled hospitalizations for the population on average and by gender. For all sub-groups the effect size is about -0.4 percentage points, but not significant. When we divide by SES, we find that retirement leads to a 0.6 percentage point reduction in the likelihood of acute hospitalization for the

low SES group. As the incidence of acute hospitalizations is 14 percent, this amounts to a 4 percent reduction in the likelihood of acute hospitalizations. The effect is significant at the 5 percent level. For the high SES group, we find an effect of 0.3, yet this is not significantly different from zero.

One way to think of these results is that retirement for the population in general leads to no short-term change in serious health-conditions. Hallberg et al. (2015) studied a targeted early retirement offer to workers in the military at age 55 and find that the number of days in inpatient care is significantly reduced at ages 61-70. One possible drawback with our method is that the regression discontinuity design only captures the short-term effect of retirement, and any potential gain of retirement is possibly not found in the subsequent months after retirement. For instance, Hallberg et al. (2015) find a 4.7 days reduction in inpatient care 6-10 years after early retirement, whereas the estimated effect is 2 days in the first years after early retirement.

To some extent, the same intuition can be found in Behncke (2012). She shows that retirement increases the risk of being diagnosed with a chronic condition in the subsequent years after retirement. However, assessments applying less acute diagnoses can be confounded for two reasons. First, the opportunity cost of seeking medical help is greatly reduced after retirement, hence increasing the likelihood of detecting such conditions. Second, the reason for seeking medical help can differ for individuals who are working and individuals who are retired. In Norway, for example, sickness absence from work for longer than the self certified absence period[19] must be certified by a physician, which means that retirees and employers most likely visit the doctor for different reasons.

5.2.3 The Effect on Mortality

The results described in the previous sections suggests that retirement leads to a short-term positive effect on subjective measures of health, whereas we find no or small effects on the number of acute hospitalizations. Given the latter findings, *a-priori*, we expect to see little or no short-term effect on mortality. In the lower panel of Table 5, we display the estimation results on mortality.

We find no short-term effect of retirement on mortality. Regardless of gender and subgroup, the estimates remain statistical indistinguishable from zero.

The question remains whether a short-term effect of retirement on relatively serious outcomes, such as mortality, is implausible in the short run. Hallberg et al. (2015) use cox-regression models to form hazard ratios and find that early retirement at age 55 reduces the risk of dying at age 70 by around 26 percent. Studying the first five years after an

[19] A medical certificate is required for spells of absence of more than three days or eight days, depending on whether the employer has signed the "IA-agreement" or not.

early retirement window in Holland, Bloemen et al. (2017) find a drop in the probability of dying of around 2.6 percent. The same effect is found in Blake and Garrouste (2013) and Kuhn et al. (2010), albeit the latter only for male blue-collar workers. However, studying the introduction of early retirement in Norway, Hernæs et al. (2013) find no effect of early retirement on mortality. They follow workers for a maximum up to 77 years of age, with eligibility for early retirement varying between 62 and 65 years of age. They conclude that early retirement in itself has no effect on mortality.

Taken together, our results show that in general there are no effects of retirement on serious health outcomes. However, as this study and several other studies show, retirement affects subjective health. What is it about these outcomes that actually makes people feel better? In the next section, we look further into the subjective physical health measure (SF-12) to assess which aspects of health that are improved by retirement.

5.2.4 Looking Further into the Effect on Physical Health

SF-12 is composed by survey responses to the following [20]: rate your health on a scale from 1-5 (self-rated health); is your health of such a character that it limits you in doing tasks like moving a table, vacuuming, hiking or gardening; is your health of such a character that it limits you from climbing several flights of stairs; has your physical health limited you in doing your daily tasks so that you have accomplished less than you wished for; has your physical health limited you from completing specific tasks; has psychological problems limited you from doing daily tasks so that you have accomplished less than you which for; has psychological problems limited you from doing these tasks as profoundly as usually; has pain limited you from doing your daily tasks; have you been feeling calm and harmonious, energized or sad during the last four weeks; and, finally, has physical or mental health limited you from socializing as much as you wanted.

Out of the 12 components that go into the SF-12, five were significantly impacted by retirement. These are the following: is your health of such a character that it limits you in doing tasks like moving a table, vacuuming, hiking or gardening (Functional); has your physical health limited you in doing your daily tasks so that you have accomplished less than you wished for (Daily); has your physical health limited you from completing specific tasks (Specific); has psychological problems limited you from doing these tasks as profoundly as usually (Mental); has pain limited you from doing your daily tasks (Pain). Each question is coded as a binary variable, where one means that health or pain is experienced as limiting. In Table 6, we present the results for these four components.

Retirement was found to reduce the experience that physical health is a limiting factor in accomplishing as much as one would like, and as a limiting factor in doing specific tasks.

[20] Translated from Norwegian by the authors.

The former holds for both men and women, whereas the latter holds for men. We find particularly strong effects on reduced pain, especially for women. Furthermore, we find that, in general, retirement reduced the limitations in doing tasks profoundly experienced due to mental health.

When we assess the different SES-groups we find that it is manual workers or lower-educated individuals who experience reduced pain and limitations from physical and mental health. We find no effects for the high SES groups. Moreover, when we divide the groups by SES, we also find that, for the low SES group, retirement reduced the limitation caused by health in doing functional tasks such as vacuuming, moving a table, hiking, or gardening. These effects are statistically significant at the 5 percent level for manual workers and at the 10 percent level for the low-educated group.

5.2.5 A Formal Test of Effect Heterogeneity

Table 7 presents the results from the formal test of heterogeneity. These are the results of a reduced form of Equation (3), where the instrument is interacted with SES groups and gender. We estimate the following:

$$Health_i = \beta_0 + \gamma 1[Age_i \geq c] \times SES_i + \beta_1 1[Age_i \geq c] + \beta_2 Age_i^B + \beta_3 Age_i^A + e_i, \quad (4)$$

where γ is the coefficient of interest and $1[Age_i \geq c]$ is the instrument indicating whether age in months is equal to or exceeds the threshold. SES is a binary indicator of either manual workers, low education or women. We apply the same +/- 10 months bandwidth in these estimations.

We see that the effects of retirement are statistically different from each other when SES is measured by occupation. Although the estimated effects differ quite substantially by educational group as shown in Table 4, the differences are not statistically significant when SES is proxied by education. Moreover, there are no statistically significant differences in the retirement effect by gender. Hence, we show that accounting for differences by socioeconomic status can be important in analyses of retirement effects on health.

5.3 Robustness Checks and Sensitivity Analysis

The results from the robustness checks are presented in the Appendix, but we provide a brief overview here. First, we show that our results on physical health and mortality are robust to different bandwidths, whereas increasing the bandwidth from 10 to 15 months yields significant, negative effects on the likelihood of having an acute hospitalization. The effects are still small, ranging from 0.7 to 1 percentage points, yielding a 5-7 percent reduction in the likelihood of an acute hospitalization. Increasing the bandwidth increases the likelihood of factors, other than retirement, affecting acute hospital

admissions. Another explanation can be that it takes some time for retirement to take effect on health issues such as stroke and acute heart conditions, thus including more post-retirement months increases the likelihood of finding significant effects.[21]

We then look for discontinuities at the retirement age threshold in a covariate that is not affected by the treatment, in this case marital status. Although, retirement can affect the likelihood of being married, it is highly unlikely in the immediate aftermath (within 10 months) of retirement. We find no retirement effects on the likelihood of living with a partner or spouse (the NorLAG data) or on being married (the register data).

Next, we perform placebo tests by checking for discontinuities in the health outcomes at values of the forcing variable and age where there should be no discontinuities. We find no discontinuities in the health outcomes at the placebo thresholds of age 61 and 73 for physical health or mortality, but we find some inconsistencies at these thresholds for acute hospitalizations. The effects are smaller than at the retirement threshold, yet significant, thus we might worry that this outcome is prone to be discontinuous at arbitrary age-thresholds.

We then test for discontinuities in the conditional density of the forcing variable to avoid self-selection or sorting into treatment or control groups. The RD design may be invalid if individuals just above the threshold are more likely to answer a survey than those just below the threshold, i.e. violating the RD assumption that the running variable is continuous at the threshold. In the Appendix, we provide histograms that display the age-in-months-distribution in the NorLAG data. There is no apparent discontinuity at the threshold in these histograms. Moreover, we applied the local polynomial density estimator for testing the null of continuous density of the forcing variable at the threshold proposed by Cattaneo et al. (2016). The p-value under this test is 0.3251.

Finally, the results for physical health and mortality are robust to the different sub-samples that are conditioned upon working or working until retirement, as described in Section 3.2. For acute hospitalizations, we find the same results as in the main analysis for all sub-groups, except for the lower SES group, where the negative impact of retirement on the likelihood of acute hospitalization is no longer found when we condition on working or working until the retirement age.

6 Conclusion

Whether retirement has a causal effect on health is a difficult question to answer because of selection into retirement. In this paper, we study the short-term health effect of

[21] When we run the entire analysis using a bandwidth of 20, we find larger and (negative) significant effects for the population as a whole, for men, and for the low educated group for this outcome. Effects sizes range from 1 to 1.5 percentage points, significant at the 5 percent level. Using this bandwidth, we still find no significant effects of retirement on mortality.

retirement using the statutory retirement age at 67 in a fuzzy regression discontinuity design. We exploit the fact that once individuals reach the statutory retirement age, the probability of claiming retirement pension drastically increase. We apply both subjective measures of health from survey data and objective health outcomes from register data, where the latter covers the entire Norwegian population.

We find that, on average, in the population, retirement has a positive effect on self-assessed physical health, but no effects on the objective measures of health: acute hospitalizations and mortality. When we assess the effects by different SES groups, we find that retirement has a large, positive impact on physical health and leads to reduced likelihood of acute hospitalization among the low SES groups. We find no significant effects for the high SES groups for any of these outcomes. For mortality, we find no significant effects for any group.

We thus confirm what has been found in several studies, namely that retirement has a positive effect on health for subjective health outcomes. How this manifests to objective outcomes is less clear as there exist little evidence using objective health outcomes, especially on the full population. In general, we find no effects on the objective outcomes, besides suggestive evidence of a retirement effect on reduced likelihood of hospitalizations for the low SES group. However, this result does not pass the robustness tests, and must therefore be interpreted with care.

We thus conclude that retirement mainly impacts subjective outcomes, not objective ones. However, we have only assessed extreme objective outcomes such as acute hospitalizations and mortality. When we assess the factors that go into the physical health measure, SF-12, we find that the positive health effect was driven by a few different factors. On the one hand, finding that retirement leads to reduced likelihood of reporting that health is limiting in managing in daily chores and in conducting specific chores profoundly, can be due to the fact that work (a possible health consuming chore), is no longer present, so health feels less limiting. This implies that the underlying health has not changed, but the presence of health consuming activities has. On the other hand, we also found that retirement reduced the presence of pain and reduced the likelihood of reporting difficulties with activities such as vacuuming, moving a table, hiking, or gardening. This indicates that retirement affects health in a more fundamental way. Future research should thus assess objective measures of health that are less extreme. In doing so, it is key to recognize that retirement necessarily coincides with reduced opportunity cost of time.

This study accentuates the importance of assessing the potential heterogeniety in the effects for individuals in different circumstances. Occupation, more than education, determines social differences in the effects of retirement on health. Our findings indicate

that the retirement reforms aimed at prolonging working life can be socially distortive due to the differential effects based on SES. We find that retirement at age 67 has positive health implications for low SES groups, but we find no effects for high SES groups. A formal test of these differences confirms that occupation matters for the health effects of retirement.

Finally, our study contributes to generalizing the positive physical health effect of retirement found in the literature across a larger age span. The current literature has mainly assessed retirement ages from late the 50s to about 65. Here, we confirm that the positive effects still hold for individuals retiring at age 67. This implies that retirements affects health also when we assess those who are healthy enough to stay in the labor force until the late 60s. Assessments of higher age thresholds are valuable for policymakers as current retirement reforms typically aim at increasing the retirement age. These reforms will likely affect relatively healthy individuals, i.e. workers who remain employed until these higher retirement ages.

References

Behncke, S. (2012). Does retirement trigger ill health? *Health Economics*, 21(3):282–300.

Blake, H. and Garrouste, C. (2013). Killing me softly: Work and mortality among French seniors. Technical report, HEDG, Department of Economics, University of York.

Bloemen, H., Hochguertel, S., and Zweerink, J. (2017). The causal effect of retirement on mortality: Evidence from targeted incentives to retire early. *Health Economics*, 26(12):204–218.

Bound, J. and Waidmann, T. (2007). Estimating the health effects of retirement. *Michigan Retirement Research Center Research Paper No. UM WP 2007-168.*

Case, A. and Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging*, pages 185–212. University of Chicago Press.

Cattaneo, M. D., Jansson, M., and Ma, X. (2016). Simple local regression distribution estimators with an application to manipulation testing. *Unpublished Working Paper, University of Michigan, and University of California Berkeley.*

Coe, N. B. and Zamarro, G. (2011). Retirement effects on health in europe. *Journal of Health Economics*, 30(1):77 – 86.

DeSalvo, K. B., Jones, T. M., Peabody, J., McDonald, J., Fihn, S., Fan, V., He, J., and Muntner, P. (2009). Health care expenditure prediction with a single item, self-rated health measure. *Medical care*, pages 440–447.

Dong, Y. (2015). Regression discontinuity applications with rounding errors in the running variable. *Journal of Applied Econometrics*, 30(3):422–446.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of Health Economics*, 43:1–12.

Godard, M. (2016). Gaining weight through retirement? Results from the SHARE survey. *Journal of Health Economics*, 45:27–46.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Hallberg, D., Johansson, P., and Josephson, M. (2015). Is an early retirement offer good for your health? Quasi-experimental evidence from the army. *Journal of Health Economics*, 44:274–285.

Heller-Sahlgren, G. (2017). Retirement blues. *Journal of Health Economics*, 54:66–78.

Hernæs, E., Markussen, S., Piggott, J., and Vestad, O. L. (2013). Does retirement age impact mortality? *Journal of Health Economics*, 32(3):586–598.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources*, 49(1):195–233.

Kudrna, G. (2017). The Norwegian pension reform: An external perspective. *CEPAR Working Paper 2017/07, CEPAR*.

Kuhn, A., Wuellrich, J.-P., and Zweimüller, J. (2010). Fatal attraction? Access to early retirement and mortality. *IZA discussion paper No. 5160*.

Lacson, E., Xu, J., Lin, S.-F., Dean, S. G., Lazarus, J. M., and Hakim, R. M. (2010). A comparison of sf-36 and sf-12 composite scores and subsequent hospitalization and mortality risks in long-term dialysis patients. *Clinical Journal of the American Society of Nephrology*, 5(2):252–260.

Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2):655–674.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Mazzonna, F. and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4):691–710.

Mazzonna, F. and Peracchi, F. (2017). Unhealthy retirement? *Journal of Human Resources*, 52(1):128–151.

McGarry, K. (2004). Health and retirement do changes in health affect retirement expectations? *Journal of Human Resources*, 39(3):624–648.

Neuman, K. (2008). Quit your job and get healthier? The effect of retirement on health. *Journal of Labor Research*, 29(2):177–201.

Røed, K. and Haugen, F. (2003). Early retirement and economic incentives: Evidence from a quasi-natural experiment. *Labour*, 17(2):203–228.

Rohwedder, S. and Willis, R. J. (2010). Mental retirement. *The Journal of Economic Perspectives*, 24(1):119–138.

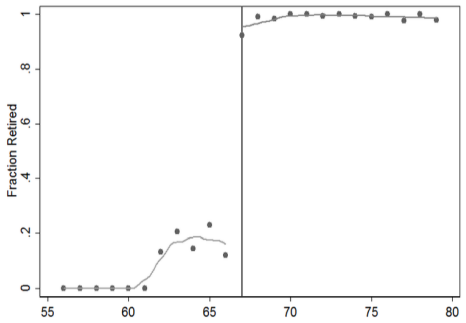
Slagsvold, B., Veenstra, M., Daatland, S. O., Hagestad, G., Hansen, T., Herlofson, K., Koløen, K., and Solem, P. E. (2012). Life-course, ageing and generations in Norway: The NorLAG study. *Norsk Epidemiologi*, 22(2).

Ware Jr, J. E., Kosinski, M., and Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34(3):220–233.

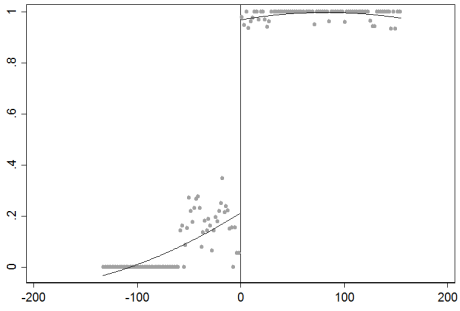
WHO (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

Figures

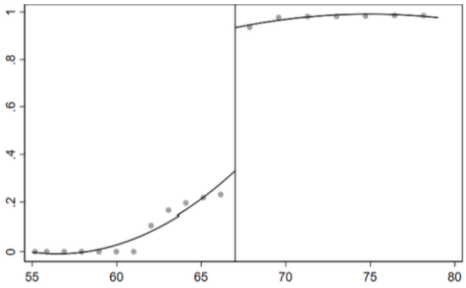
Figure 1: Discontinuity in Retirement at the Retirement Age Threshold



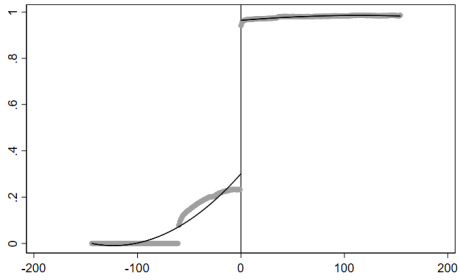
(a) NorLAG Data: Fraction Retired - Age in Years



(b) NorLAG Data: Fraction Retired - Age in Months



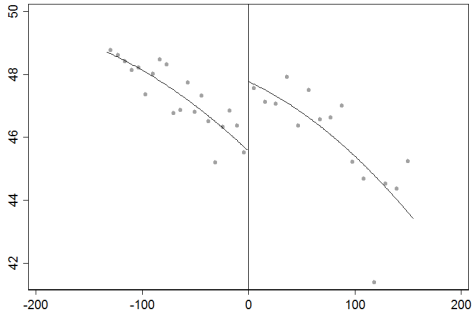
(c) Adm. Data: Fraction Retired - Age in Years



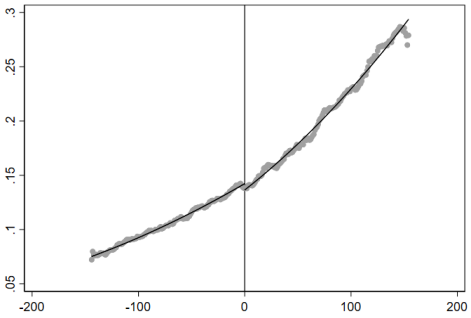
(d) Adm. Data: Fraction Retired - Age in Months

Note: The graphs show the fraction retired by age from the two datasets. The upper graphs are based on the survey data, whereas the two lower graphs are based on administrative data. All graphs depicts the fraction retired across the age span 55-79. The x-axis on the left two graphs depicts age in years, whereas the x-axis in the graphs to the right depicts age in months, relative to the retirement eligibility age-in-months (805 months).

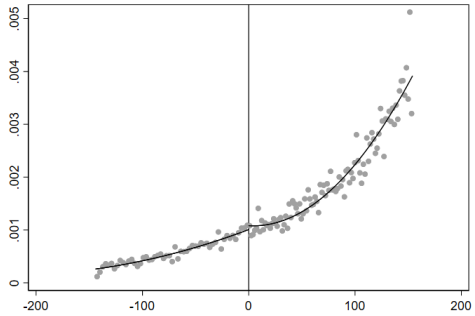
Figure 2: Discontinuity in Health at the Retirement Age Threshold



(a) NorLAG Data: Physical Health (SF-12)



(b) Administrative Data: Acute Hospitalizations



(c) Administrative Data: Mortality

Note: The graphs present the age-health relationship for physical health, acute hospital admissions and mortality. The scale for physical health are points on the SF-12 scale and the scale of acute hospital admissions and mortality corresponds to the incidence in the population. The x-axis displays age-in-months relative to the retirement age threshold at 805 months.

