Understanding Socioeconomic Differences in Health

An Economic Approach

ISBN: 978 90 3610 175 2

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **469** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Understanding Socioeconomic Differences in Health

An Economic Approach

Sociaaleconomische verschillen in gezondheid

Een economische analyse

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 29 april 2010 om 13:30 uur.

door
Johannes Laurentius Wilhelmus van Kippersluis

geboren te IJsselstein

Promotiecommissie

Promotoren:	Prof.dr. E.K.A. van Doorslaer
	Dr. O.A. O'Donnell
Overige leden:	Prof.dr. B. Jacobs
	Prof.dr. M. Lindeboom
	Prof.dr. D.J.C. van Dijk

Acknowledgements

When starting my PhD track some $3\frac{1}{2}$ years ago I did not really know what to expect. If you read books or talk to people about pursuing a PhD one persistent stereotype will always follow: "it is a solitary track". Looking back, I can say that the greatest risk I have incurred during the last couple of years was too much sociability rather than solitude! To a large extent this is due to my extremely nice fellow PhD-students and colleagues, and the atmosphere at the Applied Economics department.

In particular I would like to mention Esen, my officemate during the whole period. You were truly the best officemate I could wish, and I would like to thank you for the nice times we had in- but also outside the walls of the university in Marbella, Beijing, Hong Kong, Bergen, IJsselstein, and Rotterdam! At this point, I would also like to thank all other (former) colleagues of the Health/Behavioural Economics group and broader Applied Economics department at the 13th floor: Teresa, Pilar, Bas, Martijn, Aurelien, Kirsten, Arthur, Ellen, Lambert, Italo, Frank, Justus, Han, Anka, Nita, Gerda, Eddy, Enrico, Ronald, Harry, and Tom.

Then I would also like to particularly thank my fellow PhD students who all remained or became good fiends: Arco for having me as his best man for his wedding, the bachelor party, Carnival, and Bulgaria, to name a few. Peter for his cheerful presence everyday, for generously scanning through my whole PhD thesis, our trip to London, and our skiing trips. I would like to thank Kar Yin for the nice tea (or latte) breaks, and our combined squash-pancake events, Eelco for our trip to London and the nice skiing holiday, Yuri for organizing the cocktail party and bowling events, and Mathijn and Sjoerd for the nice lunches inside and other events outside the Erasmus university.

A second stereotype, which I cannot invalidate, is that anyone writing a PhD thesis will face difficulties at some point and goes through unproductive and inefficient periods. Obviously, I went through such periods as well, but due to my outstanding team of supervisors and co-authors, these periods were few and short. I owe most to my supervisors Eddy and Owen, who are truly the best team I can think of in supervising a PhD thesis.

Apart from your great expertise and international network from which I both benefited a lot, I appreciated even more your kindness, accessibility, availability, confidence, and encouragement during the last couple of years. Finally, I greatly enjoyed our karaoke- and ski adventures in Beijing and Crans Montana!

I would also like to thank my co-authors Titus, Tom and Pilar. Titus thanks a lot for your kindness, hospitality, and patience. I learned a lot from our discussions and it is a big pleasure to work together. I can say now that my visit to RAND, together with my teaching experiences, turned the scale of staying in academia. In that sense, you have successfully fulfilled Eddy's request of persuading me! Tom thanks for your kindness, and great accessibility, especially in the first part of my PhD track. I learned a great deal from you and appreciate the rigor with which you employ things! Pilar, thanks for our nice cooperation during the last part of my track. I mostly appreciated your endless stream of ideas and the way in which you can bring complicated matters back to the (much easier) essence!

Then I would like to thank Dick van Dijk and Christiaan Heij for triggering my interest in pursuing a PhD, and bringing me into contact with Eddy, Arie Kapteyn for facilitating my stay at RAND, and Rene Segers for generously sharing his LaTeX code. Also, I would like to thank all my relatives and friends from study, soccer, and former high school, for their interest in my work and for the nice times we had the last couple of years.

Finally, a big word of thanks goes out to my family and Vivian. Vivian for her encouragement, trust, and for being extremely patient and sweet. Peter for his encouragement to start a PhD track, and critically reading the introduction and conclusion, and finally my parents for their encouragement, flexibility, and support whenever necessary.

By and large, it was a very good decision to enter into a PhD track, and I am especially grateful for the possibility of pursuing my PhD in the field of Health Economics at the Erasmus University. I learned a great deal about (health) economics, enjoyed the research part, and the first acquaintance with teaching was very pleasant. I look forward to a new phase at the Erasmus University and would like to thank, again, Eddy, Owen and Justus for allowing me this opportunity.