Tables

Table 1: Labor Market Participation for Individuals Aged 56-79 in 2007

	Working		Retired		ER		DI	
Age Group	Men	Women	Men	Women	Men	Women	Men	Women
56 - 61	79%	79 %	-	-	-	-	19 %	28 %
62 - 66	59%	49 %	-	-	16 %	13 %	31 %	41 %
67 - 69	17%	9 %	89 %	92 %	-	-	-	-
70 - 79	18%	2 %	98 %	98 %	-	-	-	-

Note: The numbers are based on own calculations using the administrative data which covers the entire population of Norway (See Section 4 for a description). Work is defined as having earnings larger than zero. The states will not sum to unity because individuals can be in two states at the same time, e.g. by combining work and partial uptake of DI.

Table 2: Descriptive Statistics

	Whole Sample	Above Threshold	Below Threshold
Source: NorLAG			
SF12	46.93	45.73	47.55
	(10.78)	(12.03)	(10.12)
Age	65.34	66.15	67.00
	(6.58)	(0.36)	(0.0)
Retired	0.44	0.18	0.96
Elementary education	0.23	0.25	0.25
High school degree	0.51	0.45	0.51
Any college	0.27	0.30	0.25
Professional	0.48	0.47	0.50
Manual	0.43	0.40	0.41
Living with partner	0.71	0.75	0.72
Female	0.48	0.47	0.50
Observations	4619	190	200
Source: Admin. Data			
Acute Hospital Admissions	0.142	0.140	0.141
Mortality	0.019	0.017	0.018
Age	64.92	66.19	67.00
	(6.67)	(0.38)	(0.00)
Retired	0.40	0.29	0.95
Elementary education	0.31	0.32	0.34
High school degree	0.45	0.46	0.45
Any college	0.24	0.23	0.21
Married	0.63	0.64	0.64
Female	0.51	0.51	0.51
Observations	1,071,068	31,751	33,752

Note: This table displays descriptive statistics for the two data sources, the NorLAG data (above) and the administrative data (below). Column (1) presents means for the entire sample, whereas the other two columns display means for the sub-sample of individuals included in the estimations (we use a bandwidth of ten months for the estimations). Column (2) displays the means for the sub-samples aged 795-804 months (control group) and Column (3) for those aged 805-814 months (treatment group). Standard deviations in square brackets.

Table 3: First-Stage Regressions

	All	Men	Women
Source: NorLAG			
Retired	0.954***	0.941***	0.961***
	(0.0362)	(0.0587)	(0.0431)
Observations	371	190	181
Source: Admin. Data			
Retired	0.720***	0.683***	0.756***
	(0.00264)	(0.00389)	(0.0356)
Observations	825605	407386	418219

Note: This table show the first-stage regressions specified in Equation (2). The reported coefficient is γ from Equation (2). Estimation is done using a bandwidth of ten months. Standard errors in parentheses are clustered at the age-in-months level for the NorLAG data and at the individual level for the administrative data. *=p<0.10, **=p<0.05, ***=p<0.01.

Table 4: Short-Term Retirement Effects on Physical Health

	All	Men	Women	Manual	Professional	Lower	Higher
Retired	5.689***	8.036***	4.053	13.16***	-0.333	8.358***	-1.952
	(1.979)	(3.026)	(3.465)	(3.508)	(3.761)	(2.415)	(5.449)
Observations	361	185	176	126	123	261	99

Note: This table displays the impact of retirement on physical health (SF12). *All* refers to the sample as a whole, *Professional* and *Manual* to type of occupation and *Lower* and *Higher* to levels of education. The reported coefficient is τ from Equation (3). Estimation is done using a bandwidth of ten months. Standard errors in parentheses are clustered at the age-in-month level. *=p<0.10, **=p<0.05, ***=p<0.01.

Table 5: Short-Term Retirement Effects on Hospitalizations and Mortality

	All	Men	Women	Lower	Higher
Hospitalizations					
Retired	-.00419	-.00417	-.00440	-.00589**	.00255
	(.00258)	(.00395)	(.00339)	(.00292)	(.00535)
Observations	825605	407386	418219	643441	182164
Mortality					
Retired	-.000123	-.0000355	-.000266	-.0000895	-.000299
	(.000204)	(.000343)	(.000236)	(.000233)	(.000399)
Observations	840239	416611	423628	655743	184496

Note: This table displays the impact of retirement on acute hospitalizations and mortality. *All* refers to the sample as a whole and *Lower* and *Higher* to levels of education. The reported coefficient is τ from Equation (3). Estimation is done using a bandwidth of ten months. Standard errors in parentheses are clustered at the individual level. *=p<0.10, **=p<0.05, ***=p<0.01.

Table 6: Short-Term Retirement Effects on Health by SF12 Components

	Functional	Daily	Specific	Mental	Pain
All	-0.0741	-0.171**	-0.229**	0.0754**	-0.232***
	(0.0791)	(0.0776)	(0.0934)	(0.0377)	(0.0783)
Observations	371	368	369	368	371
Men	-0.0344	-0.325***	-0.401**	0.0953	-0.180*
	(0.140)	(0.123)	(0.168)	(0.0803)	(0.100)
Observations	190	189	189	188	190
Women	-0.126	-0.0378	-0.0693	0.0562	-0.328***
	(0.105)	(0.128)	(0.152)	(0.0419)	(0.117)
Observations	181	179	180	180	181
Manual	-0.258**	-0.558***	-0.551***	0.0703	-0.503***
	(0.125)	(0.128)	(0.132)	(0.123)	(0.135)
Observations	127	126	127	127	127
Professional	-0.0110	0.0937	0.0160	-0.0455	0.0342
	(0.166)	(0.123)	(0.157)	(0.0486)	(0.178)
Observations	123	123	123	123	123
Low Education	-0.158*	-0.284***	-0.292***	0.121***	-0.314***
	(0.0864)	(0.110)	(0.0971)	(0.0441)	(0.103)
Observations	270	267	268	267	270
High education	0.124	0.104	-0.0597	-0.0215	0.0183
	(0.200)	(0.156)	(0.194)	(0.0723)	(0.159)
Observations	100	100	100	100	100

Note: This table presents the impact of retirement on selected components of the physical health outcome (SF-12). The reported coefficient is τ from Equation (3). Estimation is done using a bandwidth of ten months. Standard errors in parentheses are clustered at the age-in-month level.*=p<0.10, **=p<0.05, ***=p<0.01.

Table 7: Formal Test of Differences by Socioeconomic Status

	Education	Gender	Occupation
Retired	4.975	3.696	6.858*
	(3.115)	(2.549)	(3.305)
Observations	361	361	249
Hospitalizations			
Retired	-.00294	.00224	
	(.00533)	(.00509)	
Observations	825605	825605	
Mortality			
Retired	.000066	.0000171	
	(.0000145)	(.000139)	
Observations	840239	840239	

Note: This table displays the interaction between retirement eligibility and SES (education and occupation (only for the NorLAG)) and gender. The first column presents the results for physical health from the NorLAG data and the second and third columns presents the results for acute hospitalizations and mortality, respectively, for the Administrative data. The reported coefficient is γ from Equation (4). Estimation is done using a bandwidth of ten months. Standard errors in parentheses are clustered at the age-in-month level for the NorLAG data and at the individual level for the Administrative data. *=p<0.10, **=p<0.05, ***=p<0.01.

Appendix - Sensitivity and Robustness

A.1 Disabled Individuals - Past Labor Income and Self-Reported Work Status

People on disability insurance are mechanically transferred from disability pension to retirement pension at age 805 months. We need to make sure that the positive physical health effects we found are not driven by these individuals. Initially, there is no reason to believe that there should be an effect for these individuals as they were not working before retirement, and should therefore have no change in circumstances. However, as the health measure contains elements of self-assessed health, one could imagine that someone who is disabled may need to justify their status as disabled, consciously or subconsciously. In this case, poor health prior to the statutory retirement age may be under-reported. Post retirement, when they are no longer in a situation where poor health is defining their labor market status, they might feel healthier, or no longer have the need to report poor health. If this scenario is plausible, we need to rule out that the results found in Section 5 are driven by this group.

The first two rows of Table A.1 displays the results on two sub-samples of the survey data (labeled "Working" and "Income"), each aimed at running the analysis only on the sub-sample that was recorded as working until the statutory retirement age. The working sub-samples are defined in Section 3.2. Finding coefficients of the same sign and magnitude, especially for the rule based on self-assessed work status, ensures us that these effects are not driven by the disability justification hypothesis. The estimations based on the income-rule yields large and insignificant coefficients, both a consequence of the small sample sizes. Yet, the direction of the effects are similar to what was found in the main analysis.

For the outcomes from the administrative data, as these are not subject to the potential justification bias, we should expect that individuals who retire formally at 67, but without any actual change in circumstances, should water down the effects. We can therefore expect that this assessment can uncover significant effect, not detected in the gross sample. The first row of Table A.2 and Table A.3 presents the estimations restricted to "workers" for acute hospitalizations and mortality, respectively. Here, we find no significant results for any of the sub-groups, besides a significant effect on mortality for women (0.2 percentage points significant at the 5 percent level). The significant result on hospitalizations found for men with low education in the main analysis, is no longer present.

A.2 Robustness Checks and Validity of the Regression Discontinuity Design

Below we assess the sensitivity of the results for different bandwidth selections; we check for discontinuities in the forcing variable, age, at the cutoff; we test for discontinuities in other outcomes that should not have been effected by the threshold; and, we check for discontinuities in the outcomes of interest at points in the age distribution where there should not be any discontinuities. This robusness assessment follows the suggestions in Imbens and Lemieux (2008) closely.

A.2.1 Bandwidth Selection

The worry in an RD application is that using a bandwidth that is too wide, allows for other things than the intervention of interest to drive differences in outcomes for those right above compared to those right below the threshold. In Table A.1 we display the results using bandwidths of 7 and 15 months for physical health. Using a bandwidth of 7 months does not alter the results, whereas increasing the bandwidths to 15 months somewhat reduces the effects. This is not surprising given the downward slope of the health trajectory across age and the upward shift in this trajectory at the retirement eligibility threshold.

The results for hospitalizations and mortality are displayed in Table A.2 and Table A.3. For acute hospital admissions, we find that increasing the bandwidth to 15 months yields significant, negative effects. The effects are still small ranging from 0.7 to 1 percentage points. As the incidence is 14 percent, this entails a 5-7 percent reduction in the likelihood of an acute hospitalization. Increasing the bandwidth increases the likelihood of factors, other than retirement, affecting acute hospital admissions. Another explanation can be that it takes some time for retirement to take effect on health issues such as stroke and acute heart conditions, thus including more post-retirement months increase the likelihood of finding significant effects. As in the main analysis, we find no effects of retirement on mortality at any of these bandwidths.

A.2.2 Continuity of the Forcing Variable

Vital to any RD application is the individual's incapability of manipulating the forcing variable. In this case, the forcing variable is age (reported by public registers), which individuals cannot manipulate in any way. It could however be the case that retired individuals are more likely to respond to the survey due to the reduced opportunity cost of time. Figure A.1 shows two histograms of age-in-months assessing potential bunching at the threshold. There is no evidence of any discontinuity in the forcing variable at the threshold. We also did a more formal test proposed by Cattaneo et al. (2016), a local

polynomial density estimator for testing the null of continuous density of the forcing variable at the threshold. The p-value under this test is 0.3251. For the population level data, this holds by construction, as people cannot manipulate their age and as all individuals in the population are represented in the data.

A.2.3 Placebo Tests

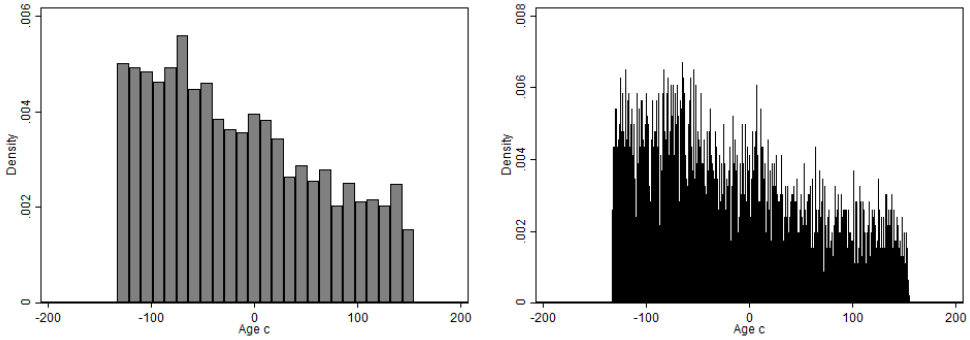
The placebo tests entails testing for discontinuities in the three health outcomes at points in the age distribution where there should be no discontinuities. A common practice is to conduct placebo tests at the median age of the two sub-samples below and above the actual cut-off. In this case, the median age below the threshold is age 62. However, some individuals can retire at this age, thus making is an unsuited placebo threshold. Consequently, we use age 61 for the lower placebo. For the upper placebo, we use age 73. No discontinuities or significant effects were found at these placebo thresholds for physical health (Table A.1). For acute hospital admissions (Table A.2), we find significant effects for both the upper and lower placebo. For the lower placebo, this could be due to some occupations having special age-limits for retirement at 61. However, we find no explanations for why the upper placebo yields significant, and even positive effects. This finding reduce the credibility of the effects found in the main analysis for this outcome. The placebo results for mortality is presented in Table A.3. There are no significant effects and the coefficients are close to zero for all sub-group at both placebo thresholds.

A.2.4 Discontinuity in Other Outcomes

Finally, we look for discontinuities in an outcome that should not be affected by retirement, at least not in the short-term. Here, we assess the likelihood of living with a partner or spouse (NorLAG) or being married (administrative data). The regression results shown in Table A.1 and Table A.4 confirm that there are no retirement effect on these outcomes.

Appendix Graphs and Tables

Figure A.1: Discontinuity of the Forcing Variable



(a) (b)

Note: The histograms show the distribution of age in months for the age-range 56-79 using the bin-width suggested by STATA (left histogram) and using one bin for each age-in-months (right histogram).

Table A.1: Robustness Checks Survey Data: Physical Health

	All	Men	Women
Conditional on income	16.42***	15.83	-1.553
	(2.966)	(10.88)	(7.264)
Observations	82	53	39
Conditional on working	6.274***	9.741***	2.523
	(2.089)	(3.758)	(7.312)
Observations	247	142	105
Bandwidth 7	9.472***	14.69***	2.623
	(2.019)	(5.206)	(4.245)
Observations	275	142	133
Bandwidth 15	5.801***	9.391***	2.623
	(2.130)	(3.109)	(6.628)
Observations	540	278	262
Placebo at 61	-1.441	.971	-5.752
	(3.665)	(4.220)	(6.628)
Observations	454	242	212
Placebo at 73	-1.111	-1.264	.628
	(1.685)	(4.786)	(2.213)
Observations	251	127	124
Living with a partner	-0.106	-0.0413	-0.162
	(0.0931)	(0.108)	(0.176)
Observations	371	190	181

Note: This table displays the various robustness checks described in the Appendix, for the physical health outcome and the NorLAG data. Standard errors in parentheses are clustered at the age-in-month level. *=p<0.10, **=p<0.05, ***=p<0.01.

Table A.2: Robustness Checks Administrative Data: Acute Hospitalizations

	All	Men	Women	Low Educ.	High Educ.
Conditional on working	0.00212	0.00266	0.00133	0.000309	0.00687
	(0.00295)	(0.00426)	(0.00397)	(0.00345)	(0.00566)
Observations	362,857	203,212	159,645	259,427	103,430
Bandwidth 7	-0.00246	-0.00117	-0.00374	-0.00343	0.00231
	(0.00227)	(0.00343)	(0.00303)	(0.00258)	(0.00459)
Observations	583,686	287,791	295,895	455,797	127,889
Bandwidth 15	-0.00722*	-0.00977*	-0.00520	-0.0101**	0.00321
	(0.00377)	(0.00584)	(0.00487)	(0.00427)	(0.00773)
Observations	1,241,687	612,603	629,084	965,278	276,409
Placebo at 61	-0.000587	-0.0000883	-0.00105*	-0.00155**	0.00199*
	(0.000510)	(0.000610)	(0.000541)	(0.000705)	(0.00104)
Observations	1,311,705	667,661	644,044	962,159	349,546
Placebo at 73	0.00106	0.00284*	-0.000557	0.00194**	-0.00321**
	(0.000676)	(0.00162)	(0.000861)	(0.000699)	(0.00153)
Observations	634,319	294,672	339,647	527,740	106,579

Note: This table displays the various robustness checks described in the Appendix, for Acute hospital admissions. Standard errors in parentheses are clustered at the age-in-month level. Standard errors in parentheses are clustered at the individual level. *=p<0.10, **=p<0.05, ***=p<0.01.

Table A.3: Robustness Checks Administrative Data: Mortality

	All	Men	Women	Low Educ.	High Educ.
Conditional on working	0.00212	0.00266	0.00133	0.000309	0.00687
	(0.00295)	(0.00426)	(0.00397)	(0.00345)	(0.00566)
Observations	362,857	203,212	159,645	259,427	103,430
Bandwidth 7	-0.00246	-0.00117	-0.00374	-0.00343	0.00231
	(0.00227)	(0.00343)	(0.00303)	(0.00258)	(0.00459)
Observations	583,686	287,791	295,895	455,797	127,889
Bandwidth 15	-0.00722*	-0.00977*	-0.00520	-0.0101**	0.00321
	(0.00377)	(0.00584)	(0.00487)	(0.00427)	(0.00773)
Observations	1,241,687	612,603	629,084	965,278	276,409
Placebo at 61	-0.000587	-0.0000883	-0.00105*	-0.00155**	0.00199*
	(0.000510)	(0.000610)	(0.000541)	(0.000705)	(0.00104)
Observations	1,311,705	667,661	644,044	962,159	349,546
Placebo at 73	0.00106	0.00284*	-0.000557	0.00194**	-0.00321**
	(0.000676)	(0.00162)	(0.000861)	(0.000699)	(0.00153)
Observations	634,319	294,672	339,647	527,740	106,579

Note: This table displays the various robustness checks described in the Appendix, for Mortality. Standard errors in parentheses are clustered at the age-in-month level. Standard errors in parentheses are clustered at the age individual level. *=p<0.10, **=p<0.05, ***=p<0.01.

Table A.4: Robustness Checks Administrative Data: Discontinuity in Marital Status

	All	Men	Women	Low Educ.	High Educ.
Married	0.00233	0.00324	0.00104	0.00156	0.00654
	(0.00229)	(0.00336)	(0.00311)	(0.00249)	(0.00576)
Observations	825,605	407,386	418,219	643,441	182,164

Note: This table displays the impact of retirement on the likelihood of being married. The reported coefficient is τ from Equation (3). Estimation is done using a bandwidth of ten months. Standard errors in parentheses are clustered at the individual level. *=p<0.10, **=p<0.05, ***=p<0.01.

Chapter 4:

Old Problem, New Evidence: Health Related Sample-Selection in Analyses of Health and Ageing

Old Problem, New Evidence: Health-Related Sample Selection in Analyses of Health and Aging

Maja Weemes Grøtting*

May 6, 2019

Abstract

This paper analyzes the extent and consequences of health-related non-response in survey data, defined as either initial non-response or attrition across waves. The first part of the paper provides a thorough description of the initial non-response and attrition between waves in the Norwegian Study on Life Course, Ageing, and generation (NorLAG), a typical survey sample. In the second part, the potential bias from sample selection is examined using econometric analyses of educational differences in health across age as a case study. This is an empirical application that is believed to be particularly prone to bias from health selection, as non-response correlates with both health, education, and age. I make use of a sample of Norwegian older adults and register data that cover the full population to make various assumptions about the selection process and find that using population weights, inverse probability weights, the Heckman selection model, and a Copula selection model does not produce substantially different estimates. Across all models, educational differences in health converge in old age, as is commonly found in analyses that use survey data. This convergence is also found using the register data. Finally, the register data reveal that healthy survivor bias among the low educated is an important factor driving the convergence in health inequalities in old age.

JEL Classification: C18, C24, I14, J14
Keywords: Sample selection, attrition, health inequalities, aging

*NOVA, Oslo Metropolitan University; email: maja.grotting@oslomet.no

1 Introduction

The impact of sample selection in analyses of survey data is an old question that has been extensively assessed. Sample selection can lead to serious bias, unless properly accounted for, if the non-response correlates with outcomes of interest for the empirical analysis in question. The problem arises because an individual's willingness and ability to participate in a survey depend not only on daily life circumstances but also on a wide range of personal characteristics, including health and well-being. The potential for bias from sample selection is particularly prominent in studies of health, and even more so in studies of health and aging.

Sample selection applies to cross-sectional and longitudinal samples, and it can be characterized as either initial selection, in which case the survey sample is not representative of the population the sample represents, or selective attrition (data loss) across waves in a longitudinal survey, in which case the sample composition changes over time in a way that is related to variables of interest. The initial non-response is usually larger than the attrition across waves, because respondents in the first wave tend, on average, to be more willing to cooperate than the target populations as a whole.

The majority of the empirical literature that assesses the impact of sample selection, has focused on attrition across waves in longitudinal surveys. Whether representability is compromised across waves in longitudinal data is an important question. Strong evidence of health-related attrition has been documented across several well-known household surveys, but is found to cause little bias (Becketti et al., 1988; Fitzgerald et al., 1998; Contoyannis et al., 2004; Jones et al., 2006). However, empirical evidence of the impact of initial selection is limited. Although it is likely that the initial and subsequent non-response are based on the same mechanism, we cannot rule out that the patterns of non-response differ. In addition, as initial non-response typically is larger than non-response at subsequent waves, initial selection can still be an important source of bias.

Empirical approaches, such as weighting, to control for initial non-response exist. How well these methods work depends on the pre-sampling information at hand. This information is often limited with respect to how well it captures the underlying mechanisms behind the non-response. Although age and variables expressing socioeconomic status (SES) usually correlate with health, uncertainty remains about how well weights based on such variables re-establish a distribution of health in the sample that mirrors the distribution of health in the population. Ideally, to investigate the bias from initial non-response, individual-level health data on the full population that the survey sample is intended to represent is needed. However, the availability of this type of data is limited, which is one reason bias from initial non-response rarely has been assessed.

In this paper, I assess the bias from health-related non-response in a typical survey sample by comparing the results based on survey data with corresponding results from register data covering the full population. A novelty of this analysis is access to sufficient individual-level health information collected from register data on services provided in primary health care to compose a generic health measure for each individual in the population. As the register data are not hampered by non-response, the registers can uncover the true age pattern in health in the population.[1]

As a case study, I apply an empirical problem that is especially prone to bias from health-related sample selection: social inequalities in health across age. This is an application in which the problem of health-related sample selection is especially salient, as the outcome of interest, health, and the explanatory variables, age and socioeconomic status (SES), are likely to correlate with non-response. To respond to a survey, health above a minimum level is usually required. Moreover, as age increases, health deteriorates, and the likelihood of response declines. This implies that those who respond at older ages are more selected with respect to health. This process applies to all SES groups. However, it is more prominent in low SES groups. As individuals in the lower SES groups have worse health outcomes and higher mortality for a given age, they are more likely to never respond to or to drop out of the sample across time or age, compared to higher order SES groups. As a result, across time or age, the remaining sub-sample in the low SES group becomes increasingly less representative of that group as a whole. Unless properly accounted for, the consequence of this sort of sample selection is empirical evidence of reduced social inequalities in health in later life. In fact, evidence of converging health inequalities in old age is a common empirical result in the social gerontological literature, which is where the health-SES-age relationship has been most widely studied (Ross and Wu, 1996; Chandola et al., 2007; Kim and Durden, 2007; Mirowsky and Ross, 2008). Whether the convergence is due to reduced social inequalities or merely is due to sample selection, is largely unresolved as the bias from sample selection has been inadequately assessed in much of this literature.

The distinction between non-response and data loss due to mortality is important in general, and for this analysis in particular. Those who fall out of a sample across time may do so because they are deceased or because they are prevented from or unwilling to participate in the survey. The first differ substantially from the latter because those who

[1] Age pattern is a rude simplification as the main empirical appliciation in this study is a cross-sectional representation of health across age; thus, the age pattern in health applied here is therefore a combination of age and cohort effects. In this paper, I also assess the age pattern in health using panel register data. For this part of the analysis, the estimated age pattern in health is a combination of age, cohort, and time effects. There are several issues with disentangling time, cohort, and age effects, despite having access to large scale panel register data, but that is a discussion that is beyond the scope of this paper. See e.g. Biørn et al. (2013) for a discussion.

are deceased are no longer in the survey's reference population, unlike the non-responders who are in the population but not in the sample. A sample can therefore be selected compared to the population the sample is drawn from (initial selection), or it can be selected across waves (selective attrition). The latter can be caused by either non-response or mortality that is non-random across subsequent waves in longitudinal surveys (selective attrition). Either way, selective attrition makes subsequent waves unrepresentative compared to the first wave. Finally, in analyses of trajectories across age, any dataset, either sample or full population data, can be selected if people in different subgroups of interest have different mortality rates for a given age (selective mortality).

Whether or not sample selection or selective mortality leads to bias, ultimately, depends on the empirical question at hand. If interest lies in describing a cross-sectional relationship as it currently is in the population, then the initial selection is the relevant dimension to correct for. If, however, the interest lies in describing a longitudinal relationship, then non-response and mortality that is non-random can lead to bias. In this paper, the primary interest lies in assessing whether the non-response in a typical health survey sample leads to bias compared with the population the sample represents. Therefore, selective mortality is of secondary importance as this happens at the population level, as well as in the sample. However, at the end of the analysis, I show how selective mortality (healthy survivor bias) can bias the empirical relationship between health, SES, and age–even in a dataset that has no non-response.

This study uses a combination of survey and register data. The survey data are the Norwegian Study on Life Course, Ageing, and Generation (NorLAG). This is a panel survey of Norwegian older adults that corresponds to well-known surveys such as the Survey of Health, Ageing and Retirement in Europe (SHARE), English Longitudinal Study of Ageing (ELSA), and the Health and Retirement Study (HRS) for the US.[2] The NorLAG currently consists of two waves (collected in 2002 (T1) and 2007 (T2)). Among the 8,298 persons initially invited to participate in this survey in 2002, 67 percent responded. In the second wave, the response rate was 60 percent among the gross sample invited to participate at T1[3] and 68 percent among those who also participated at T1. Population weights, derived by Statistics Norway, are available for each wave. These weights are based on individual-level data for measures such as education, age, gender, region, and centrality, that are known for the entire population.

The second data source consists of full population individual-level register panel data (2006 to 2014). This is the reference population on which the NorLAG sample is based.

[2]Other examples include Berlin Aging Study (BASE), Longitudinal Aging Study Amsterdam (LASA), and Swedish National Study on Aging and Care (SNAP).

[3]In addition, 737 of the initial non-responders participated in the second wave.

To compare whether the NorLAG reproduce the relationship between health, SES, and aging in the population, a measure of health from the registers that is comparable to the health measure applied from the survey data is needed. From the NorLAG data, I apply the Short Form 12 (SF12) which is a generic health measure often applied in survey analyses. From the registers, I propose the following measure of health: the annual reimbursements claimed by primary care physicians (PCPs) for patients. In Norway, practically all citizens are associated with a PCP of their choice through a patient list system.[4] PCPs are responsible for all initial assessments and treatments. In addition, treatments in secondary care are via PCP referral. Except for a small fixed patient charge, consultations are free of charge for the patient, and the PCP is reimbursed by the state per consultation based on the duration of the consultation and the specific treatment given. I argue that the reimbursements claimed by the physician (PCPR) can be seen as a generic health measure because it is a general expression of health and functional ability, as with the SF12. In addition, as a PCP consultation often is initiated by the patient as a response to self-assessed health, PCPR contains an element of self-assessment as well. I show that PCPR have the same patterns of distribution as the SF12 across dimensions that are central to this analysis, and thus, can arguably be used as a benchmark for how the SF12 would be distributed in the population across the variables that are of interest in this study: gender, age, and education.

In addition to comparing results using survey data to results using register data, I employ more conventional sample selection correction methods, such as weights and selection models. This implies assessing attrition from one wave to the next in the NorLAG sample which enables comparisons with the previous literature. In addition, it is likely that attrition and initial non-response are based on similar non-response mechanisms; thus, comparing selection corrections due to attrition and due to the initial selection is interesting on its own. Because the true nature of the pattern of health-related non-response (the selection mechanism) in the data is unobserved, it is not clear what the best method to correct for sample selection is. Therefore, I make various assumptions about the selection mechanism and assess how estimates are affected by the corresponding selection correction method. The selection correction methods are those most common in the applied microeconometrics literature such as population weights, Inverse probability weights (IPW), and the Heckman selection model (hereafter, Heckman; (Heckman, 1979)). In addition, I apply the more recently developed Copula selection model (hereafter, Copula; (Smith, 2003)).

The aim of the study is to shed new light on the potential bias from sample selection

[4]Only 0.5 percent of Norwegian citizens were not on a PCP list in 2007 (The Norwegian Directorate of Health, 2009).

in analyses of health and aging, and the study makes two contributions to the existing literature. First, a novel feature of this paper is the presentation of a register-based generic measure of health, available for the full population. This measure serves as base for evaluating health-related non-response in the survey data. Second, although IPWs have commonly been used in similar settings, few examples use selection models in analyses of the bias from health-related sample selection. This is especially true for the Copula selection model.

In the first part of the analysis, I document the presence of sample selection in the NorLAG data. Then, I establish the familiar pattern of converging health inequalities across age in later life. Suspecting that health-related non-response could be a driver of this result, I apply the various sample selection correction methods. Analyses based on the register data confirm the convergence in health inequalities across age. Moreover, the two weighting approaches, the Heckman, and the Copula provide trajectories similar to a baseline of no selection correction. However, unlike the Heckman, the Copula provides evidence of selection bias. Despite this, the bias from selection is small as the pattern and sizes of the educational differences in health across age are similar across models and specifications. Finally, using the full panel of the register data, I show that although there is no non-response in the register data, there is a strong case for mortality selection driving the convergence in health inequalities in later life in these analyses.

The remainder of this paper is organized as follows: Section 2 presents a literature review, and Section 3 presents the data. Section 4 provides an assessment of the selection mechanism in the survey data, and Section 5 presents the case study and the different selection correction approaches applied. Section 6 presents the results, and Section 7 concludes.

2 Literature

In this section, I present the literature on sample selection, before I present the literature on social inequalities in health in later life. The latter is primarily based on social gerontology literature.

2.1 Controlling for Sample Selection

A large body of literature proposes alternative solutions for handling sample selection, and a common approach has been to weight the data by IPWs (Wooldridge, 2002; Fitzgerald et al., 1998; Jones et al., 2006). IPWs are derived by estimating the likelihood of response based on observable characteristics and then weighting each observation by the inverse of the likelihood of response. The selection correction entails giving more weight to observa-

tions with a small likelihood of appearing in the data, based on observable characteristics. The implicit assumption is that responders who are similar to non-responders in terms of observable characteristics also have similar values of the missing outcome variable.

Another group of selection correction methods are selection models (Hausman and Wise, 1979; Heckman, 1979). Selection models imply modeling the selection process explicitly and controlling for this in the main outcome equation. Here, sample selection is considered an omitted variable bias problem. Thus, including a control for being selected into the sample in the main health equation is assumed to restore the population-level conditional health distribution. Selection models have been widely applied in the context of labor market outcomes (Vella, 1998). In particular, the Heckman has been popular in applied work (Puhani, 2000).

The assumption about the selection mechanism in the data is less strong in the selection model approaches compared to in the weighting approaches, as the selection is allowed to be based on unobservable factors. However, selection models typically rely on multivariate normality for consistent estimates. Moreover, these models are particularly sensitive to these distributional assumptions (Smith, 2003). A growing number of approaches relax the joint normality assumption.[5] One example is the Copula, which can be estimated by maximum likelihood while the assumption of joint normality is relaxed (Smith, 2003).

Among empirical assessments of the bias from, or impact of, sample selection, IPWs have been more commonly applied than selection models. Several studies analyze the impact of sample selection in survey data on estimates of the realtion between health and socioeconomic status by using IPW (Becketti et al., 1988; Fitzgerald et al., 1998; Contoyannis et al., 2004; Jones et al., 2006; Michaud et al., 2011). Although these studies find evidence of health-related attrition in several well-known surveys, such as the British Household Panel Survey (BHPS) (Contoyannis et al., 2004; Jones et al., 2006), the Panel Study of Income Dynamics (PSID) (Becketti et al., 1988; Fitzgerald et al., 1998), and European Community Household Panel (ECHP) (Jones et al., 2006), selective attrition is found to cause no or little bias when IPWs are used as to correct for sample selection. Similar findings have been reported in studies assessing income dynamics and other labor market outcomes (Lillard and Panis, 1998; Zabel, 1998; Ziliak and Kniesner, 1998).[6] Also using IPWs, Michaud et al. (2011) show that bringing respondents who have not responded to a particular wave back into the sample in later waves can substantially

[5] An extensive body of research has developed non- and semi-parametric methods for relaxing the distributional assumptions. These methods are not applied here, but several are covered in the review by Vella (1998).

[6] See the special issue of *Journal of Human Resources on Attrition in Longitudinal Surveys* (Manski and Altonji, 1998).

reduce selection bias in analyses of wealth, home ownership, and labor force participation.

For selection models, there is limited evidence of the impact of health-related non-reponse in survey data. However, the results in Zweifel et al. (1999), showing that health care expenditures do not increase with age if proximity to death is controlled for, sparked a debate about the identifying assumptions in selection models applied to health care expenditures (Salas and Raftery, 2001; Seshamani and Gray, 2004; Zweifel et al., 2004).[7]

Cheng and Trivedi (2015) apply a copula-based selection model to assess the impact of selective attrition in a survey on medical practitioners' income. They find that selection correction by the Copula has little impact on the empirical assessments of income for general practitioners but some impact on the earnings of specialists, thus indicating selection bias in the latter.

Finally, a related approach is by Heiss et al. (2014). They assess how selective mortality across waves in the HRS affects health trajectories for different ethnic groups. Missing data due to mortality is different from other forms of non-response. An important distinction is that the deceased has no health level, unlike other non-responders for whom the health level exists but is unobserved. Careful consideration as to what the selection correction implies about the health levels for deceased observations is therefore important. Heiss et al. (2014) apply a model that allows estimation of health and mortality jointly, to correct for mortality selection in the health-age trajectories, without having to define the health level for those who are deceased. They show that mortality selection is what causes the observed convergence in health between ethnic groups. Moreover, when they correct for mortality selection, they find that the health gap between ethnic groups are in fact increasing across age.

In general, the literature on sample selection has found little bias from sample selection when standard models of selection correction have been applied. One reason can be that the selective non-response simply does not bias the estimates. Another reason can be insufficient data or that the assumptions needed for consistent estimates are violated. In addition, as shown by Heiss et al. (2014), mortality selection can produce large biases in longitudinal estimates of health trajectories across age.

2.2 Social Inequalities in Health

The relationship between SES and health across age has been scarcely assessed in economics, but a comprehensive body of research, primarily in social gerontology, documents this relationship. How these studies deal with sample selection varies greatly. Although most studies acknowledge the likely presence of health selection in the data applied, it is

[7] In addition to poor identification of the selection equation, the critique of the Zweifel et al. (1999) paper concerns endogeneity issues with the explanatory variables.

commonly ignored (Ross and Wu, 1996), or resolved using ad hoc solutions, such as giving the deceased a health score of zero (Benzeval et al., 2011) or the lowest observed value Herd (2006). Another approach entails simply including binary controls for attrition or mortality in longitudinal analysis (Xu et al., 2014). Among more complex approaches are latent growth curve modeling (Kim and Durden, 2007; Brown et al., 2012) or random effects models (Herd, 2006).

In this field of research, there are generally two opposing theories of how the relationship between SES and health changes across age. The Cumulative advantage (CAD) theory predicts diverging inequalities with age due to the socially stratified accumulation of possibilities and risks over the life span (Ross and Wu, 1996). In contrast, the Age-as-Leveler (AAL) theory predicts converging inequalities in later life. The convergence is assumed to be caused by four mechanisms: the existence of a biological ceiling, after which good health can no longer be maintained causing health to deteriorate at a higher rate for the high SES group in old age; more egalitarian public policies among older individuals such as pensions and Medicaid; the fading out of past experiences that were causing health inequalities, such as work conditions in old age; and, finally, selective mortality (Hoffmann, 2011).[8] Most empirical evidence supports the CAD theory for early until mid-adult life (Mirowsky and Ross, 2008). Whereas for later life, there is no consensus. Although the majority of empirical evidence supports the AAL theory in old age, poor correction of the bias from sample selection makes definite conclusions about what drives the convergence hard to reach; that is, the sample selection caused by non-response as well as selective mortality could be driving the results.

3 Data

3.1 The NorLAG Survey Data

The survey data applied in this study are from the NorLAG panel study[9]. This is a survey carried out on a representative sample of Norwegian older adults. The NorLAG contains individual-level data on a range of health outcomes, as well as information about SES. Data collection was carried out by Statistics Norway,[10] and the first two waves were collected in 2002 (age 40–80) and 2007 (age 45–85).[11] All survey responses are merged with administrative records of birth year and education from the population database and national education database, as well as data on welfare transfers and income from

[8] If selective mortality is the cause of reduced inequalities in health, then this is, in reality, support for the CAD theory; that is, the health inequalities have gotten so large that people in the low SES group have higher mortality rates than people in the high SES groups for a given age.

[9] See Slagsvold et al. (2012) for a thorough description.

[10] Using computer-assisted telephone interviews.

[11] An updated sample of individuals aged 40–44 is also available in 2007 but is not included here.

the social security and income registers. The panel contains 10,057 observations divided by 6,307 individuals, of whom 3,765 responded to both waves. The response rates are presented in more detail in Figure 1 below. Statistics Norway provided population weights based on individual-level register data for gender, age, educational level, province, and centrality.

Health status is measured as the physical component of the SF12 (Ware Jr et al., 1996). Among the components of the SF12 are self-rated health, measures of the degree to which an individual is able to perform tasks like vacuuming, moving a table, or climbing stairs, whether there are certain tasks that could not be performed due to health limitations, or whether pain limits daily activities. Thus, the SF12 is an expression of general health comprising both self-rated health and health limitations in daily life. Moreover, the SF12 has been found to be a strong predictor of hospitalization, job loss due to health, future use of medical health services, and depression (Jenkinson and Layte, 1997; Ware Jr et al., 1996; Brazier and Roberts, 2004). The SF12 score is standardized on a scale from 0 to 100 with a mean of 50 and a standard deviation of 10 using the U.S. population as a reference.

3.2 The Register Data

The second data source consists of individual-level register data provided by Statistics Norway that covers the entire resident population over the period 2006–2014. These registers include the same demographic information that was used to construct population weights for the NorLAG sample in addition to variables such as cohabitation status, date of birth and death, and various health registers. Individuals can be linked across all registers by unique identifiers.