Hans van Kippersluis
Rotterdam, February 2010

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"Health is (. . .) without doubt the first good and the foundation of all the other goods of this life" as was recognized by René Descartes already in 1637. Health has both intrinsic and instrumental value, while it is central to our well-being and a prerequisite to the functioning as an agent (Anand, 2002; Sen, 2002). In other words, health is a universal human aspiration and a basic human need (Marmot, 2007). The right to the highest attainable level of health is enshrined in the charter of the World Health Organization (WHO) and many international treaties (e.g. United Nations, 2000). Despite this right, a majority of people in the world do not enjoy the good health that is biologically possible (CSDH, 2008).

Most notably, there is a strong association between health and socioeconomic status (SES), with the socially and economically disadvantaged enjoying worse health; irrespective of whether it is measured by morbidity, disability or mortality (see Antonovsky, 1967; Adler *et al.*, 1994; Smith, 1999; Marmot, 1999; Cutler *et al.*, 2008 for overviews). Inequalities are substantial and striking. For example, Case and Deaton (2005) show that in the US at age 20, men in the bottom quartile of household income already report worse health, on average, than do men in the top quartile at age 50. In Glasgow, UK, life expectancy of men in the most deprived areas was 54 years, compared with 82 years in the most affluent (Hanlon *et al.*, 2006). Even in an egalitarian country such as the Netherlands striking differences in health and longevity by socioeconomic status exist. Individuals with a university degree live, on average, 6 to 7 years longer than those who finished only primary school, while the difference in life expectancy in good health can be as much as 16 to 19 years (CBS, 2008).

Inequalities in, for example, income are to some extent tolerated for economic reasons since income incentives are inevitable to stimulate effort, skill, and enterprise. Inequality in income has the potential to increase the size of total income from which, in principle,

the society as a whole can gain through taxation (Anand, 2002). Also, income clearly can be more easily redistributed than health. Health, in contrast, is a special good, a basic necessity of life, and as such inequalities in health constitute inequalities in people's capability to function. This is considered as a denial of equality of opportunity (Rawls, 1971; Anand, 2002). Recently, the Commission on Social Determinants of Health of the World Health Organization (WHO) vociferously defended the view that inequalities in health by social and economic status are to a large extent avoidable by reasonable action, such that they can be labeled as inequities (CSDH, 2008). The lack of opportunity to achieve good health because of inadequate social arrangements is considered a serious social injustice (Sen, 1985; 2002), and it is an ethical imperative to tackle inequities in health (CSDH, 2008).

Apart from the argument of social justice, addressing the dependence of health on socioeconomic position, and thereby improving average population health might have economic motives too. First, there could be substantial savings on age-specific health care expenditures. Better health is also able to increase labor supply, productivity, educational attainment, and savings and investment (Bloom and Canning, 2000). Historically, health has been the major contributor — even above educational attainment — of economic growth, also in developed countries (Barro, 1997; Suhrcke *et al.*, 2006). Finally, an unequal society in which large health inequalities exist has less social cohesion, less solidarity and more stress, which is potentially detrimental to economic development (Deaton, 2003). In sum, breaking the connection between low education, poor health and early labor force exit could raise national wealth. Due to these strong social and economic motives, it is no surprise that the primary health-related goal of current governments all over the world is the reduction of health inequalities.

To this aim, it is vital to understand why socioeconomic differences in health exist, but this is generally not easy. Both health and socioeconomic status are multidimensional and dynamic, and the relationship between them may reflect different effects and feedbacks over the course of life. Does low economic status in early life lead to the development of health problems in middle and later life? Or, does poor health interfere with the acquisition of education and, subsequently, the chances of securing, or holding onto, a well-paid job? Over the years progress has been made in characterizing the relationships between the various dimensions of socioeconomic status and health over the life cycle and in understanding the relative importance and directions of causal pathways. Before 1985 scholars thought that poverty and access to medical care were the main drivers of inequalities in health (Adler and Ostrove, 1999). Above a certain poverty line, income and other socioeconomic variables were not expected to have any influence on health.

This threshold model was challenged most forcefully by the seminal Whitehall studies (Marmot *et al.* 1984; 1991), in which even among British civil servants, all with a salary way beyond the poverty line, socioeconomic status as measured by occupational hierarchy had a negative impact on mortality. Since then the relationship between health and socioeconomic status has been labeled a gradient, since mortality and ill-health continue to drop as one goes up the SES ladder (Adler *et al.*, 1994; Deaton, 2002a).