The measure of health status from the registers is acquired from detailed information about PCP consultations. The payment to a PCP for a consultations is a combination of a fixed patient charge and per-consultation reimbursement from the state based on the duration of the consultation and the specific treatment given. All reimbursements are recorded by the Control and Distribution of Health Reimbursement database (KUHR) and available for the years 2006–2014.

In this analysis, health status is measured by the sum of the reimbursements for all PCP consultations for an individual on a yearly basis. The sum of the reimbursements can arguably be interpreted as the presence and extent of health limitations in a given year. In addition, I provide an ad hoc solution for individuals living in nursing homes. Because these care facilities usually have in-house physicians, the resident's consultations are not registered in the KUHR. Therefore, I provide all individuals living in nursing homes in the current year with the average of the reimbursements claimed by individuals

who show up in a nursing home the next year.[12]

I do not include utilization of the specialist health care services in the health measure as specialist services are less suited as a measure of health in this setting because it has been shown to be endogenous to SES. Individuals with higher SES are found to get better access to and more treatment in the specialist services for, provided the same levels of health (Grasdal and Monstad, 2011; Syse et al., 2012; Elstad, 2018). One explanation could be that communication is better within the SES groups; that is that the PCP and higher educated individuals understand each other better, thus making the PCP more likely to refer the high SES patient to a specialist than a patient from a lower SES group. Another explanation could be that gaining better access to specialist care services demands skills more common among people with high SES, such as navigating heavy bureaucracy. The worry is that educational differences in health measured as utilization of specialist care services will be biased toward zero as individuals with high SES are better at accessing this type of care.

As the PCP system covers the entire population (privately funded primary health care services were uncommon in 2007), I argue that, overall, PCPR provides a credible measure of the extent of individual health problems around a given age, and thus, is an indicator of an individual's health level. In tables 2 through 4 below, I explore how PCPR are distributed across SES, gender, and age, compared to the distribution of the SF12 from the NorLAG along these dimensions.

3.3 Other Variables

Records of education, which is the measure for SES in this analysis, and age, are based on the same administrative records in the survey and register data (the population register and the national education database). Age is divided into 5-year age groups: "40–44", "45–49",..., "80–84", while education is divided into three levels: "lower secondary education" (8–10 years), "upper secondary education" (11–14 years), and "higher education" (14+ years). Cohabitation, defined as being married or living with a registered partner, is based on the population register and the survey response in the NorLAG data and on the population register in the register data. Summary statistics for the second wave (2007) of the NorLAG data and the register data in 2007 are presented in Tables 1 through 4 below in Section 4.

[12] As a sensitivity analysis, I perform the main analysis using reimbursements excluding those who are institutionalized to see how much this ad hoc solution affects the outcome. The results are presented in Graph A2 in the Appendix.

4 Sample Selection in the NorLAG Survey Data

The aim in this section is to uncover as much information about the sample selection mechanism in the NorLAG data as possible with the data available. I start by describing the response rates, before I compare relevant observable characteristics in the NorLAG and register data. I then compare the patterns of distribution of the two health measures, SF12 and PCPR, and mortality across the same observable characteristics. Finally, I assess what observable characteristics at T1 predict non-response at T2 and whether there is a correlation between the number of waves responded to and health, after controlling for variables such as age and education.

The response rates in the NorLAG are 66.9 percent at T1 and 60.6 percent at T2. Among those who responded at T1, 67.5 percent responded at T2 (the conditional response rate). The main reason for non-response at both waves was refusal to participate, whereas a small fraction was prevented from responding or contact was not established (Koløen et al., 2013; Bjørshol et al., 2010). Figure 1 provides an overview of the response rates and sample sizes. The gross sample invited to participate at T1 was also invited to participate at T2, and about 30 percent of the non-responders at T1 responded at T2. This implies that the full T2 sample of responders can be seen as a cross-sectional sample of the population. Thus, we can assess the non-response at T2 either compared to the full population (initial selection) or compared to at T1 (attrition). This is illustrated by arrow **B** and arrow **C** in Figure 1, respectively. The health registers are not available in 2002 for comparison with the T1 sample. In addition, as this study implies comparing estimates across a range of selection correction methods, including comparing estimates based on the survey sample to estimates based on full population registers, it is convenient for the purpose of the comparisons to base all estimates on the same wave of data.

Throughout this analysis, two different sub-samples of the T2 are applied: an unbalanced sample and a balanced sample. The unbalanced sample consists of all responders at T2. That includes those who responded at both T1 and T2, in addition to those who were invited to participate at T1 but responded only at T2. The balanced sample consists of only those who responded at T2 who also responded at T1. The balanced sample is assumed to be more selected than the unbalanced sample as these individuals have agreed to, or been able to, participate twice. Because the unbalanced sample can be seen as a cross-sectional survey sample, it is interesting to compare results based on this sample with results from the register data. However, for some of the selection correction approaches, data on observable characteristics are needed, also for the non-responders at T2, for example to make IPWs or estimate Heckman's lambda. For these approaches,

the balanced sample must be applied.

The summary tables 1 to 4 show some important preconditions for this analysis. Table 1 compares the distribution of central covariates in the NorLAG to the underlying population of interest: all Norwegian adults aged 45–85 in 2007 measured by the register data. Table 2 and Table 3 show how the distribution of the health measure from the register data compares with the distribution of the health measure from the NorLAG survey. Finally, Table 4 shows how mortality from 2002 to 2007 compares across central covariates in the two data sources.

Table 1 presents summary statistics for the unbalanced and the balanced T2 samples, as well as the means weighted by population weights for the unbalanced sample and by IPWs (derived below) for the balanced sample. The final row contains summary statistics for the population-level register data. Table 1 shows that praimarily those with low education are under-represented in the NorLAG sample.[13] The fraction of low educated for the unweighted means of the unbalanced and balanced samples compared to the actual fraction in the population are 8–9 percentage points lower for men and 9–10 percentage points lower for women. Accordingly, among the higher educated, the pattern is reversed with an over-representation of the higher educated by 6–7 percentage points for men and 9–11 percentage points for women. The balanced sample is more selected than the unbalanced sample, as the educational means differ more for this sub-sample. The population weights and the IPWs do a fairly good job at correcting for this education-based selection.

The NorLAG sample is more representative when it comes to the age distribution. There is a slight under-representation of the younger age groups and the weights correct for the under-representation. For the oldest group (age 80+), the means by the population weights correspond less with the population mean than the other means. On the one hand, the older age groups are oversampled to have enough individuals in the NorLAG sample at these higher ages. On the other, the oldest age group has a poorer response rate compared to the younger groups. In sum, the fraction of older adults in the sample becomes fairly representative compared to the underlying population. For this group, the weights provide a poor correction.

How well the NorLAG sample compares to the full population regarding health has not previously been assessed. This is in part because there are no overlapping measures of health in the two data sources.[14] Here, I show that the PCPR is comparable to the

[13] This was also shown by Lappegård and Veenstra (2010).

[14] There is a measure of mortality in both data sources, but unfortunately, mortality measured by the NorLAG is limited to information about mortality at any point in time between T1 and T2, conditional on responding at T1. Thus this measure is too limited to provide a meaningful basis for regression analysis.

SF12 along the dimensions that are important for this study.

Table 2 shows that similar to the SF12 (scales are in the opposite direction), health measured by PCPR deteriorates with age. In addition, men have higher health levels than women, higher educated individuals are more healthy than lower educated individuals, and those in single households are less healthy than those who live with a partner or spouse. Thus, PCPR and the SF12 have the same pattern of distribution across these explanatory variables. This gives reason to believe that PCPR captures several of the same underlying health dimensions as the SF12 in relation to the explanatory variables of interest in this study, and that PCPR therefore can be used to assess how the SES–health relationship changes across age in a way that is comparable to the SF12.

In addition, the SF12 and the PCPR are both generic health measures, as they contain elements of functional ability and general health, unlike more disease-specific register health measures. The PCPR and the SF12 contain some element of self-assessed health because the initiative to visit a PCP is based on how health is experienced by the individual, whereas for the SF12, the survey respondents are asked about how they experience health in several settings. In addition, both measures contain some elements that are more concrete expressions of health, such as the PCP's assessment of the patient's health and thus, the services provided. In the SF12, this is incorporated by the questions of a more technical nature that include the respondent's ability to climb stairs, vacuum, go hiking, or garden. Importantly, the PCPR and the SF12 are also very different and, therefore, unsuited to compare health *levels* across measures, or to assess whether the educational or gender differences in health are larger or smaller in either data source. These differences imply that we cannot use the PCPR to assess the magnitude of health selection in the NorLAG sample. In Table 3, I present standardized means to further facilitate comparisons between the two health measures.

Finally, a measure of health that is available in both data sets is mortality. Unfortunately, the year of death is not available in the NorLAG data; therefore, comparing mortality hazard rates is not an option. However, in Table 4, I show how mortality from 2002 to 2007 is allocated across age, education, and cohabitation status in the two data sources.

Mortality in the NorLAG sample is defined as the fraction deceased at T2 (2007) of those who responded at T1 (2002). In the register data, mortality is defined as the fraction deceased in 2007, contingent on being alive in 2002. Overall, mortality corresponds fairly well in the two data sources. However, the fraction who died between 2002 and 2007 is somewhat larger in general, compared to the NorLAG, indicating that those who respond at T1 are likely to be selected on good health, as they have a lower likelihood of dying within the next five years. In addition, the lower educated in the NorLAG sample have

lower rates of mortality indicating that the NorLAG respondents in the lowest educational group are more selected with respect to good health compared to this group in the population.

In sum, Tables 1 to 4 provide evidence supporting that the NorLAG survey is fairly representative compared to the underlying population, except for education, in which case the weights do a good job of correcting the means. Moreover, the health measures from the two data sources follow the same distribution over age, education, and cohabitation status. Finally, the mortality patterns in the two data sources suggest that the NorLAG sample is positively selected with respect to health, especially for the lower educated group.

The previous results assessed the initial non-response in the NorLAG, that is, selection compared to the underlying population. I now turn to assessing attrition across waves in the NorLAG data. This is interesting for three reasons. First, assessing whether the data loss happening across time correlates with health is important for longitudinal analyses. Second, the data loss across waves is a combination of time gone by and of aging of the sample. Thus, the data loss occurring across waves can be indicative of the selection mechanism occurring across age in a cross-sectional sample. Finally, the data loss from one period to the next potentially occurs along the same dimensions as the initial non-response. Provided this is true, it implies that the selective attrition mechanism (partially) corresponds to the initial sample selection mechanism. Thus, finding evidence of selective attrition is indicative not only of potential bias in longitudinal analyses using the survey data but also of sample selection across age compared to the full population.

To assess the magnitude of selective attrition in the NorLAG survey, I perform two exercises: probability models of non-response at T2 based on relevant observable characteristics at T1 and the variable-added test suggested by Verbeek and Nijman (1992) in line with Jones et al. (2006).

First, to assess how non-response at T2 correlates with covariates of interest, in particular past health, the following linear probability model is estimated:

$$P(a_i = 1|x_i) = x_i\beta, \tag{1}$$

where $a_i = 1$ if individual i fell out of the sample from T1 to T2. The dependent variable, thus, is the likelihood of non-response at T2, conditional on responding at T1. x_i is a vector of explanatory variables. I employ three versions of x_i. In the first model "Simple", I include no controls besides the age and educational dummies. In the second model "P. Weights", I employ the same controls as those applied by Statistics Norway to make the population weights. This implies that x_i includes *centrality* and *region of residence*

in addition to the age and educational dummies. Finally, the model "Health" adds the SF12 to the x_i. This provides a systematic overview of what predicts non-response at T2 (conditional on response at T1). The results are presented in Table 5.[15]

Table 5 shows that age and education are strong predictors of attrition. Those who are older and have less education are more likely to fall out of the sample. The other variables that go into making the population weights, centrality and region, do not predict attrition. This does not necessarily mean that these variables do not predict initial non-response. Lappegård and Veenstra (2010) show that region and centrality do predict initial non-response, however, only slightly. Finally, those with worse health at T1 are less likely to respond at T2, thus indicating that there is health-related sample selection in the NorLAG data.[16] However, according to the adjusted R^2, including health adds some explanatory power in predicting attrition between the waves for men, but none for women.

In the second test of selective attrition, the variable "number of waves responded to" is added to the model of the main outcome (derived in Equation 2 below) according to Verbeek and Nijman (1992) and Jones et al. (2006) by using a pooled sample of all responders at T1 and T2, thus, providing a test of whether the number of waves an individual responded to (one or two) is associated with health after controlling for relevant variables, in this case, the age and education dummies. The assumption is that those who respond to more waves have better health than those who respond to fewer waves, thus showing that health-related attrition across waves is a potential problem.

The results from the variables added-test are presented in Table 6. The coefficients "No. of waves" are significant for both men and women, meaning that there is a correlation between health and the number of waves responded to, even when age, education, and their interactions are controlled for. Moreover, this correlation persists when those who fell out of the sample at T2 due to mortality are excluded, ensuring that the correlation between health and non-response is not driven by mortality.

5 Case Study and Sample Selection Bias Corrections

In this section, I present the empirical case study applied to test the selection bias, and the various selection correction methods that are applied in this analysis. Despite the large amount of literature on sample selection, no method is considered superior. This

[15] A fuller probit model, based on the derivation of the IPWs described below, and a simpler probit, which is the first stage of the Heckman, are presented in Table A1 and Table A2 in the Appendix.

[16] As shown in Table A1 in the Appendix, there is larger selection based on past self-rated health compared to past SF12. For this reason, as a robustness check, the main analysis (outlined in the next section) was conducted using self-rated health as the health outcome. These results does not change the main conclusions reached and are presented in the Appendix.

is in part because the impact of the sample selection depends on the question at hand and the available data, and thus, the correct method for selection correction corresponds to the selection mechanism in the data and the empirical question. The exact nature of the sample selection in the NorLAG data is unobserved. Therefore, I make various assumptions about the selection mechanism and assess how the estimates in a regression of SES and age on health are affected by the corresponding selection correction. For comparability, I assess the sample selection at T2 across all specifications.

I start by assuming that the selection is based on observable factors. Under this assumption, consistent estimates can be obtained by applying weights (Fitzgerald et al., 1998; Wooldridge, 2002). Then, by assuming that the sample selection is based on unobservable factors, I apply the Heckman and Copula selection models. Finally, I compare the results from the survey data to the results using data without non-response, namely full population register data. This implies that we can assess the impact of initial non-response, not just attrition, in the NorLAG data. It is plausible that the initial selection is stronger than the selection that takes place across waves, and finding that the bias caused by the latter is small does not necessarily mean that the former sort of selection is not problematic. The register data approach can thus uncover some implications of initial selection for analyses of health in old age. None of these approaches are without problems, but the idea is that the sum of selection corrections will bring us closer to definite answers about whether non-response in survey data causes bias in estimates of health inequalities in later life using a typical sample of older adults.

5.1 Case Study: Social Inequalities in Health across Age

To assess the bias from health-related sample selection established above, an empirical application is needed. Here, I employ the relationship between health and education across age. This is a case study particularly prone to bias from health-related non-response as health, education, and age correlate with survey non-response. I assume the following relationship:

$$Health_i = \alpha + \sum_{j=2}^{8} \delta_j Age_{ij} + \sum_{k=2}^{3} \beta_k SES_{ik} + \sum_{j=2}^{8} \sum_{k=2}^{3} \gamma_{jk} Age_{ij} \times SES_{ik} + u_i, \qquad (2)$$

for k=1–3 and j=1–8. $Health_i$ is the health outcome for individual i. Age include $j-1$ age group indicators (50–54, 55–59, .., 80–84, with the age group 45–49 serving as the reference category, and SES are education level indicators (upper secondary and higher education, respectively, with lower secondary education serving as the reference level). $Age_{ij} * SES_{ik}$ are interaction terms that allow the health–education relationship

to vary across age. This specification does not include any controls as this is merely a descriptive analysis of differences in patterns of health across age for different educational groups. Moreover, this is a cross-sectional representation of health across age, and not a representation of an individual's health trajectory across the life-course in later life.

The measure of interest is the marginal effect of education on health, evaluated at different age groups, and is given by:

$$\frac{\partial Health}{\partial SES_k} = \beta_k + \gamma_{kj} Age_j, \tag{3}$$

for k=2–3 and j=1–8.

Importantly, as non-response was shown to correlate with age, education, and health (even after conditioning on education and age), there is a correlation between u_i and Age_{ij} and SES_{ik} in Equation 2, and therefore, omitted variable bias. Hence, the coefficients in Equation 2 and 3 cannot be identified unless sample selection is properly accounted for.

5.2 Correcting for Sample Selection Bias

5.2.1 Selection on Observables

To formalize, let r_i be response in a given wave, h_i the health level, and x_i the observable controls for individual i. If we have $P(r_i = 1|h_i, x_i) = P(r_i = 1)$, then the likelihood of response does not vary systematically with the health level (or controls), and the non-response can be ignored.[17] OLS estimation of Equation 2 will yield unbiased coefficients. This is the assumption for the "baseline" approach.

If, instead, we have that:

$$P(r_i = 1|h_i, x_i, z_i) = P(r_i = 1|x_i, z_i), \tag{4}$$

then, conditional on x_i and z_i, r_i and h_i do not vary systematically.[18] For Equation 4 to hold, the conditional independence condition (CIC) must be satisfied. The CIC requires that there exists some variable, z_i, that predicts r_i and is endogenous to h_i (must correlate with h_i after controlling for x_i). In addition, z_i must be left out of the main equation. If the CIC hold, Equation 2 can be consistently estimated by weighted least squares (WLS) (Fitzgerald et al., 1998; Jones et al., 2006; Wooldridge, 2002). This approach is referred to as selection on observables (Fitzgerald et al., 1998).

In practice, weighting entails giving more weight to observations that are less likely to be in the sample, based on their observable characteristics. Weighting applies to both initial selection and to attrition. In the first case, population weights are appropriate. In

[17] This is often referred to as data missing completely at random (MCAR) (Fitzgerald et al., 1998).
[18] This is often referred to as data missing at random (MAR) (Fitzgerald et al., 1998)

the latter, IPWs are appropriate. IPWs imply calculating the likelihood of responding, based on the observable characteristics from past waves, and weighting the data by the inverse of these probabilities (Wooldridge, 2002; Michaud et al., 2011).

In this study, I estimate inverse probability weights based on the probability of response in the second wave of the NorLAG, given observable characteristics. The following probit is estimated:

$$P(r_i = 1|x_i, z_i) = \Phi(x_i\beta + z_i\gamma), \tag{5}$$

where $r_i = 1$ if individual i responded at T2. Thus, the dependent variable is the likelihood of response at T2, conditional on responding at T1. x_i includes age group indicators (50–54, 55–59, .., 80–84) and education indicators (upper secondary and higher education, respectively). The z_i vector comprises past health (the SF12, indicators for each level of self-rated health, an indicator for physical health limitations, and an indicator for recently severely reduced health), an indicator for manual occupation, if one's mother is alive, being retired, living with a partner, income by quartiles, and an indicator for whether the individual was working or had worked before retirement. The models are estimated separately for men and women.

What have been suggested as good instruments for z_i are lagged variables of the dependent variable and intermediary factors on the pathway from x_i to h_i (Fitzgerald et al., 1998). In this case, the first is past health, and the latter is occupation, income, retirement, disability, and work status. Clearly, the latter correlates with education (and age) and health. This approach is therefore primarily applicable to structural problems. That being problems where the interest lies in a theoretical relationship between h_i and x_i, rather than a reduced form problem as the exclusion of z_i would create bias in the latter. In this study, I claim no causal relations as I merely assess average health by education and age group. Thus, leaving the z_i out of the main equation (Equation 2) does not cause bias. Finally, this approach controls only for health risks that are in the history of the individual, not for idiosyncratic shocks to health that cause someone to drop out of the sample.

Whereas the IPWs correct for attrition from the first to the second wave, the population weights derived by Statistics Norway can be applied to assess initial selection. Population weights account for the sample's lack of representativeness compared to the population of interest, in this case, the entire Norwegian population aged 45–85 in 2007. The population weights are based on observable characteristics that are known in the entire population. In short, all individuals are initially given the same weight and then the weights are calibrated based on gender, age, education, region, and centrality to match

the observed shares in the population.[19] These weights are not derived to control for health-related sample selection in particular. Instead, they are derived with the purpose of making the sample averages comparable to the underlying population of interest. The assumption needed for this analysis is that these variables correlate with health; thus, applying the population weights will partially correct for the health-related non-response.

Estimation by the population weights provides insight into how sensitive the estimates are to initial non-response, whereas the IPW estimation can provide some insight into whether the attrition happening over time affects the estimates. In both weighting approaches, Equation 2 is estimated using weighted least squares. For the weighting approach to provide consistent estimates of Equation 2, the CIC must be satisfied. For the population weights, this implies that region and centrality are endogenous to health and correlate with selection (as these weights are already created by Statisticts Norway, and education, age, and gender are a part of x_i). The latter is barely satisfied. As Lappegård and Veenstra (2010) show, region and centrality only slightly predict non-response. The IPWs, on the other hand, are derived with the purpose of satisfying the CIC.

Finally, as the IPWs imply selection correction of the balanced sample and the population weights imply correction for the unbalanced sample, I provide the marginal effects (Equation 3) derived by the OLS estimates of Equation 2 based on each of these samples. In addition, comparing the OLS estimates based on these sub-samples provides a test of sample selection bias (Jones et al., 2006), under the assumption that those who respond to both (all) waves are more selected, in this case with respect to health, than those who respond only to one. If the estimates differ, it is likely that the non-response will cause omitted variable bias.

Moreover, comparing the WLS estimates to the OLS estimates is a test of selection bias–provided the selection is based on observables. The assumption that the sample selection is based only on observables is strong. We cannot exclude that the selection mechanism is based on unobservable factors, in which case Equation 4 does not hold. In this case, selection models are more appropriate.

5.2.2 Selection on Unobservables

Selection models rely on assumptions about the conditional underlying health distribution (latent health) and about the sample selection mechanism, and are given by the two equations:

$$h_i^* = x_{1i}\beta_1 + u_{1i}, \tag{6}$$

[19]The weights are calculated by Statistics Norway using CLAN. As Statistics Norway has individual-level register data covering these variables for the entire Norwegian population, these weights are of very high quality.

where h_i^* is the underlying health distribution in the population, x_{1i} are the controls (education and age group dummies), and u_{1i} is the error. As we observe health only for those who answer the survey, h_i^* is unobserved (latent). Instead, we observe h_i according to the selection equation:

$$r_i = x_{2i}\beta_2 + u_{2i}, \tag{7}$$

where r_i is the likelihood of response, which takes the value 1 if i responded at T2, x_{2i} are the controls, and u_{2i} is the error, possibly correlated with u_{1i}. x_{1i} is nested within x_{2i}.

The conditional expectation function is

$$E[h_i^*|x_{1i}] = E[h_i|x_{1i}, r_i \geq 0] = x_{1i}\beta_1 + E[u_{1i}|u_{2i} \geq -x_{2i}\beta_2], \tag{8}$$

which is the conditional mean for the survey responders. If there is no health-related selection, that is, if the errors, u_{1i} and u_{2i}, are uncorrelated, Equation 8 can be estimated by OLS without the non-response causing bias. However, as shown above, there is correlation between health and non-response, conditional on the controls (Table 5), which implies that the sample selection can be seen as an omitted variables bias problem. The insight by Heckman (1979) was that as such, Equation 8 can be consistently estimated by a linear regression of h_i on x_{1i} and an additive non-linear function of x_{2i}, where the latter is estimated by the inverse of Mill's ratio, given by $\lambda(x_{2i}\beta_2) = \phi(x_{2i}\beta_2)/(1 - \Phi(x_{2i}\beta_2))$ Accordingly, we have

$$E[h_i^*|x_{1i}] = E[h_i|x_{1i}, r_i \geq 0] = x_{1i}\beta_1 + \sigma\lambda(-x_{2i}\beta_2) \tag{9}$$

which is an estimable equation. Heckman (1979) proposed a two-step estimator that separately estimates the selection equation in a probit to obtain estimates of the inverse Mill's ratio, before controlling for this in an OLS regression of x_{1i} on h_i.[20]

Identification of the sample selection, estimated by $\sigma\lambda(-x_{2i}\beta_2)$, hinges on the non-normality of the λ, which is approximately linear over a large range of its argument. If x_{1i} and x_{2i} contain the same variables, the selection equation is poorly identified due to multicollinearity problems. Therefore, it is common to include an exclusion restriction to increase the robustness of the identification. This involves finding a variable that predicts non-response but is uncorrelated with the error in the main equation.[21] Here, I propose

[20] The standard errors must be corrected accordingly, but this is typically implemented in the statistical software.

[21] Unlike in the case of selection on observables where the CIC was satisfied by a variable that is correlated with selection and health but is excluded from the main equation.

the year in which the interview took place at T1. Respondents who were interviewed in 2003 are substantially less likely to respond at T2 compared to respondents interviewed in 2002. The reason is that those who responded in 2003 were less inclined to respond to the survey and had to be contacted several times before a response was obtained. They are marginal responders. The year of the survey response, however, is not correlated with health at T1. I show this in a regression of the year of response at T1, education, and age, on health. The results are presented in Table 7. This table also includes results showing the coefficients for how the year of response at T1 correlates with the likelihood of responding at T2, holding age and education constant. Estimation is done separately for men and women.

All though the Heckman model has been criticized, it is widely applied in empirical work (Puhani, 2000). Moreover, it is considered the textbook example of sample selection models (Cameron and Trivedi, 2010; Wooldridge, 2010), and is therefore included in this study. A central critique is that the Heckman two-step estimator is less efficient than maximum likelihood estimation (MLE) (Nelson et al., 1984; Puhani, 2000). Thus, I include results estimated by maximum likelihood in addition to the two-step estimates. Moreover, the MLE estimates are appropriate for comparing estimates derived by the Copula outlined below. The likelihood is given by

$$\mathcal{L} = \prod_{i=1}^{N} (Pr(r_i < 0))^{1-r_i} (f(h_i|r_i > 0) Pr(r_i > 0))^{r_i} \tag{10}$$

For consistent estimates, the errors (u_{1i}, u_{2i}) must be bivariate normal and independent of x_{2i}, with correlation ρ, and variances σ^2 and 1, respectively.[22]

A central critique of the Heckman is the bivariate normality assumption. If the errors are not bivariate normal, then the model does not sufficiently control for selection, leading to inconsistent estimates. The Copula is a Heckman type of selection model estimated by maximum likelihood and where the joint normality assumption is relaxed. The copula approach implies replacing the bivariate distribution by two marginal distributions bound together by a copula function.[23] This implies specifying the marginal selection distribution, $f(r|x_2)$, and the marginal outcome distribution, $f(h|x_1)$, separately, and a copula function that describes the relationship between the two distributions, $C[f(r|x_2), f(h|x_1), \theta]$, where θ is the dependence parameter in the copula that governs the degree of dependence.

According to Sklar's theorem, all multivariate distributions have a copula representa-

[22] Assuming unity for the variance of u_{2i} can be done without loss of generality, as the selection equation is without scale in the estimations.

[23] This is due to Sklar's theorem which states that any multivariate distribution can be described by its underlying marginal distributions bound together by a copula function (Smith, 2003).

tion. A few families of couplas has been identified by Smith (2003) as particularly well suited for selection models due to their mathematical properties, such as joint asymmetry and excess joint skewness.[24] These families of couplas, and a few others, have been implemented in statistical software. To estimate the Copula model, I use the *heckmancopula*[25] command developed by Hasebe et al. (2013).[26]

The limitations of the copula approach lies in the ability of the copula to describe the dependence pattern between the marginal distributions of the selection and the main outcome. Choosing the right copula, requires some *a priori* insight into the dependence between the two. If the dependence between these two marginal distributions is unknown, then a proposed solution is to apply various copulas and choose the copula associated with the largest log-likelihood according to the Baysian Information Criteria (BIC), presupposing that the number of parameters does not differ across models (Smith, 2003; Hasebe et al., 2013).

Here, the actual dependence between health and the likelihood of responding at T2 is unobserved, as is common in selection models due to the latent structure of the dependent variable. Therefore, I apply the various families of copulas that are currently developed for selection models and available in the statistical software. These copula families include the Product, Gaussian, Ali-Mikhail-Haq, Gumbal, Plackett, Clayton, Frank, Joe, and Fairlie-Gumbel-Morgenstern (Hasebe et al., 2013). The Gaussian, Plackett, Frank, and Fairlie-Gumbel-Morgenstern exhibit various degrees of symmetric tail dependence, whereas the Ali-Mikhail-Haq, Gumbal, Clayton, and Joe exhibit various degrees of asymmetric tail dependence. The Product assumes independence between the two marginal distributions. The copula that is best suited according to the BIC is then included in the main analysis. In addition, I assess whether the best-fitting copula is an improvement compared to the Heckman MLE by applying the Vuoung test (Hasebe et al., 2013). This is described in more detail in Section 6.1.1 below.

5.2.3 Comparing Results Based on Survey Data to Results Based on Register Data

The final part of the bias assessment entails comparing the marginal effects (Equation 3) using the NorLAG data to the results using the full population register data. Therefore, having a measure of health from the registers that is comparable to the SF12 is essential. As the register data contain only administrative records of health care utilization (and

[24] These include the Ali-Mikhail-Haq, Clayton, Frank, Gumbel, and Joe.

[25] Copula estimation using the gaussian copula and assuming that the marginal distributions of health and selection are normal is the same as estimation by heckman MLE.

[26] See Smith (2003) for more about copula selection models in general and Hasebe et al. (2013) for copula estimation in Stata.

mortality), no perfect counterpart to the SF12 exists in the registers. For the purpose of this analysis, I argue that the sum of reimbursements for all PCP consultations in a given year can be applied. First, similar to the SF12, PCPR is a generic measure of health. Second, reimbursements are shown to have patterns of distribution across age, SES, and gender that are comparable to those of the SF12 (Table 2). As, the register data do not have non-response issues, the distribution of PCPR reflects the actual distribution of health in the population and therefore, can serve as an unbiased distribution of health with which we can compare to the SF12. If SES and health are shown to have the same relationship across age in the survey as in the register data, then we have an indication that selective non-response is not what is causing the observed convergence in health in later life. Rather, this is something that happens at the population level as well.

As explained in Section 3, secondary health care utilization is excluded because of the worry that the expression of utilization reflects not only health but also SES. However, as a robustness check, I provide results using hospital admissions as the main health outcome. Hospitalization is a sub-group of the specialist care services that arguably has smaller bias because admission to the hospital is not granted unless the patient is severely ill. However, hospitalizations, either acute or planned, are granted based on the medical staff's assessment of need. Therefore, hospitalizations are likely to reflect differences in utilization due to SES in addition to the presence and extent of severe health problems. This implies a downward bias in the educational differences in hospitalizations. In addition, being hospitalized is a non-generic and extreme health outcome in that it reflects very poor health, and therefore, is not well suited for a comparison with the SF12. Results for hospitalizations are shown in Appendix Graph A1.

6 Results

6.1 Sample Selection Correction in the Survey Data

The results from the different approaches using the NorLAG survey data are presented in Figure 2 (men) and Figure3 (women). The estimation tables for all specifications are shown in Appendix Tables A3 through A5. All graphs show the marginal effects (Equation 3) of education on health across 5-year interval age groups. The graphs show the difference in health for upper secondary (solid line) and higher (dotted line) education compared to lower secondary education (reference), respectively, as the aim here is to study convergence in health between educational groups across age.[27]

Graph A. Unbalanced Sample, in Figure 2, shows the estimation results using OLS

[27]In addition, I ran the whole analysis excluding individuals who fell out of the sample due to mortality. This had no impact on the estimates. Only 265 individuals (4.8 percent) of the responders at T1 died between T1 and T2.

on the gross sample of responders at T2 for men. In general, a pattern of converging health inequalities, consistent with the AAL theory, is found. Inequalities are largest for the 60- to 64-year-old age group, after which they decrease across the older age groups, except a spike for the 75- to 79-year-old age group for the highest educated group.

Graph B. Balanced sample, shows the OLS estimation results using only the sub-sample who responded to both waves at T2. These results confirm the AAL for the "more selected" sub-sample. The educational differences in health range from about 7 points (on the SF12 scale) for the 60- to 64-year-old group, to almost zero for the oldest age groups. Assuming that the balanced sample is more selected than the unbalanced sample, we would expect smaller educational differences in Graph B compared to Graph A. In general, this is not the case. However, there is an apparent faster convergence of the inequalities for the age groups 60- to 64-year-old and 65- to 69-year-old age groups in Graph B.

In this application, sample selection is assumed to bias the estimates of the educational differences toward zero, and this bias is assumed to increase with age. This implies that we would expect larger educational differences and weaker convergence in the bias-corrected estimates, provided the proposed selection correction methods are appropriate. Specifically, assuming that the AAL pattern is caused by sample selection the selection correction would imply a smaller or no convergence in the educational differences in health for the older age groups.

The two weighting approaches are shown in Graph C. Pop. Weights and Graph D. IPW. The coefficients are estimated using weighted least squares, as outlined in Section 5. Importantly, the population weights are derived with respect to the gross sample at T2, whereas the IPWs are derived with respect to the balanced sample at T2, as we need the information at T1 to allocate the weights to each observation at T2. The weighted estimates do not differ substantially from the OLS estimates, except a smaller convergence for the oldest age group for the higher educated using the population weights. However, this does not change the empirical support for the AAL theory.

The two lower graphs, E. Heckman and F. Copula, show the marginal effects estimated by the Heckman and the Copula (more about the choice of copula in Section 6.1.1 below). Because data on the explanatory variables are needed for the estimation of both selection models, the selection models are based on responders and non-responders at T2, provided information about age and education was obtained at T1.[28]

The inequalities estimated by the Heckman are marginally larger for all age groups

[28]In this application, educational group is based on the education reported at T1. It is unlikely that someone in these cohorts undertake an amount of education that would change education from one category to another.

compared with the OLS estimates of the balanced sample. This results is consistent with the assumption that selection causes the estimates to be biased toward zero. Despite larger inequalities in all age groups, the Heckman estimates still provide evidence of converging inequalities for the older age groups. The effects estimated by the Copula are slightly larger than those by the Heckman, and there is still evidence of convergence, albeit less clear. The choice of copula is not important for the pattern of convergence.

The results for women (Figure 3) provides fairly similar conclusions. Except for the highest educated in the oldest age group. The results for this group should be handled with care as very few of these women have higher education (this group contain only 96 observations for the unbalanced sample and 51 for the balanced sample). The convergence in health inequalities across age is visually stronger for women than for men, particularly due to the large differences between both higher and upper secondary education and lower secondary education for the age group 55-59. As for men, the Copula results for women display a slightly weaker convergence compared to the other models.

6.1.1 The Choice of Copula

The main difference between the copulas is how they capture the dependence between health and non-response. The copula that provided the best fit according to the BIC (the largest log-likelihood) was the Joe, and thus, this is the copula shown in Figure 2 and Figure 3. This copula exhibits asymmetric, right tail joint dependence, which complies with the assumption that those with higher levels of health have a higher likelihood of responding.

Moreover, according to the Vuoung test, the Copula using the Joe copula is a statistically significant improvement compared to the Heckman MLE. The Vuoung test implies calculating the contribution to the log-likelihood function of each observation in the Heckman MLE and in the Coupla MLE, and then testing whether there is a statistically significant difference between the contributions to the log-likelihood in these two models. As shown in Table 8, the log-likelihood of the Joe copula estimation compared to the Heckman was -8048.569 compared to -8128.966 for men and -8859.473 compared to -8813.156 for women. The OLS test of the difference between the contribution of each observation to the log-likelihood function is statistically significant with a t-value of -16.67 for men and -14.07 for women. The results for this test are presented in Table A6 in the Appendix.

In addition, unlike in the Heckman, the Copula provides evidence of sample selection. The p-value under the test of independence between u_{1i} and u_{2i} is below 0.001 for men and women. In contrast, the p-values based on the Heckman are 0.499 and 0.831 for men and women, respectively. This is also evident by the statistical significance of the θs,

whereas the λs are not statistically significance.

6.2 Comparing Results Based on Survey Data to Results Based on Register Data

The results for men and women using the panel data are displayed in Figure 4. The figure presents the marginal effects in Equation 3 estimated by OLS using the register data in 2007. The dotted line represents higher education, and the solid line represents upper secondary education. Lower secondary education is the reference category. As the outcome here (PCPR) is expressed in monetary terms, higher values indicate worse health outcomes. The higher educated have better health outcomes at all age groups, but these differences decrease across age. Thus, the results from the register data supports the finding in the survey data and contradict the hypothesis that bias due to selective non-response drives the convergence in later life health in studies using survey data. This result is also supported by the analysis using hospitalizations (shown in the Appendix).

A few cautions are worth mentioning at this point. First, the estimated trajectories cannot necessarily be ascribed to an age effect on health due to the cross-sectional nature of the data. Rather, the observed health-age relationship is a combination of age and cohort effects. Second, the results do not provide identification of whether social inequalities in health are diverging or converging across age in later life due to the cross-sectional and descriptive nature of the estimated health trajectories. Finally, it is possible that neither of the selection approaches fully controls for sample selection bias in the empirical example provided, as they all rely on assumptions that are not necessarily satisfied. The weighting approach hinges on selection being based on the observable factors that go into making the weights, and the Heckman on joint normality of the errors, both of which are strong assumptions. The results from the Copula show that there is a correlation between the errors in the selection and the main outcome equations, thus indicating that selection is based on unobservables and contrasting the results from the Heckman. The results from the Heckman and the weighting approaches, therefore, are likely to suffer from selection bias due to insufficient selection correction. Despite this, the bias from sample selection is small as the pattern and sizes of the educational differences in health are similar across models, including the Copula. In addition, the register approach hinges on the comparability of the SF12 and the PCPR. Without having access to exactly the same measure of health in the register and survey data, it is difficult to be specific about the selection bias.

Although register data do not suffer from non-response, there is the possibility that the convergence in health results from mortality selection. Individuals with poor health have a higher likelihood of premature death and are concentrated among the low educated.

This causes the remaining survivors in the low educated group to be less representative of that group as a whole compared to the higher order SES groups. Those alive at older ages are likely to have had higher health levels compared to the whole group of low SES being alive, at any past age, thus causing healthy survivor bias. This assumption is tested in Figure 5.

In these graphs, I exploit the full panel of the register data and create health trajectories across 2007–2014 for the cohorts born in 1927, 1937, and 1947. This entails trajectories of average health across the age spans 80–88, 70–78, and 60–68, respectively. For ease of presentation, I include only the results for the educational groups higher education and lower secondary education (upper secondary education is excluded). I divide each group into two groups where one group contains all individuals in the respective cohort alive at each age (the gross sample in each year 2007–2014), and the other group contains only those who are alive across the whole age span (those who are alive until at least 2014). The latter group contains only those who are healthy enough to survive the whole period, that is only the healthy survivors.[29]

As evident in Figure 5, those who survive the whole period have better health at all ages compared to the gross sample. The curves for the low educated survivors (Surv. Low Edu) and the higher educated survivors (Surv. High Edu) are lower than the curves for the gross sample at each age group for the low educated group (All Low Edu) and higher educated group(All High Edu), respectively. The difference is largest in the beginning of the panel, 2007, and falls across age/time as those with poor health falls out of the gross sample (out of the "All" groups). In the final year, 2014, the health levels between the two groups have leveled, as only those alive throughout the panel are left in the gross sample. The difference between the survivors and "all" is largest for the oldest age groups and for the lowest educated group. Whereas there are practically no differences between the survivors and "all" for the highest educated group born in 1947, the difference is larger for the lowest educated group of this cohort. Moreover, for the oldest cohort, born in 1927, there are larger differences also among the high educated.