In recent years, epidemiological and sociological research has used longitudinal studies to examine the role of behavioral, material, psychosocial and healthcare related pathways in explaining SES-health associations (Marmot *et al.*, 1997b; Van Lenthe *et al.*, 2002; Lynch, 2003; van Oort *et al.*, 2005; Skalicka *et al.*, 2009). The focus of these studies has been mostly on the effect of SES on health, while economists have recently emphasized the importance of the reverse impact of health on SES through ability to work (Smith, 1999; 2005a; 2007; Case and Deaton, 2005). Obviously, it could be that there are so-called third factors influencing both socioeconomic status and health, triggering a spurious correlation between them. Potential confounders are for example time preference (Fuchs, 1982; Barsky *et al.*, 1997), and intelligence and cognitive ability (Auld and Sidhu, 2005; Deary, 2008). Although the importance of third factors is non-negligible, the direct causal pathways from SES to health and vice versa seem more important (Elo and Preston, 1996; Cutler and Lleras-Muney, 2008). Two causal pathways are beginning to emerge as particularly important in explaining the SES-health gradient. First, there is a non-negligible life-long effect of education on health (Oreopoulos, 2006; Smith, 2007; Silles, 2009). Second, there is a large effect of health on income that operates through employment (Smith, 1999; 2005a; Case and Deaton, 2005). In contrast, although large health differences across income groups are visible, economists argue that these do not derive from a causal impact of income on health (Adams *et al.*, 2003; Contoyannis *et al.*, 2004; Frijters *et al.*, 2005; Smith, 2007; Cutler *et al.*, 2008), although some epidemiological scholars disagree with this tentative conclusion (e.g. Marmot, 2002; Herd *et al.*, 2009).

Despite the wealth of literature available and improved knowledge on socioeconomic differences in health, very little robust causal evidence is available. For example, the effect of education on health outcomes has been found in a couple of studies (Oreopoulos, 2006; Silles, 2009) but was found insignificant in numerous other studies (Clark and Royer, 2007; Mazumder, 2008; Albouy and Lequien, 2008). Moreover, the underlying mechanisms responsible for these potential pathways are largely unclear. For example, even if education is found to be protective for health, it is not known why the more educated are healthier. Without knowledge of the mechanisms, it is difficult to design policies that are effective in reducing disparities (Deaton, 2002a). In sum, two important

gaps remain in the literature: (i) robust causal evidence on the most important pathways, and (ii) knowledge of the mechanisms that underlie these causal pathways.

The aims of this thesis are to help fill these two gaps in the literature by taking an economic approach. We will exploit econometric techniques to estimate the causal effects, while we will use and develop an economic model to help understand the underlying mechanisms. The contribution of this thesis is twofold. First, exogenous variation in SES and health is exploited to estimate robust causal evidence of the pathways anticipated to be most important in generating SES differences in health: (i) the effect of education on health and mortality, and (ii) the effect of health on employment and income. By exploiting quasi-natural experiments we are able to control for possible third factors influencing both socioeconomic status and health. Second contribution is a theoretical framework of health and socioeconomic status over the life cycle that captures many important mechanisms that potentially create health disparities across socioeconomic groups. This helps us to understand the underlying mechanisms responsible for the causal pathways indicated above.

The first part of this thesis — chapters 2 and 3 — provides a descriptive analysis how health evolves over the life cycle and whether the pattern is different across socioeconomic groups. Cutler *et al.* (2008) noted that *"(. . .) differential patterns of causality make a single theory of socioeconomic gradients in health difficult to imagine. We suspect, though, that the right theory will emphasize the lifecycle."* Inspired by this idea, chapter 2 aims to give an idea how health and health inequality evolve over the life cycle, correcting for possible cohort effects deriving from differences between subsequent generations. These differences between generations could for instance derive from changes in health systems, medical technology and broader social and economic developments. Cohort effects are not just a nuisance distorting the life cycle picture but are important for predicting the consequences of population ageing and are of intrinsic interest. For example, cohort effects show whether average health and inequality in health have increased or decreased over generations. It turns out that in the Northern-European countries average self-reported health has been stable over generations, such that cohort effects do not distort the life cycle profile of average health in these countries. Overall inequality has decreased over generations, yet socioeconomic inequality remained stable. Apparently, despite efforts to improve population health and reduce socioeconomic inequalities in health, over the last generations this was not particularly successful. The remaining chapters of this thesis try to understand the nature of socioeconomic differences in health, such that policy can be more intelligently and successfully conducted.

Chapter 3 reviews the economic, sociological and epidemiological literature on socioeconomic inequalities in health over the life cycle and presents a descriptive analysis of how health evolves over the life cycle for different socioeconomic groups in the Netherlands. The latter allows identification of the timing of events in health and socioeconomic status, and helps gauge the extent to which data are consistent with various causal pathways. The analysis serves as a precursor to the causal analyses later in this thesis. The results reveal that the life cycle profile of socioeconomic differences in health in the Netherlands is very similar to that in the US. This suggests that fundamental mechanisms linking age, health, income, and education are relatively unresponsive to policy parameters. On the basis of this analysis, and the related literature, two mechanisms seem particularly important in generating the large association between health and socioeconomic status. First, there appears to be a large effect of health on income that operates through employment. Second, there seems to be a non-negligible life-long effect of education on health.