Finally, the differences between the educational groups are larger for the younger cohorts compared to the older cohorts, supporting the AAL theory. However, this result can also be due to cohort effects. For the cohort born in 1947, the educational differences are large, both between the surviving and the gross samples. For the cohort born in 1927, however, the group of (healthy) survivors in the lowest educated group have better health than the gross sample with high education. This is evident by the line for "All High Edu", lying above the line for "Surv. Low Edu". Although there is a strong tendency towards reduced health inequalities across age across the graphs (cohorts), within each cohort,

[29]Of course, idiosyncratic shocks to health can also cause someone to drop out.

there is little evidence of convergence besides for the oldest cohort. Therefore, whether the convergence is related to increasing age or a result of cohort effects, remains an open question.

7 Conclusion

Statistical inference about health is typically based on survey data. This is due to the limited availability of population-level data containing information about health and the feasibility of obtaining generic measures of health in surveys (compared to, e.g., administrative registers). Health might be a particularly important predictor of non-response in survey samples, and health-related non-response becomes even more pronounced in samples of older individuals. Studies of health and aging based on survey data, therefore, are particularly prone to bias due to selective non-response.

In this paper, I assess the bias from health-related sample selection in a survey of health in the second half of life. Specifically, I assess whether the (well-established) pattern of converging inequalities in health in old age is caused by sample selection. Bias from selective non-response in the second wave of the NorLAG survey is assessed compared to the survey's reference population using full population register data and population weights, and compared to those who responded to the first wave (attrition) using more conventional sample selection methods such as IPWs and selection models.

I find that the health trajectories estimated by the different weighting and selection models do not differ substantially from the trajectories estimated by OLS. This result corresponds to the literature on sample selection, where, despite evidence of selective non-response in the data, the bias is found to be small or negligible. For all specifications, I find that the educational differences in health are falling across age in later life, as is a common result in the social gerontological literature referred to as the AAL theory. Moreover, the analysis using the full population register data confirms the pattern of converging health inequalities in old age. Thus, the survey and the full population register data lead to the same conclusion.

Taken together, these results suggest that selective non-response in the survey data does not cause the empirical evidence of converging health inequalities in later life. Due to the cross-sectional and descriptive nature of the age trajectories, the results in this analysis do not provide identification of whether social inequalities in health are diverging or converging across age in later life. This analysis suggests that survey data are well suited to describe patterns of health inequalities as they are manifested in the population, as the bias due to selective non-response does not change the observed patterns. However, as suggested by the analyses using the full panel register data, selective mortality leads to stronger healthy survivor bias in the lower SES groups compared to higher SES groups.

An important issue in analyses of health and aging, is that the population alive at older ages provides a poor representation of the population alive at younger ages. Mortality selection, present even in the most perfect dataset, is a concern in analyses attempting to measure the relationship between groups with different mortality rates across time or age, or in analyses of later life outcomes in general, provided mortality is correlated with the outcome of interest.

The impact and magnitude of health-related initial non-response in survey data are still largely unresolved. This paper highlights the importance of allowing survey data to be combined with outcomes from health registers that are available for the full population. This would not only allow assessing the bias from health-related non-response more explicitly, as the exact health distribution in the population compared to in the survey data would be known. This would also allow assessing whether those in low SES groups and/or older age groups have lower response rates because these groups also have worse health outcomes, or if these characteristics are associated with lower response rates in themselves.

References

Becketti, S., Gould, W., Lillard, L., and Welch, F. (1988). The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics*, 6(4):472–492.

Benzeval, M., Green, M. J., and Leyland, A. H. (2011). Do social inequalities in health widen or converge with age? Longitudinal evidence from three cohorts in the West of Scotland. *BMC Public Health*, 11(1):947.

Biørn, E., Gaure, S., Markussen, S., and Røed, K. (2013). The rise in absenteeism: disentangling the impacts of cohort, age and time. *Journal of Population Economics*, 26(4):1585–1608.

Bjørshol, E., Høstmark, M., and Lagerstrøm, B. (2010). Livsløp, generasjon og kjønn. LOGG 2007. Dokumentasjonsrapport. *Notater*, 19:2010.

Brazier, J. E. and Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care*, 42(9):851–859.

Brown, T. H., O'Rand, A. M., and Adkins, D. E. (2012). Race-ethnicity and health trajectories: Tests of three hypotheses across multiple groups and health outcomes. *Journal of Health and Social Behavior*, 53(3):359–377.

Cameron, A. C. and Trivedi, P. K. (2010). *Microeconometrics Using Stata*, volume 2. Stata press College Station, TX.

Chandola, T., Ferrie, J., Sacker, A., and Marmot, M. (2007). Social inequalities in self reported health in early old age: Follow-up of prospective cohort study. *BMJ*, 334(7601):990.

Cheng, T. C. and Trivedi, P. K. (2015). Attrition bias in panel data: A sheep in wolf's clothing? A case study based on the mabel survey. *Health Economics*, 24(9):1101–1117.

Contoyannis, P., Jones, A. M., and Rice, N. (2004). The dynamics of health in the british household panel survey. *Journal of Applied Econometrics*, 19(4):473–503.

Elstad, J. I. (2018). Educational inequalities in hospital care for mortally ill patients in norway. *Scandinavian Journal of Public Health*, 46(1):74–82.

Fitzgerald, J., Gottschalk, P., Moffitt, R., et al. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *Journal of Human Resources*, 33(2):251–299.

Grasdal, A. L. and Monstad, K. (2011). Inequity in the use of physician services in norway before and after introducing patient lists in primary care. *International Journal for Equity in Health*, 10(1):25.

Hasebe, T. et al. (2013). Copula-based maximum-likelihood estimation of sample-selection models. *The Stata Journal*, 13(3):547–573.

Hausman, J. A. and Wise, D. A. (1979). Attrition bias in experimental and panel data: The Gary income maintenance experiment. *Econometrica: Journal of the Econometric Society*, pages 455–473.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Heiss, F., Venti, S. F., and Wise, D. A. (2014). The persistence and heterogeneity of health among older americans. Technical report, National Bureau of Economic Research.

Herd, P. (2006). Do functional health inequalities decrease in old age? Educational status and functional decline among the 1931-1941 birth cohort. *Research on Aging*, 28(3):375–392.

Hoffmann, R. (2011). Illness, not age, is the leveler of social mortality differences in old age. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 66(3):374–379.

Jenkinson, C. and Layte, R. (1997). Development and testing of the UK SF-12. *Journal of Health Services Research*, 2(1):14–18.

Jones, A. M., Koolman, X., and Rice, N. (2006). Health-related non-response in the british household panel survey and european community household panel: Using inverse-probability-weighted estimators in non-linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3):543–569.

Kim, J. and Durden, E. (2007). Socioeconomic status and age trajectories of health. *Social Science & Medicine*, 65(12):2489–2502.

Koløen, K., Lima, I., and Veenstra, M. (2013). Non-response and attrition in the NorLAG panel study. *Oslo: Norwegian Social Research*.

Lappegård, T. and Veenstra, M. (2010). Life-course, generation and gender: Logg 2007. *Field report of the Norwegian Generations and Gender Survey*.

Lillard, L. A. and Panis, C. W. (1998). Panel attrition from the panel study of income dynamics: Household income, marital status, and mortality. *Journal of Human Resources*, pages 437–457.

Manski, C. F. and Altonji, J. G. (1998). Special Issue: Attrition in Longitudinal Surveys. *The Journal of Human Resources*, 33(2).

Michaud, P.-C., Kapteyn, A., Smith, J. P., and Van Soest, A. (2011). Temporary and permanent unit non-response in follow-up interviews of the health and retirement study. *Longitudinal and Life Course Studies*, 2(2):145–169.

Mirowsky, J. and Ross, C. E. (2008). Education and self-rated health: Cumulative advantage and its rising importance. *Research on Aging*, 30(1):93–122.

Nelson, F. D. et al. (1984). Efficiency of the two-step estimator for models with endogenous sample selection. *Journal of Econometrics*, 24(1-2):181–196.

Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14(1):53–68.

Ross, C. E. and Wu, C.-L. (1996). Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior*, pages 104–120.

Salas, C. and Raftery, J. P. (2001). Econometric issues in testing the age neutrality of health care expenditure. *Health Economics*, 10(7):669–671.

Seshamani, M. and Gray, A. (2004). Ageing and health-care expenditure: The red herring argument revisited. *Health Economics*, 13(4):303–314.

Slagsvold, B., Veenstra, M., Daatland, S. O., Hagestad, G., Hansen, T., Herlofson, K., Koløen, K., and Solem, P. E. (2012). Life-course, ageing and generations in Norway: The NorLAG study. *Norsk Epidemiologi*, 22(2).

Smith, M. D. (2003). Modelling sample selection using archimedean copulas. *The Econometrics Journal*, 6(1):99–123.

Syse, A., Lyngstad, T. H., and Kravdal, O. (2012). Is mortality after childhood cancer dependent on social or economic resources of parents? A population-based study. *International Journal of Cancer*, 130(8):1870–1878.

The Norwegian Directorate of Health (2009). Styringsdata for fastlegeordningen, 4. kvartal 2009. Online; accessed 11 October 2018

https://helsedirektoratet.no/Documents/Statistikk%20og%20analyse/Fastlegestatistikk/Fastlegestatistikk%20hovedtallsrapport-2009.pdf.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, pages 127–169.

Verbeek, M. and Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, pages 681–703.

Ware Jr, J. E., Kosinski, M., and Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34(3):220–233.

Wooldridge, J. M. (2002). Inverse probability weighted m-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, 1(2):117–139.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Xu, X., Liang, J., Bennett, J. M., Botoseneanu, A., and Allore, H. G. (2014). Socioeconomic stratification and multidimensional health trajectories: Evidence of convergence in later old age. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 70(4):661–671.

Zabel, J. E. (1998). An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. *Journal of Human Resources*, pages 479–506.

Ziliak, J. P. and Kniesner, T. J. (1998). The importance of sample attrition in life cycle labor supply estimation. *Journal of Human Resources*, pages 507–530.

Zweifel, P., Felder, S., and Meiers, M. (1999). Ageing of population and health care expenditure: A red herring? *Health Economics*, 8(6):485–496.

Zweifel, P., Felder, S., and Werblow, A. (2004). Population ageing and health care expenditure: New evidence on the "red herring". *The Geneva Papers on Risk and Insurance-Issues and Practice*, 29(4):652–666.

Tables

Table 1: Summary Statistics for the Survey and Register Data

	Unbalanced	Pop. weights	Balanced	IPW	Register
Men:					
Age	61.32	59.55	61.63	62.59	60.31
Age 45-49	0.16	0.19	0.14	0.14	0.18
Age 50-59	0.30	0.33	0.31	0.30	0.34
Age 60-69	0.30	0.28	0.30	0.27	0.27
Age 70-79	0.18	0.19	0.18	0.21	0.15
Age 80+	0.06	0.01	0.06	0.09	0.06
Low. sec. educ	0.19	0.25	0.18	0.24	0.26
Upp. sec. educ	0.50	0.48	0.51	0.50	0.50
Higher educ	0.30	0.27	0.31	0.26	0.24
Cohabitation	0.77	0.77	0.78	0.75	0.75
Observations	2227	2107	1832	1774	819418
Women:					
Age	61.50	60.88	61.28	62.56	61.63
Age 45-49	0.16	0.16	0.15	0.14	0.16
Age 50-59	0.31	0.32	0.33	0.30	0.32
Age 60-69	0.28	0.28	0.29	0.27	0.26
Age 70-79	0.18	0.23	0.17	0.19	0.17
Age 80+	0.07	0.02	0.06	0.10	0.09
Low. sec. educ	0.24	0.32	0.23	0.29	0.33
Upp. sec. educ	0.46	0.43	0.45	0.45	0.46
Higher educ	0.30	0.26	0.32	0.26	0.21
Cohabitation	0.62	0.62	0.63	0.61	0.64
Observations	2367	2227	1933	1872	868506

Note: Mean values for covariates of interest in the NorLAG sample and register data in 2007. Columns 1–4 present results for the NorLAG, where the column "unbalanced" presents means for all T2 responders. Column "Pop. weights" presents unbalanced sample means weighted by population weights. Column "Balanced" presents unweighted means for the responders at T2 who also responded at T1, and columns "IPW" presents balanced sample means weighted by IPW. Column "Register" presents means for the register data.

Table 2: Mean Health Outcomes across SES, Age, and Gender

	SF12		PCPR	
	Men	Women	Men	Women
All	48.70	46.49	678	840
	(9.86)	(11.41)	(1258)	(1335)
Age 45-49	51.50	49.70	421	648
	(8.14)	(9.88)	(1156)	(1307)
Age 50-59	50.05	48.05	520	723
	(9.33)	(10.84)	(1114)	(1234)
Age 60-69	47.58	45.91	698	805
	(10.33)	(11.53)	(1174)	(1245)
Age 70-70	47.13	44.38	1039	1049
	(9.89)	(11.57)	(1458)	(1434)
Age 80+	44.42	39.65	1330	1295
	(11.07)	(12.30)	(1644)	(1602)
Low. sec. educ	45.51	43.23	864	1003
	(11.37)	(12.21)	(1502)	(1502)
Upp. sec. educ	48.61	46.26	668	818
	(9.82)	(11.43)	(1212)	(1280)
Higher educ	50.80	49.47	494	635
	(8.32)	(9.86)	(1003)	(1126)
Single	47.87	45.03	789	1034
	(10.65)	(12.11)	(1517)	(1543)
Cohabitation	48.94	47.39	640	732
	(9.61)	(10.86)	(1154)	(1192)
Observations	2195	2314	819418	868506

Note: Mean values (unweighted) for the health outcomes in the NorLAG unbalanced sample at T2 (SF12) and register data (PCPR) in 2007 at different values of the covariates of interest. For the SF12 (scale: 0-100), higher values indicates better health. Reimbursements (scale: 0-50000, extreme outliers are dropped) are measured in inflation adjusted Norwegian Kroner. Lower numbers indicate better health. Standard deviations in parentheses.

Table 3: Standardized Mean Health Outcomes

	SF12		PCPR	
	Men	Women	Men	Women
All	-0.00	-0.00	0.00	0.00
Age 45-49	0.28	0.28	0.20	0.14
Age 50-59	0.14	0.14	0.13	0.09
Age 60-69	-0.11	-0.05	-0.02	0.03
Age 70-79	-0.16	-0.18	-0.29	-0.16
Age 80+	-0.43	-0.60	-0.52	-0.34
Low. sec. educ	-0.32	-0.29	-0.15	-0.12
Upp. sec. educ	-0.01	-0.02	0.01	0.02
Higher educ	0.21	0.26	0.15	0.15
Single	-0.08	-0.13	-0.09	-0.15
Cohabitation	0.02	0.08	0.03	0.08
Observations	2195	2314	819418	868506

Note: Standardized mean values for the SF12 (NorLAG unbalanced sample at T2) and PCPR (register data in 2007) at different values of the covariates of interest. Standardization is done separately for men and women.

Table 4: Mortality from 2002 to 2007 across SES, Gender, and Age

	NorLAG		Register	
	Men	Women	Men	Women
All	0.06	0.04	0.07	0.05
Age 45-49	0.01	0.00	0.01	0.01
Age 50-59	0.02	0.01	0.02	0.01
Age 60-69	0.04	0.02	0.05	0.03
Age 70-79	0.18	0.08	0.13	0.08
Age 80+	0.33	0.30	0.28	0.18
Low. sec. educ	0.09	0.07	0.11	0.08
Upp. sec. educ	0.05	0.03	0.06	0.04
Higher educ	0.03	0.02	0.03	0.02
Single	0.11	0.06	0.05	0.04
Cohabitation	0.04	0.02	0.02	0.01
Observations	2,699	2,856	898,252	925,463

Note: Mortality between 2002 and 2007 in the NorLAG panel and the register data at different values of the covariates of interest. For the NorLAG data, mortality in 2007 is conditional on response in 2002. For the register data, mortality in 2007 is conditional on being alive in 2002.

Table 5: Attrition between the Waves

	Men			Women		
	Simple	P. Weights	Health	Simple	P. Weights	Health
45-49	0.003	-0.000	-0.005	-0.004	-0.004	-0.014
	(0.0330)	(0.0331)	(0.0330)	(0.0308)	(0.0310)	(0.0309)
50-54	0.000	-0.001	-0.012	-0.002	-0.002	-0.014
	(0.0324)	(0.0326)	(0.0326)	(0.0307)	(0.0308)	(0.0309)
55-59	-0.090***	-0.091***	-0.109***	-0.021	-0.019	-0.031
	(0.0316)	(0.0317)	(0.0315)	(0.0312)	(0.0313)	(0.0316)
60-64	-0.049	-0.050	-0.077**	0.012	0.013	-0.006
	(0.0331)	(0.0332)	(0.0331)	(0.0331)	(0.0332)	(0.0333)
65-69	0.027	0.022	-0.004	0.024	0.026	0.003
	(0.0366)	(0.0368)	(0.0365)	(0.0363)	(0.0364)	(0.0365)
70-74	0.176***	0.172***	0.141***	0.092**	0.093**	0.069*
	(0.0382)	(0.0383)	(0.0383)	(0.0369)	(0.0370)	(0.0372)
75-79	0.225***	0.224***	0.177***	0.244***	0.245***	0.215***
	(0.0406)	(0.0406)	(0.0409)	(0.0395)	(0.0397)	(0.0405)
80+	0.263***	0.261***	0.208**	0.378***	0.373***	0.330***
	(0.0908)	(0.0909)	(0.0921)	(0.0763)	(0.0781)	(0.0811)
Upp. sec. ed.	-0.150***	-0.149***	-0.130***	-0.107***	-0.101***	-0.088***
	(0.0230)	(0.0232)	(0.0233)	(0.0217)	(0.0221)	(0.0223)
Higher ed.	-0.267***	-0.264***	-0.225***	-0.229***	-0.218***	-0.200***
	(0.0245)	(0.0256)	(0.0261)	(0.0230)	(0.0238)	(0.0243)
Centrality		0.005	0.003		-0.017	-0.019
		(0.0213)	(0.0212)		(0.0218)	(0.0218)
Region		0.003	0.001		0.006	0.006
		(0.0040)	(0.0039)		(0.0038)	(0.0038)
SF12			-0.006***			-0.003***
			(0.0009)			(0.0008)
Constant	0.436***	0.423***	0.734***	0.393***	0.377***	0.516***
	(0.0300)	(0.0390)	(0.0593)	(0.0278)	(0.0391)	(0.0534)
Observations	2668	2654	2639	2844	2828	2804
Adj. R^2	0.089	0.088	0.105	0.074	0.074	0.078
Mean	0.317	0.317	0.316	0.321	0.321	0.318

Note: The likelihood of non-response at T2, conditional on responding at T1. Column "Simple" includes only the explanatory variables of main interest in this study. Column "P. Weights" includes the variables that go into making the population weights. Column "Health" adds controls for health. Coefficients are probit estimates. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 6: Variables-Added Test

	Men All	Women All	Men Alive	Women Alive
No. of waves	3.168***	2.028***	1.991***	1.618***
	(0.4667)	(0.5067)	(0.4737)	(0.5135)
Observations	2653	2820	2494	2722
Adjusted R^2	0.106	0.103	0.082	0.092
Mean	48.287	45.862	48.910	46.188

Note: The correlation between "number of waves responded to" and health at T1, when age, education, and their interactions are controlled for, i.e. coefficient is based on OLS estimation of Equation 2 (derived below) adding "number of waves responded to" as a covariate. The sample includes all those who responded to T1, T2, or both. To save space, the table displays only the coefficients for the number of waves. The first two columns include all respondents at T1, whereas the latter two exclude those who died between T1 and T2. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 7: Exclusion Restriction for the Heckman Selection Model

	SF12 at T1		Response at T2	
	Men	Women	Men	Women
Survey year T1	0.468	0.644	-0.191***	-0.185***
	(0.5971)	(0.6920)	(0.0264)	(0.0267)
Observations	2667	2829	2668	2844
Adjusted R^2	0.088	0.098	0.106	0.089
Mean	48.292	45.846	0.683	0.679

Note: This table shows the coefficients of the variable "Year of interview at T1" from an OLS regression on health at T1 and on the likelihood of responding at T2. Age and educational dummies are also included in the regressions but are suppressed in the table to save space. Standard errors in parentheses. *Mean* implies the mean of the outcome variables. * p<0.1, ** p<0.05, *** p<0.01.