The second part of this thesis — chapters 4 and 5 — makes use of quasi-natural experiments to investigate whether this descriptive evidence derives from genuine, causal effects. Chapter 4 exploits the change in the Dutch compulsory schooling law in 1928, which raised the compulsory years of education from 6 to 7, to provide exogenous variation in education that allows identification of a causal effect of schooling on mortality. By comparing individuals born in 1916, which were allowed to drop out of school after 6 years of schooling, with individuals born in 1917, who were the first obliged to stay in school for the full 7 years, in a Regression Discontinuity Design (RDD) we are able to estimate the causal effect of an extra year of education on the probability of dying between ages 81 and 89.

Chapter 5 exploits acute hospitalizations as exogenous variation in health to estimate the causal effect of ill-health on employment and disposable income. Since shocks to one's health are not randomly distributed across the population, we employ a propensity score matching approach combined with differences-in-differences to correct for the non-experimental nature of the data. It is perhaps not surprising that health impacts on employment and, given that replacement rates of disability insurance are smaller than 100 percent, personal earnings. It is more interesting to investigate the effects on disposable income after behavioral responses by the individual that experienced the health shock and his/her other household members. In other words, chapter 5 aims to investigate the formal and informal mechanisms of income protection after a health shock. If health shocks reduce disposable income permanently, this potentially has serious consequences for household welfare, through reduced consumption (Attanasio and Davis, 1996; Deaton, 2002b; Blundell *et al.*, 2008).

For the policy purpose of reducing socioeconomic health inequalities it is extremely interesting to analyse whose employment and disposable income are least protected against health deterioration. The protection may for example differ by education since highly educated professionals have more human capital that is not contingent on physical health. Lower educated, unskilled workers rely more on their physical health. Should they lose this, their relative earnings capacity will fall by more. More highly educated individuals may also be better at managing their disease, for example, through adherence to medication (Goldman and Smith, 2002), such that it impinges less on their earning capacity. Differences by employment sector may also be anticipated. In the public sector, where there is less exposure to competition and consequently salaries are more regulated and less tied to productivity, employment and earnings should be less responsive to ill-health. The employer, rather than the disability insurance scheme, protects against the economic consequences of ill-health.

The quasi-natural experiments exploited in chapters 4 and 5 confirm the descriptive analysis that there is a strong effect of (i) education on mortality, and (ii) health on income through constrained employment. Regarding (i), even among an elderly population of over 80 we find a significant protective effect of education on mortality. The probability of dying between ages 81 and 89 is reduced by 2 to 3 percentage points, which translates into 4-6 percent relative to the baseline probability. We can conclude that at least part of the strong association between education and health outcomes derives from a causal effect of years of schooling on mortality. There is obviously pre-sample selective mortality and our estimates cannot be taken as indicative of the mortality effect of education at all ages. However, if education has a non-negative effect on mortality at all ages, a relatively larger group of lower educated individuals will have passed away before the age of 80, compared to their higher educated peers. In that case our estimates constitute a lower bound on the total effect of education on life expectancy.

Regarding (ii), a significant effect of ill-health on employment and disposable income is confirmed. Disposable income on average drops by 8 percent three years after a sudden deterioration of health, which represents a 92% replacement rate. Two other important considerations follow from the study. First, while Disability Insurance (DI) is developed to insure personal income, it cannot prevent disposable household income to drop. Striking is that household income drops by a larger amount than personal income, which indicates spill-over effects to other household members, possibly reflecting a fall in the earnings of the spouse as a result of meeting the care needs of the disabled person. Second consideration is that the drop in disposable income is not evenly spread over society. Groups in the lowest socioeconomic classes — those in the bottom quartile of personal income, and

the lower educated — are the most vulnerable to lose their jobs. At the same time this is translated into higher relative drops in personal and household disposable income. The average actual replacement rate of 80% faced by individuals in the lower income quartile compared to the average of 92%, does probably imply that the experience of a health shock by the poor would not only raise the observed income inequalities, but also put a non-negligible proportion of individuals in this group at risk of poverty.

Despite these advances in knowledge of the main causal pathways determining the SES-health gradient, little is known about the mechanisms that produce them. Without a better understanding of the mechanisms that lead to the observed health disparities, identification of policies that are effective in reducing them is typically hard. Case and Deaton (2005) argue that it is extremely difficult to understand the relationships between health, education, income and labor force status without some guiding theoretical framework. The standard economic model of the demand for health (Grossman, 1972a; 1972b) has proved extremely useful in providing a framework for the demand for health and medical care but cannot explain a number of the most salient features of the SES health gradient. It is no surprise then that recent reviews of the literature (e.g., Cutler *et al.*, 2008) point to the absence of a unifying theory of SES and health over the life cycle and emphasize the importance of developing one.

The last part of this thesis — Chapter 6 — aims to help fill this gap. Our starting point is the Grossman model (Grossman, 1972a; 1972b) and the extended version presented by Case and Deaton (2005). Our contribution is as follows. First, we address three significant problems identified in the literature with the Grossman model by introducing decreasing-returns-to-scale (DRTS) in the health production function (as in Ehrlich and Chuma, 1990). This addresses (i) the indeterminacy problem for investment in medical care, (ii) the inability of the model to reproduce the observed negative relation between health and the demand for medical care (e.g., Zweifel and Breyer, 1997), and (iii) the model's lack of capacity to explain differences in the effective rate of health deterioration (not just the level of health) between different socioeconomic groups (Case and Deaton, 2005). Yet, utilization of medical services and access to care explain only part of the association between SES and health (e.g., Adler *et al.*, 1993). Our second contribution is therefore to incorporate many potential mechanisms in the model that could explain disparities in health by SES and to include a multitude of potential bi-directional pathways between health and dimensions of SES. One important concept in our work is "job-related health stress", which can be interpreted broadly and can range from physical working conditions (e.g., hard labor) to the psychosocial aspects of work (e.g., low status, limited control, repetitive work, etc). The notion here is that job-related health stress can include any

aspect of work that is detrimental to health and as such is associated with a wage premium (a compensating wage differential). Other important features of the model are lifestyle factors such as preventive care, and healthy and unhealthy consumption. The model integrates a life cycle approach, and the concepts of financial, human and health capital (Muurinen and Le Grand, 1985). The focus is on understanding the SES-health gradient as the outcome of rational (constrained) individual behaviour, and the framework applies to individuals who have completed their education and participate (or have participated) in the labor-force.

We find that greater initial wealth, permanently higher income (over the life cycle) and a higher level of education induce individuals to invest more in curative and in preventive care, shift consumption toward healthy consumption, and enable individuals to afford healthier working environments (associated with lower levels of physical and psychosocial health stresses) and living environments. The mechanism through which initial wealth, permanent income and education operate is by increasing the marginal cost of, and demand for, curative care. The greater marginal cost of curative care in turn increases the health benefit of (and hence demand for) preventive care and healthy consumption, and the health cost of (and hence reduced demand for) unhealthy working and living environments, and unhealthy consumption. Jointly these gradually lead to cumulative health advantage with age. Our model thus holds considerable promise in explaining empirical health patterns. Such a model has not been available before and economists have highlighted the significance of its development (e.g., Case and Deaton, 2005; Cutler *et al.*, 2008).

Part I

Socioeconomic Differences in Health over the Life Cycle

Chapter 2

Health and Income across the Life Cycle and Generations in Europe

An age-cohort decomposition applied to panel data identifies how the mean, overall inequality and income-related inequality of self-assessed health evolve over the life cycle and differ across generations in 11 EU countries. There is a moderate and steady decline in mean health until the age of 70 or so and a steep acceleration in the rate of deterioration thereafter. In southern Europe and Ireland, where development has been most rapid, the average health of generations born in more recent decades is significantly better than that of older generations. This is not observed in the northern European countries. In almost all countries, health is more dispersed among older generations indicating that Europe has experienced a reduction in overall health inequality over time. Although there is no consistent evidence that health inequality increases as a given cohort ages, this is true in the three largest countries — Britain, France and Germany. In the former two countries and the Netherlands, at least for males, the income gradient in health peaks around retirement age, as in the US. In most European countries, unlike the US, there is no evidence that income-related health inequality is greater among younger than older generations.

This chapter is based upon:

Van Kippersluis, H., T. Van Ourti, O. O'Donnell, and E. van Doorslaer (2009), "Health and income across the life cycle and generations in Europe", *Journal of Health Economics* 28(4): 818-830.

2.1 Introduction

How does the distribution of health evolve over the life cycle and is it changing across generations? How do socioeconomic disparities in health change as individuals age and are they narrowing, or widening, across generations? These questions are addressed using comparable panel data from 11 European countries. The analysis is pertinent not only to gauging the consequences of the rapid population ageing occurring in Europe and elsewhere, but also to determining the causes of the socioeconomic gradient in health and the extent to which European countries have been successful in reducing such disparities over time. Knowledge of how the distribution of health changes over the course of the life cycle is key to understanding individual behaviour with respect to retirement, saving, health insurance and the utilisation of health care, and, consequently, to the formation of public policy concerning pensions, health financing, and health and social care. Evidence of generational differences in health is essential for accurate monitoring and projection of trends in population health, and can signal potential determinants of health, such as living conditions in childhood.

Beyond a certain age it is anticipated that health will begin to decline. But at what age, on average, does this decline set in? What is the rate of decline? And from which age does this accelerate such that the deterioration in health becomes rapid? The answers to these questions have important implications for a wide range of policies including, for example, increases in the retirement age that are being implemented, or considered, in a number of European countries. Besides identifying the average rate of change in health across the life cycle, it is also important to establish whether health becomes more or less dispersed as a cohort of individuals ages. Deaton and Paxson (1998) argue that if shocks to health are permanent, then their cumulative effect will result in health being more widely dispersed at older ages. In this case, there would be welfare gains from pooling risks across periods to provide protection against the cumulative deterioration of health, or at least its financial consequences, but these may not be realisable given the enforceable insurance contracts currently offered by markets (Diamond, 1992; Cochrane, 1995; Pauly *et al.*, 1995). Divorcing health financing contributions from risks, as is done in most European social insurance and tax financed systems, may partially solve this problem but relies on within, as well as between, generation solidarity that could become increasingly strained as a cohort ages and redistribution from the healthy to the unhealthy increases. A further implication if health dispersion does indeed increase with age is that ageing of the population would lead to greater total inequality in health, providing there were no offsetting differences across generations.