Table 8: Model Characteristics

	Heckman		MLE		Copula	
	Men	Women	Men	Women	Men	Women
Lambda	4.243	1.434	0.709	0.309		
	(2.7651)	(3.1899)	(0.9487)	(1.3774)		
Theta					2.644***	2.242***
					(0.1280)	(0.1199)
LL			-8128.966	-8859.473	-8048.569	-8813.156
p-value			0.499	0.831	0.000	0.000
Survey						
year T1	-0.525***	-0.501***	-0.530***	-0.502***	-0.473***	-0.436***
	(0.0771)	(0.0765)	(0.0773)	(0.0765)	(0.0678)	(0.0690)
Observations	2682	2853	2682	2853	2682	2853

Note: This table presents model characteristics of the different selection model estimations, including the coefficients of the λs from the Heckman estimations and the θs from the Copula estimations, the log-likelihoods (LL) from the MLEs and the p-value of the test of no independence between u_{1i} and u_{2i} from the MLE. In addition, the coefficient on the exclusion restriction, the survey year at T1, is presented. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Figures

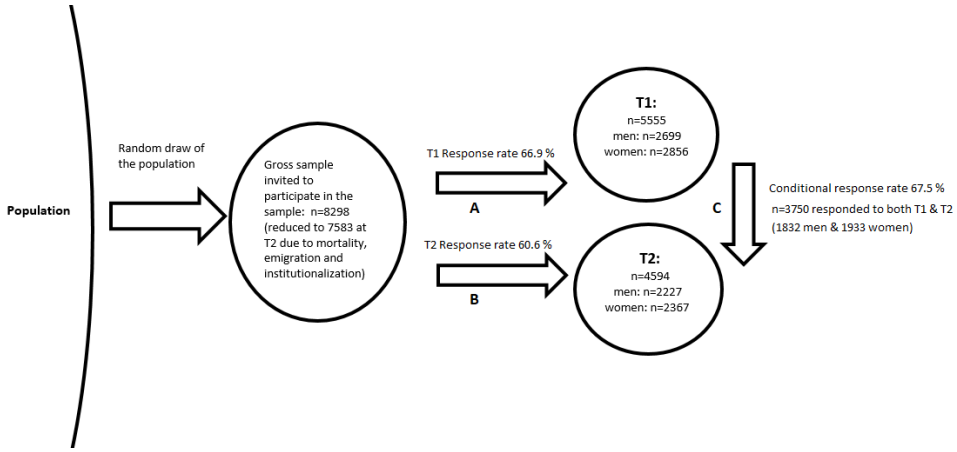


Figure 1: Response Patterns in the NorLAG Survey Sample

Note: This figure displays the sample sizes and response rates in the two waves of the NorLAG survey. 265 individuals (4.8 percent) of the responders at T1 died between T1 and T2.

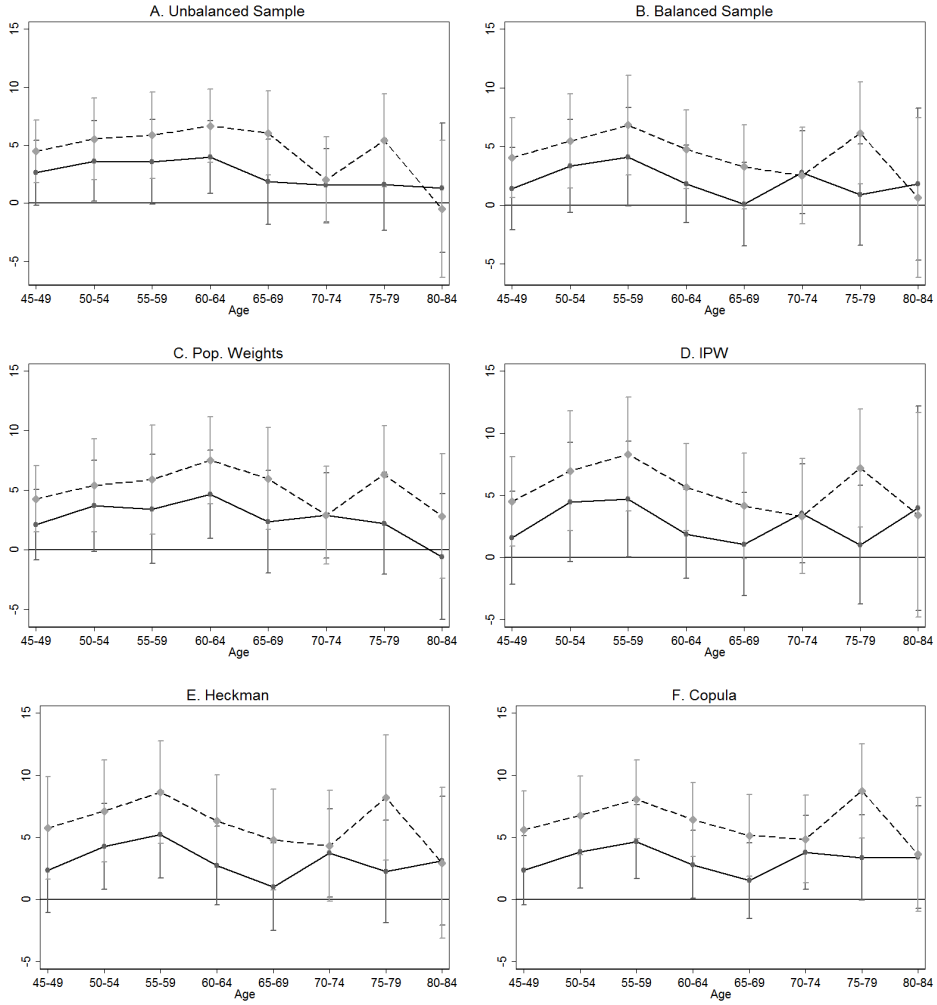


Figure 2: Marginal Health Effects by Education and Age for Men

Note: Marginal effects of education on health (SF12) for upper secondary and higher education compared to lower education for men. The dotted line represents higher education, and the solid line represents upper secondary education. Lower secondary education is the reference category. The effect is estimated for each age group, 45–49, ..., 80–84. The brackets indicate 95 percent confidence intervals. All estimates are based on Equation 2, and the reported coefficients are as derived in 3. Graph A displays OLS estimates based on the unbalanced sample of responders at T2. Graph B displays OLS estimates of the balanced sample at T2. Results in graph C are estimated by WLS using population weights on the unbalanced sample at T2, and graph D presents WLS estimates using the IPWs on the balanced sample at T2. Graph E displays the estimates based on the Heckman two-step, and graph F displays the Copula estimates using the Joe copula. For graph E and graph F, estimations are based on responders and non-responders at T2, provided they responded at T1.

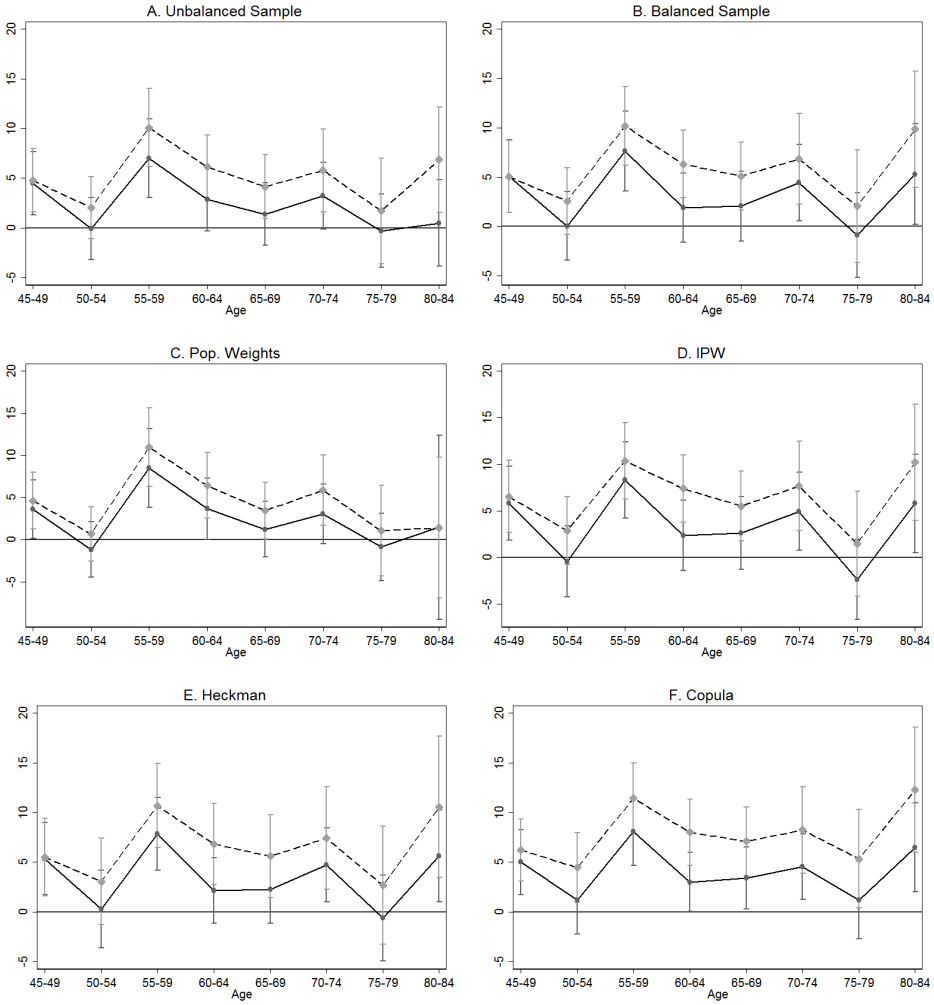


Figure 3: Marginal Health Effects by Education and Age for Women

Note: Marginal effects of education on health (SF12) for upper secondary and higher education compared to lower education for women. The dotted line represents higher education, and the solid line represents upper secondary education. Lower secondary education is the reference category. The effect is estimated for each age group, 45–49, ..., 80–84. The brackets indicate 95 percent confidence intervals. All estimates are based on Equation 2, and the reported coefficients are as derived in 3. Graph A displays OLS estimates based on the unbalanced sample of responders at T2. Graph B displays OLS estimates of the balanced sample at T2. Results in graph C are estimated by WLS using population weights on the unbalanced sample at T2, and graph D presents WLS estimates using the IPWs on the balanced sample at T2. Graph E displays the estimates based on the Heckman two-step, and graph F displays the Copula estimates using the Joe copula. For graph E and graph F, estimations are based on responders and non-responders at T2, provided they responded at T1.

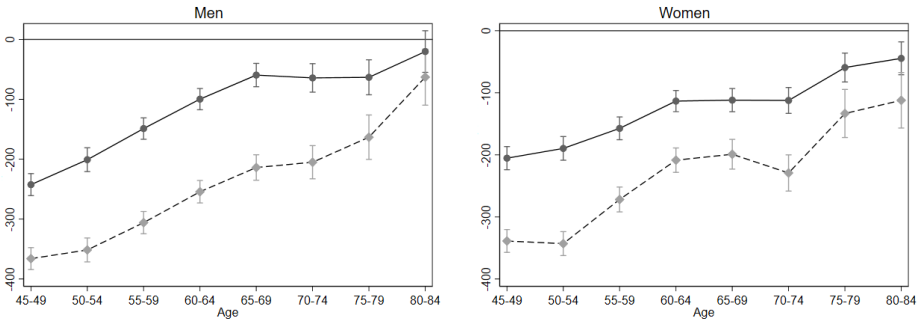


Figure 4: Marginal Health Effects by Education and Age based on Register Data

Note: Marginal effects of education on health (PCPR) for upper and higher education compared to lower education. The dotted line represents higher education, and the solid line represents upper secondary education. Lower secondary education is the reference category. The effect is estimated at each age group, 45–49, ..., 80–84. The brackets indicate 95 percent confidence intervals. All estimates are by OLS and are based on Equation 2, and the reported coefficients are as derived in 3. Data are the register data in 2007.

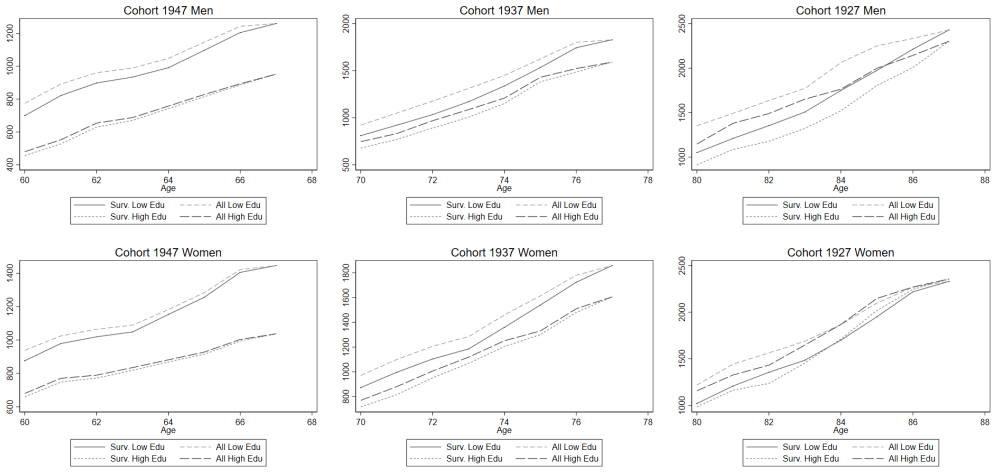


Figure 5: Mortality Selection: Healthy Survivors

Note: Average reimbursements across age for the high and lower secondary educated groups for cohorts born in 1947 (age 60–68), 1937 (age 70–78), and 1927 (age 80–88) for men and women. "Surv." indicates those surviving throughout the period (2007-2014). "All" indicates the gross sample alive at each age (each year 2007-2014).

Appendix

Table A1: Probit Estimates for Making the IPW

	Men		Women	
45-49	-0.012	(0.1029)	0.051	(0.1005)
50-54	-0.018	(0.1027)	0.063	(0.1007)
55-59	0.299***	(0.1144)	0.129	(0.1058)
60-64	0.180	(0.1233)	0.100	(0.1130)
65-69	-0.085	(0.1525)	0.048	(0.1485)
70-74	-0.495***	(0.1903)	-0.127	(0.1892)
75-79	-0.568***	(0.1953)	-0.470**	(0.1945)
80+	-0.782***	(0.2975)	-0.787***	(0.2863)
Upp. sec. educ	0.300***	(0.0660)	0.180***	(0.0622)
Higher educ	0.604***	(0.0863)	0.495***	(0.0842)
Inc Q2	0.011	(0.0895)	0.115*	(0.0685)
Inc Q3	0.096	(0.0900)	0.157*	(0.0819)
Inc Q4	0.196**	(0.0942)	0.245**	(0.1023)
Central	-0.032	(0.0593)	0.068	(0.0588)
Cohabitation	0.171***	(0.0630)	0.045	(0.0596)
Retired	-0.002	(0.1408)	0.019	(0.1420)
Working	0.089	(0.0867)	0.088	(0.0805)
SF12	-0.003	(0.0063)	-0.002	(0.0054)
Cronic	-0.130	(0.1196)	0.020	(0.1061)
SRH = 2	0.242**	(0.1192)	0.209*	(0.1093)
SRH = 3	0.398***	(0.1326)	0.092	(0.1237)
SRH = 4	0.501***	(0.1475)	0.272*	(0.1410)
SRH = 5	0.575***	(0.1608)	0.286*	(0.1535)
Physical lim.	0.092	(0.1157)	0.186*	(0.1015)
Moth alive	-0.181**	(0.0717)	0.034	(0.0673)
Inc drop	0.009	(0.0583)	-0.094*	(0.0567)
Inc SD	-0.000**	(0.0000)	-0.000	(0.0000)
Constant	-0.286	(0.3274)	-0.318	(0.2724)
Observations	2591		2736	
Pseudo R^2	0.104		0.070	

Note: Variables applied for making the IPW. The coefficients are the probit estimates of the likelihood of response in wave 2 based on observable characteristics in wave 1 (Equation 5). Standard errors in parentheses in column 2 and 4 (SE). * p<0.1, ** p<0.05, *** p<0.01.

Table A2: First Stage Heckman Selection Model

	Men	Women
50-54	0.026	0.035
	(0.0987)	(0.0968)
55-59	-0.011	-0.027
	(0.0981)	(0.0951)
60-64	0.240**	0.054
	(0.1025)	(0.0966)
65-69	0.113	-0.055
	(0.1045)	(0.0992)
70-74	-0.099	-0.045
	(0.1054)	(0.1065)
75-79	-0.469***	-0.289***
	(0.1049)	(0.1063)
80+	-0.743***	-0.758***
	(0.1112)	(0.1051)
Upp. sec. educ	0.390***	0.298***
	(0.0627)	(0.0586)
Higher educ	0.770***	0.679***
	(0.0761)	(0.0723)
Survey year T1	-0.525***	-0.501***
	(0.0771)	(0.0765)
Constant	0.218**	0.300***
	(0.0854)	(0.0808)
Observations	2668	2844
Pseudo R^2	0.091	0.079

Note: First stage of the Heckman two-step selection model. The coefficients are the probit estimates of the likelihood of response in wave 2 based on observable characteristics in wave 1. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A3: Marginal Health Effects of Education by Age for Men - Weighting Approach

	Unbalanced	Balanced	Pop. Weights	IPW
Upp. sec. educ				
45-49	2.618*	1.394	2.104	1.597
	(1.4165)	(1.7921)	(1.5134)	(1.9080)
50-54	3.626**	3.338*	3.682*	4.453*
	(1.7703)	(2.0156)	(1.9518)	(2.4559)
55-59	3.546*	4.113*	3.426	4.687**
	(1.8588)	(2.1430)	(2.3367)	(2.3686)
60-64	3.971**	1.809	4.650**	1.875
	(1.5984)	(1.6815)	(1.8976)	(1.8291)
65-69	1.848	0.100	2.357	1.066
	(1.8721)	(1.8144)	(2.2030)	(2.1103)
70-74	1.550	2.789	2.887	3.552*
	(1.6098)	(1.7943)	(1.8280)	(2.0254)
75-79	1.635	0.880	2.220	1.016
	(2.0167)	(2.2030)	(2.1847)	(2.4356)
80+	1.329	1.810	-0.581	3.961
	(2.8414)	(3.3025)	(2.6921)	(4.2011)
Higher educ				
45-49	4.472***	4.053**	4.271***	4.511**
	(1.3783)	(1.7299)	(1.4233)	(1.8365)
50-54	5.539***	5.460***	5.398***	6.959***
	(1.7997)	(2.0428)	(1.9843)	(2.4599)
55-59	5.853***	6.824***	5.887**	8.314***
	(1.8942)	(2.1658)	(2.3412)	(2.3439)
60-64	6.669***	4.770***	7.522***	5.652***
	(1.6195)	(1.7065)	(1.8665)	(1.7949)
65-69	6.051***	3.263*	5.978***	4.158*
	(1.8498)	(1.8348)	(2.1772)	(2.1696)
70-74	2.015	2.532	2.909	3.318
	(1.8987)	(2.0923)	(2.0937)	(2.3653)
75-79	5.419***	6.139***	6.326***	7.209***
	(2.0390)	(2.2092)	(2.0936)	(2.4251)
80+	-0.501	0.634	2.815	3.404
	(3.0080)	(3.4656)	(2.6618)	(4.1992)
Observations	2185	1799	2067	1753
Adjusted R^2	0.072	0.060	0.067	0.087
Mean	48.700	48.785	48.928	47.759