The existence of socioeconomic inequality in health in Europe and elsewhere is firmly established (Adler *et al.*, 1994; Van Doorslaer *et al.*, 1997; Mackenbach *et al.*, 1997; 2008; Smith, 1998; 1999; Van Doorslaer and Koolman, 2004), but its causes are not yet well understood. Examination of how the socioeconomic gradient in health varies across the life cycle can help reveal its origins (Case *et al.*, 2002; Smith, 2007; Currie *et al.*, 2007). For example, there is some US evidence of the gradient peaking around retirement age (Elo and Preston, 1996; Smith and Kington, 1997; Deaton and Paxson, 1998; Smith, 2005a). This is consistent with the gradient in large part reflecting income losses from illness-induced interruptions to work, which obviously cease after retirement. While a couple of European studies confirm this inverse U-shape in the age profile of the health-income gradient (Van Ourti, 2003; Kamrul Islam *et al.*, 2009), others show that it continues to prevail post retirement (Gerdtham and Johannesson, 2000; Burström *et al.*, 2005), albeit less so for an occupation based measure of socioeconomic status (Marmot and Shipley, 1996), and even to increase in early old age (Chandola *et al.*, 2007). This is relevant to an ongoing debate, to which we aim to contribute, over whether the relationship between health and socioeconomic status follows a process of cumulative advantage, with early life disparities in health becoming magnified over the life cycle (Ross and Wu, 1996; Lynch, 2003; Wilson *et al.*, 2007; Kim and Durden, 2007), or whether health problems that inevitably arise in the course of time act as a leveller and so narrow socioeconomic disparities in old age (Kunst and Mackenbach, 1994; Elo and Preston, 1996; Deaton and Paxson, 1998; Beckett, 2000; Case and Deaton, 2005; Herd, 2006; Kim and Durden, 2007).

The distribution of health may differ between generations because of changes in the health system, such as the extension of coverage, or advances in medical technology that are effective for the treatment of age-specific conditions. Differences in the health, economic and social conditions experienced in early childhood are also potentially important contributors to inter-generational differences in health. There is considerable evidence supporting a strong link between early childhood, even intrauterine, experiences and health in later life (Barker, 1995; Case *et al.*, 2005; van den Berg *et al.*, 2006; Bozzoli *et al.*, 2007). The socioeconomic gradient in health may differ across generations in response to changes in wider social and economic disparities (Pappas *et al.*, 1993; Preston and Elo, 1995; Deaton and Paxson, 1998). These cohort effects potentially confound age effects, but they are also of interest in their own right. Not least for the purpose of predicting the consequences of population ageing, one wants to know whether younger generations are, on average, healthier than older generations, whether health inequality is increasing or decreasing, and whether the socioeconomic gradient in health is becoming more or less steep over time.

In this paper we describe how the distribution of health and income-related health inequality evolve over the life cycle and differ across generations in 11 European Union (EU) countries. This is the first study to disentangle age and cohort effects for the mean level of self-assessed health, as well as for overall and income-related health inequality, across a large number of European countries. It enables us to establish whether the life cycle evolution of health and its socioeconomic gradient are consistent across countries. Any observed differences in the age profiles may reflect differences in welfare systems and their consequences for the way in which age, health and income interact. The 11 countries studied span the north and south of Europe and so differ in the scale and timing of the economic and social development, as well as changes in public health and nutrition conditions, experienced over the lives of the adults represented in the data. We examine whether this heterogeneity is reflected in the size and nature of generational differences in the distribution of health.

The analysis is closely related to that of Deaton and Paxson (1998), who describe how average self reported health, its variance and its correlation with income vary with age in US cohorts. They find that average health declines with age at a surprisingly constant rate. This finding is confirmed for Britain by Sutton (2004) and for Sweden by Burström *et al.* (2005). The US and European studies are less consistent with respect to the direction of the cohort effect. Deaton and Paxson (1998) find that average health is lower for older cohorts but there is no difference between cohorts born after 1945, while Sutton (2004) and Burström *et al.* (2005) find that younger cohorts report lower health.