Note: Marginal effects of education on health for upper secondary and higher education compared to lower secondary education for men. The effect is estimated at each age-group 45-49, ..., 80-85. All estimates are based on Equation 2, and the reported coefficients are as derived in Equation 3. The first column (Unbalanced), presents OLS estimates based on the ubalanced T2 sample. The second column presents OLS estimates of the balanced T2 sample. The third column is estimated by WLS using the population weights on the unbalanced T2 sample, and the fourth column presents WLS estimates using the IPW and the balanced T2 sample. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A4: Marginal Health Effects of Education by Age for Women - Weighting Approach

	Unbalanced	Balanced	Pop. Weights	IPW
Upp. sec. educ				
45-49	4.483***	5.109***	3.588**	5.804***
	(1.6166)	(1.8792)	(1.7592)	(2.0346)
50-54	-0.064	0.032	-1.189	-0.445
	(1.5854)	(1.7758)	(1.6878)	(1.9415)
55-59	7.021***	7.618***	8.469***	8.314***
	(2.0190)	(2.0609)	(2.3839)	(2.0799)
60-64	2.880*	1.880	3.662**	2.338
	(1.6242)	(1.7918)	(1.8567)	(1.9250)
65-69	1.385	2.055	1.228	2.615
	(1.6052)	(1.8040)	(1.6912)	(1.9867)
70-74	3.191*	4.435**	3.042*	4.933**
	(1.7140)	(1.9823)	(1.8097)	(2.1304)
75-79	-0.304	-0.890	-0.866	-2.388
	(1.8856)	(2.1839)	(2.0412)	(2.1887)
80+	0.462	5.297**	1.453	5.781**
	(2.2184)	(2.6111)	(5.5799)	(2.7026)
Higher educ				
45-49	4.788***	5.039***	4.611***	6.513***
	(1.6137)	(1.8626)	(1.7214)	(1.9746)
50-54	2.022	2.581	0.677	2.884
	(1.5809)	(1.7237)	(1.6463)	(1.8617)
55-59	10.086***	10.200***	10.967***	10.353***
	(2.0090)	(2.0328)	(2.3959)	(2.0838)
60-64	6.138***	6.332***	6.424***	7.384***
	(1.6459)	(1.7481)	(1.9758)	(1.8351)
65-69	4.149**	5.112***	3.448**	5.519***
	(1.6479)	(1.7553)	(1.7112)	(1.9071)
70-74	5.792***	6.850***	5.866***	7.654***
	(2.1278)	(2.3375)	(2.1212)	(2.4411)
75-79	1.693	2.046	1.025	1.450
	(2.7107)	(2.9009)	(2.7464)	(2.8624)
80+	6.851**	9.856***	1.398	10.210***
	(2.7084)	(3.0085)	(4.2583)	(3.1840)
Observations	2311	1894	2180	1837
Adjusted R^2	0.079	0.085	0.058	0.106
Mean	46.500	46.766	46.716	45.872

Note: Marginal effects of education on health for upper secondary and higher education compared to lower secondary education for women. The effect is estimated at each age-group 45-49, ..., 80-85. All estimates are based on Equation 2, and the reported coefficients are as derived in Equation 3. The first column (Unbalanced), presents OLS estimates based on the ubalanced T2 sample. The second column presents OLS estimates of the balanced T2 sample. The third column is estimated by WLS using the population weights on the unbalanced T2 sample, and the fourth column presents WLS estimates using the IPW and the balanced T2 sample

Table A5: Marginal Health Effects of Education by Age - Selection Model

	Men			Women		
	Heckman	MLE	Copula	Heckman	MLE	Copula
Upp. sec. educ						
45-49	2.359	1.555	2.343*	5.338***	5.159***	5.009***
	(1.7543)	(1.6610)	(1.4209)	(1.8597)	(1.8025)	(1.6726)
50-54	4.247**	3.490**	3.821**	0.280	0.086	1.162
	(1.7660)	(1.6846)	(1.4868)	(1.9804)	(1.9175)	(1.7503)
55-59	5.211***	4.297**	4.638***	7.846***	7.667***	8.061***
	(1.7904)	(1.6720)	(1.5221)	(1.8625)	(1.8059)	(1.7232)
60-64	2.721*	1.962	2.800**	2.125	1.933	3.002**
	(1.6202)	(1.5206)	(1.3962)	(1.6759)	(1.6020)	(1.5125)
65-69	1.019	0.254	1.505	2.290	2.106	3.401**
	(1.8010)	(1.7164)	(1.5668)	(1.7524)	(1.6881)	(1.5903)
70-74	3.736**	2.947*	3.788**	4.725**	4.498**	4.542***
	(1.8086)	(1.7252)	(1.5178)	(1.9077)	(1.8176)	(1.6900)
75-79	2.238	1.107	3.360*	-0.596	-0.827	1.160
	(2.1050)	(1.9648)	(1.7544)	(2.2064)	(2.1281)	(1.9878)
80+	3.118	2.028	3.412	5.608**	5.364**	6.472***
	(2.6542)	(2.5866)	(2.1049)	(2.3661)	(2.2861)	(2.2821)
Higher educ						
45-49	5.747***	4.337**	5.609***	5.516***	5.142***	6.228***
	(2.1148)	(1.8435)	(1.6027)	(1.9921)	(1.7467)	(1.5979)
50-54	7.115***	5.736***	6.767***	3.069	2.686	4.458**
	(2.0891)	(1.8270)	(1.6177)	(2.2184)	(1.9904)	(1.7989)
55-59	8.638***	7.128***	8.065***	10.684***	10.304***	11.441***
	(2.1057)	(1.7945)	(1.6229)	(2.1475)	(1.9157)	(1.7984)
60-64	6.309***	5.027***	6.429***	6.831***	6.440***	7.997***
	(1.9054)	(1.6474)	(1.5107)	(2.0883)	(1.8320)	(1.7076)
65-69	4.806**	3.521*	5.148***	5.601***	5.217***	7.106***
	(2.0687)	(1.8364)	(1.6715)	(2.1240)	(1.8831)	(1.7572)
70-74	4.307*	2.828	4.858***	7.417***	6.972***	8.233***
	(2.2810)	(2.0063)	(1.7997)	(2.6330)	(2.3733)	(2.2133)
75-79	8.206***	6.485***	8.744***	2.672	2.181	5.344**
	(2.5677)	(2.2551)	(1.9375)	(3.0422)	(2.7715)	(2.5301)
80+	2.934	1.019	3.623	10.557***	10.007***	12.280***
	(3.1017)	(2.8109)	(2.3428)	(3.6299)	(3.3492)	(3.2048)
Observations	2682	2682	2682	2853	2853	2853

Note: Marginal effects of education on health for upper secondary and higher education compared to lower secondary education using the Heckman selection model and the Copula selection model. The effect is estimated at each age-group 45-49, ..., 80-85. Estimation is based on responders and non-responders at T2, provided information about gender, age, and education were obtained at T1. The column "Heckman" presents results estimated by the *heckman two-step*, the column "MLE" presents results of the heckamn model estimated by maximum likelihood. The column "Coupla" presents results estimated by the *heckmancopula* with copula *Joe* in Stata. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A6: The Vuoung Test

	OLS Test	
	Men	Women
Constant	-0.176***	-0.128***
	(-16.67)	(-14.07)
Observations	2,682	2,853

Note: The Vuong test of the likelihood contributions of each observation in the *heckman* MLE estimation compared to the *heckmancopula* with copula Joe. Estimates are the constant term in an OLS regression of the difference between the contribution of each observations to the log likelihood function in the Heckman MLE compared to the Copula MLE.

Table A7: Marginal Health Effects of Education by Age - Register Data

	Men 2007	Women 2007
Upp. sec. educ		
45-49	-242.4***	-205.3***
	(-25.85)	(-21.54)
50-54	-200.8***	-189.6***
	(-19.79)	(-19.45)
55-59	-148.8***	-157.3***
	(-16.33)	(-16.76)
60-64	-99.67***	-113.3***
	(-11.01)	(-13.02)
65-69	-59.47***	-111.8***
	(-5.97)	(-11.73)
70-74	-64.23***	-112.3***
	(-5.30)	(-10.56)
75-79	-63.24***	-59.24***
	(-4.25)	(-4.99)
80+	-20.22	-44.49**
	(-1.14)	(-3.29)
Higher educ		
45-49	-366.0***	-338.8***
	(-39.13)	(-35.99)
50-54	-351.5***	-343.0***
	(-34.49)	(-34.88)
55-59	-305.9***	-271.9***
	(-32.25)	(-26.67)
60-64	-254.3***	-208.5***
	(-26.58)	(-20.78)
65-69	-213.8***	-199.0***
	(-19.75)	(-16.26)
70-74	-205.0***	-229.2***
	(-14.50)	(-15.39)
75-79	-163.2***	-133.3***
	(-8.64)	(-6.76)
80+	-62.91**	-112.0***
	(-2.63)	(-4.91)
Observations	816178	865759
Adjusted R^2	0.0473	0.0245
Mean	671.7	831.1

Note: Marginal effects of education on annual PCPR for upper secondary and higher education compared to lower education for men and women. The effect is estimated at each age-group 45-49, ..., 80-85. Estimates are based on OLS estimation of Equation 2 using the register data in 2007, and the reported coefficients are the marginal effects as derived in 3. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

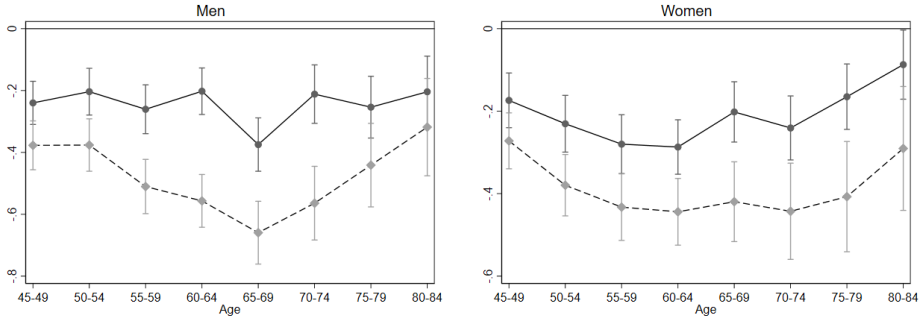


Figure A1: Marginal Effects of Education on days of Hospitalization across Age

Note: Marginal effects of education across age on the likelihood of having a hospital admission. The dotted line represents higher education, and the solid line represents upper secondary education. Estimation is by a Tobit regression using the register data in 2008 on a binary variable based on having a record of a hospitalization in the current year. Records of hospitalizations are not available for 2007. The brackets indicate 95 percent confidence intervals.

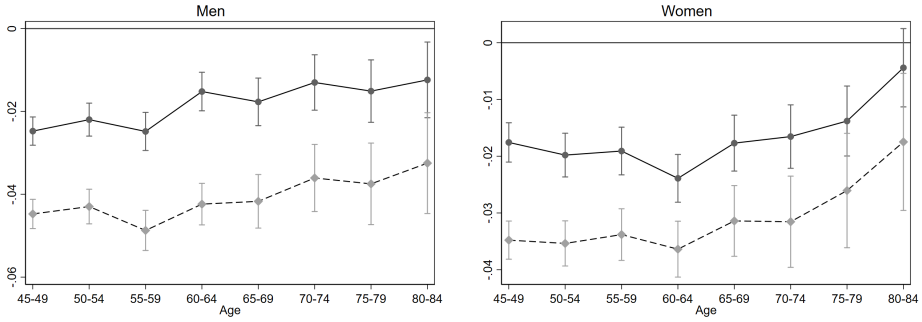


Figure A2: Marginal Health Effects of Education w/o Institutionalized

Note: Marginal effects of education on Reimbursements across age without the ad hoc reimbursements for the institutionalized. The dotted line represents higher education, and the solid line represents upper secondary education. Estimation based on the register data in 2007. The brackets indicate 95 percent confidence intervals.

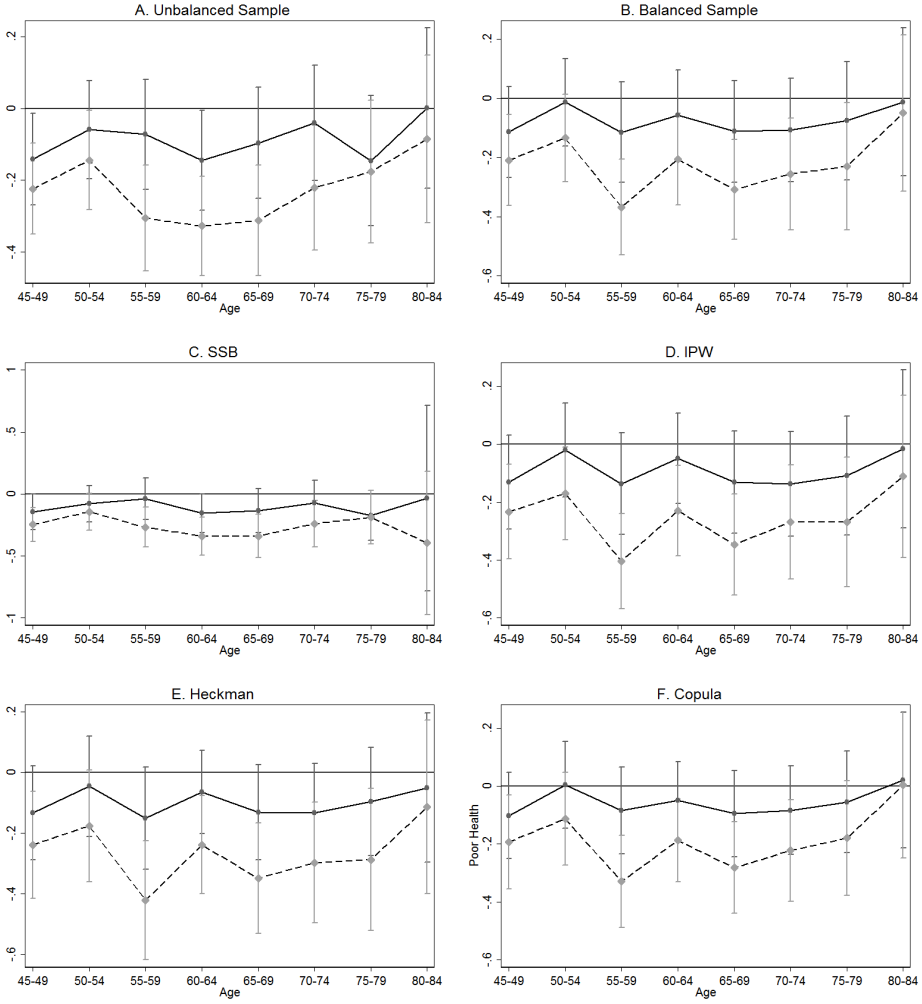


Figure A3: Poor Health: Marginal Health Effects by Education and Age for Men

Note: Marginal effects of education on the likelihood of having poor health for upper secondary and higher education compared to lower education for men. The dotted line represents higher education, and the solid line represents upper secondary education. Lower secondary education is the reference category. The effect is estimated for each age group, 45–49, ..., 80–84. The brackets indicate 95 percent confidence intervals. All estimates are based on Equation 2, and the reported coefficients are as derived in 3. Graph A displays OLS estimates based on the unbalanced sample of responders at T2. Graph B displays OLS estimates of the balanced sample at T2. Results in graph C are estimated by WLS using population weights on the unbalanced sample at T2, and graph D presents WLS estimates using the IPWs on the balanced sample at T2. Graph E displays the estimates based on the Heckman two-step, and graph F displays the Copula estimates using the Joe copula. For graph E and graph F, estimations are based on responders and non-responders at T2, provided they responded at T1.

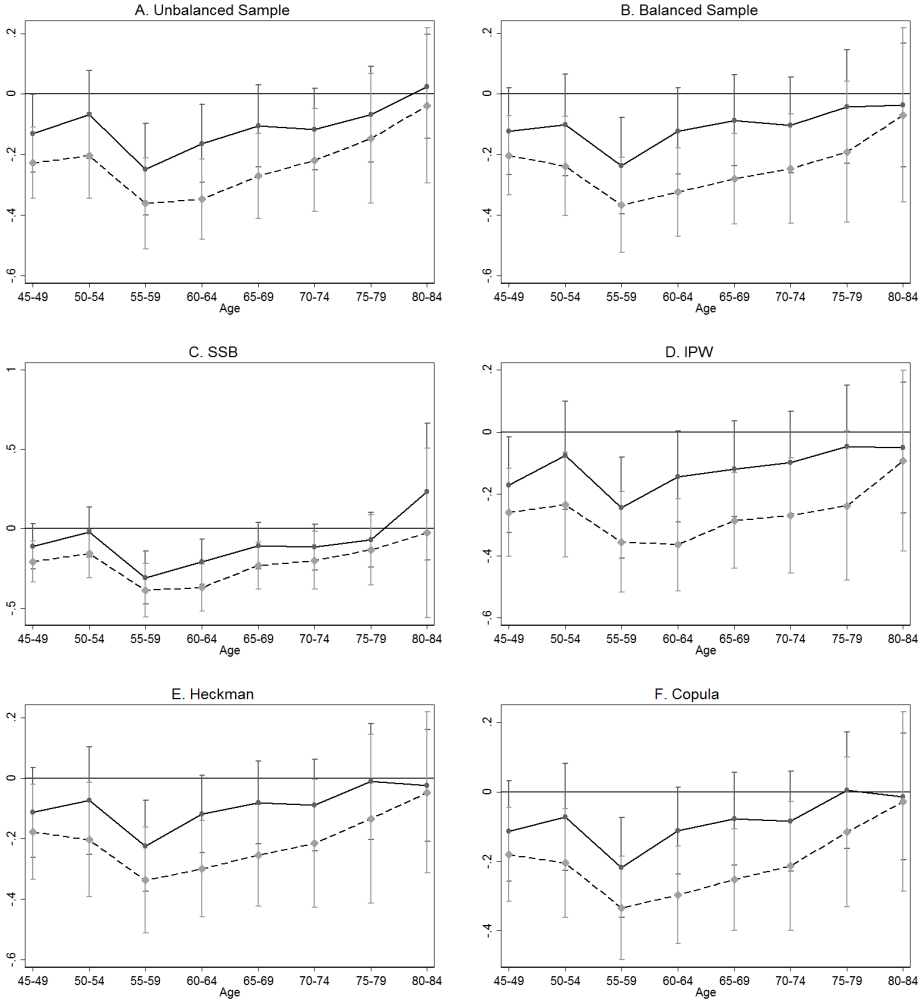


Figure A4: Poor Health: Marginal Health Effects by Education and Age for Women

Note: Marginal effects of education on the likelihood of having poor health for upper secondary and higher education compared to lower education for men. The dotted line represents higher education, and the solid line represents upper secondary education. Lower secondary education is the reference category. The effect is estimated for each age group, 45–49, ..., 80–84. The brackets indicate 95 percent confidence intervals. All estimates are based on Equation 2, and the reported coefficients are as derived in 3. Graph A displays OLS estimates based on the unbalanced sample of responders at T2. Graph B displays OLS estimates of the balanced sample at T2. Results in graph C are estimated by WLS using population weights on the unbalanced sample at T2, and graph D presents WLS estimates using the IPWs on the balanced sample at T2. Graph E displays the estimates based on the Heckman two-step, and graph F displays the Copula estimates using the Joe copula. For graph E and graph F, estimations are based on responders and non-responders at T2, provided they responded at T1.

Grafisk design: Kommunikasjonsavdelingen, UiB / Trykk: Skipnes Kommunikasjon AS

uib.no

ISBN: 978-82-308-3010-9