Deaton and Paxson (1998) find that the variance in health is increasing up to the age of 60, after which it remains constant. This is only partially consistent with their prediction that, if health shocks are permanent and not perfectly correlated across individuals, then the variance should be monotonically rising with age for a fixed cohort of individuals. The incomplete empirical verification may be due to selective mortality but it could also indicate invalidity of the assumptions made about the evolution of health. Deaton and Paxson assume that health shocks are additive and independent, and so health follows a random walk. If instead one assumes multiplicative shocks deriving from the depreciation of health capital (Grossman, 1972b), then the prediction of increasing variance with age no longer holds[1]. We add to Deaton and Paxson's previously sole analysis of the relationship

[1]Deaton and Paxson (1998) propose that the evolution of health is given by $h_{i,t} = h_{i,t-1} + u_{i,t}$ (Deaton and Paxson, 1994). Assuming zero covariance between lagged health and the health shock (u), $var(h_{i,t}) - var(h_{i,t-1}) = \sigma_u^2 > 0$ if health shocks are not perfectly correlated across individuals. But if health evolves according to a depreciation process, $h_{i,t} = h_{i,t-1}(1 - \delta_{i,t})$, where $\delta_{i,t} \in [0, 1)$ represents the stochastic rate of depreciation that is assumed independent of the level of health, then $var(h_{i,t}) - var(h_{i,t-1}) = -2\bar{\delta}_t var(h_{i,t-1}) + E(h_{i,t-1}^2)E(\delta_{i,t}^2) - (E(h_{i,t-1}))^2 \bar{\delta}_t^2$ (Goodman, 1960). The first

between age and the variance of health with the purpose of improving understanding of how health evolves over the life cycle. We also seek to establish whether Deaton and Paxson's finding of a smaller variance among later born cohorts, indicating that the distribution of health in the US is becoming more compressed, is confirmed for European countries.

Deaton and Paxson (1998) find that the income gradient in health is greater among younger cohorts in the US, such that socioeconomic inequality in health has been rising while total health inequality, measured by the variance, has been falling. Kamrul Islam *et al.* (2009) find that socioeconomic inequalities in reported health have been increasing over time in Sweden, but Ferrie *et al.* (2002) and Burström *et al.* (2005) find little or no evidence of increasing socioeconomic inequality in morbidity in the UK and Sweden respectively[2]. We substantially extend the evidence on how socioeconomic inequality in health is changing across Europe through analysis of common measures of health and income from 11 EU countries.

The remainder of this paper is organized as follows. Section 2.2 describes the data and the measurement of health using utility scores to scale self-reported health categories. This procedure avoids the assumption, made by Deaton and Paxson (1998), that health declines linearly across categories and results in a more plausible age profile of health. The implications for the measurement of health inequality are also discussed here. Section 2.3 presents the methodology for separating age from cohort effects using the panel data. In section 2.4 we present the results and in section 2.5 their robustness is checked to different aspects of the methodology. In the final section we summarise the results and discuss their implications.

2.2 Data and measurement of health

We use data taken from the full eight waves (1994-2001) of the European Community Household Panel (ECHP). The ECHP was designed and coordinated by the European Statistical Office (EUROSTAT). It consists of a panel of private households providing data on socioeconomic, demographic and health characteristics of non-institutionalised individuals aged 16 or older. We restrict attention to adults aged 18 and above. The

term in the latter equation is negative, while the second term is positive; which is larger depends upon the distribution of $\delta_{i,t}$. For example, if the depreciation rate is a constant, then the variance will be decreasing over time (Deaton and Paxson, 1994).

[2]Burström *et al.* (2005) do report that inequalities in life expectancy and QALYs have been increasing over time. Pappas *et al.* (1993) also find increasing socioeconomic inequalities in mortality in the US.

questionnaire is standardized across all countries. We use all waves that are available for 11 EU member states: Belgium, Denmark, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain and the United Kingdom. Luxembourg is not included due the small size of the sample and Austria, Finland and Sweden are not analyzed due to the limited number of available waves (respectively 7, 6 and 5). For Germany and the UK, we do not use the original ECHP samples (which only ran for the first three waves) but instead use the corresponding waves from the German Socio-economic Panel (GSOEP) and the British Household Panel Survey (BHPS)[3]. One wave (9) of the BHPS is dropped because of an inconsistency in the health question (Hernández-Quevedo *et al.*, 2008) and the 2002 wave (12) is added as a replacement. As a result, we analyse exactly eight waves for all countries. Table 2.2 in the appendix lists the sample sizes and some descriptive statistics from the data. All analyses are conducted separately for each of the 11 countries.

The two key variables in the analysis are health and income. The ECHP income measure is annual disposable (i.e. after-tax) household income, including income from work, investments, property, private transfers, pensions and other direct social transfers. Indirect social transfers (e.g. reimbursement of medical expenses), in kind benefits and imputed rent from owner-occupied accommodation are not included. Income is divided by the OECD modified equivalence scale in order to account for differences in household size and composition[4]. In the BHPS an additional file is used to derive the annual disposable income (Bardasi *et al.*, 2007). Concerns about the appropriateness of disposable income as an indicator of the living standards of the elderly (see e.g. Van Ourti, 2003; Fahey *et al.*, 2004) are less worrisome here as it is only used to rank individuals within cohorts/generations.

Information on health is from the question, "How is your health in general?" with a five-point response scale ranging from very good to very bad. A potentially important difference in the UK — to which we return later — is that the BHPS question instructs respondents to rate their health relative to others of the same age[5]. This self-assessed health (SAH) variable is widely used and is known to be a very good predictor of other health outcomes, including mortality (see e.g. Idler and Benyamini, 1997). One important drawback is that it is ordinal and so statistics, such as the mean, variance and other

[3]The BHPS only covers Great Britain, but for convenience we refer to the UK throughout the text.

[4]The OECD scale gives a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 14 in the household.

[5]The SAH question in France and Germany was consistent with the other ECHP countries but a 6 and 10 point scale respectively was used and responses were recoded into the common 5 point scale by Eurostat.

inequality measures, cannot be computed directly from it. Some scaling is required. A simple 1 to 5 scale for very bad to very good health is obviously arbitrary and imposes a number of restrictions, including that movements between subsequent SAH categories always represent the same change in health, which may distort patterns observed in summary statistics of the scaled health variable. Our strategy is to transform the ordinal SAH information onto an interval scale and then adopt summary measures of inequality that are suitable for a variable with this measurement property.

We attach utility scores to the SAH categories using data from the 2001 Canadian Community Household Survey (CCHS), which is a large scale household survey that contains data on both the five-point SAH question available in the ECHP and a general health measure on an interval scale — the Health Utility Index Mark 3 (HUI3) (Furlong *et al.*, 2001; Feeny *et al.*, 2002). The HUI3 uses a multi-attribute function to transform an individual's health status measured on eight domains into a utility score derived from community preferences over health states (Furlong *et al.*, 2001; Feeny *et al.*, 2002). The index is scaled such that a value of 0 represents death and 1 indicates perfect health. It has been shown that the profile across SAH categories of HUI3 is concave with the health difference between very good and good SAH being smaller than that between good and fair (see Van Doorslaer and Jones, 2003). We attach the SAH category specific means of HUI3 computed from the Canadian sample to the respective SAH categories in the ECHP data. Obviously this is somewhat restrictive in that it does not allow for any differences between the EU countries and Canada in the profile of mean health utility across SAH categories. But this is no more restrictive than imposing the same arbitrary scale on SAH categories across all countries and there is no information available that would allow the scale to differ across countries. Moreover, a multi-attribute utility function estimated from French data proved to be very similar to the original function estimated from the Canadian data (Le Galès *et al.*, 2002).

The interval scale property of the HUI3 has consequences for the summary statistics used to measure overall and income-related health inequality. The concentration index has been a popular measure of income-related health inequalities (Wagstaff *et al.*, 1991; Van Doorslaer *et al.*, 1997), but in a recent contribution Erreygers (2009a) has shown that its validity crucially depends upon the scaling of the underlying health variable. If the scaling is interval (as is the case for the HUI3), the concentration index will have the following undesirable properties: (i) its bounds will depend on the minimum, maximum and average value of health, (ii) its value will differ depending upon whether health (HUI3) or ill-health (i.e. one minus HUI3) is examined, and (iii) its value will not be

invariant to a positive linear transformation[6]. These properties may obscure the life cycle profile and generational differences in income-related health inequality. For example, the first property implies that the life cycle profile of average health will influence the life cycle profile of the concentration index; and the third property implies that changing the location of the HUI3, which is essentially arbitrary, may change the generational differences in the concentration index. Using an axiomatic approach, Erreygers (2009a) derives an adjusted concentration index that does not have these drawbacks, but preserves other desirable properties of the concentration index as a measure of income-related health inequality[7]. A similar argument holds for measures of total health inequality. For example, the variance, the coefficient of variation and the Gini index all require that the underlying health variable has ratio scale properties. Application of one of these three measures to the interval scaled HUI3 leads to the same three undesirable properties. Erreygers (2009b) develops an adjusted Gini index that does not have these deficiencies. It is similar to the adjusted concentration index, except that individuals are ranked by health, rather than income. The issue does not arise for mean health. Although mean health is only uniquely defined up to a positive linear transformation of the interval scaled HUI3, its life cycle profile is not affected. The latter follows from the fact that the mean of a positive linear transformed variable equals the positive linear transformation of the mean, which is not the case for the traditional measures of total and income-related health inequalities[8].

For a given SAH category, older individuals tend to have lower health as measured by HUI3 (Lindeboom and van Doorslaer, 2004). It is likely that this reflects both the inability of a crude ordinal measure such as SAH to capture all the variation in health and age differences in the thresholds of true health at which health is reported to be very poor, poor, etc (ibid). To allow for variation in HUI3 by age and sex within SAH category, and so simultaneously correct for age-sex heterogeneity in reported SAH, we assign the SAH category, age-sex specific means of HUI3 from the Canadian data to ECHP observations

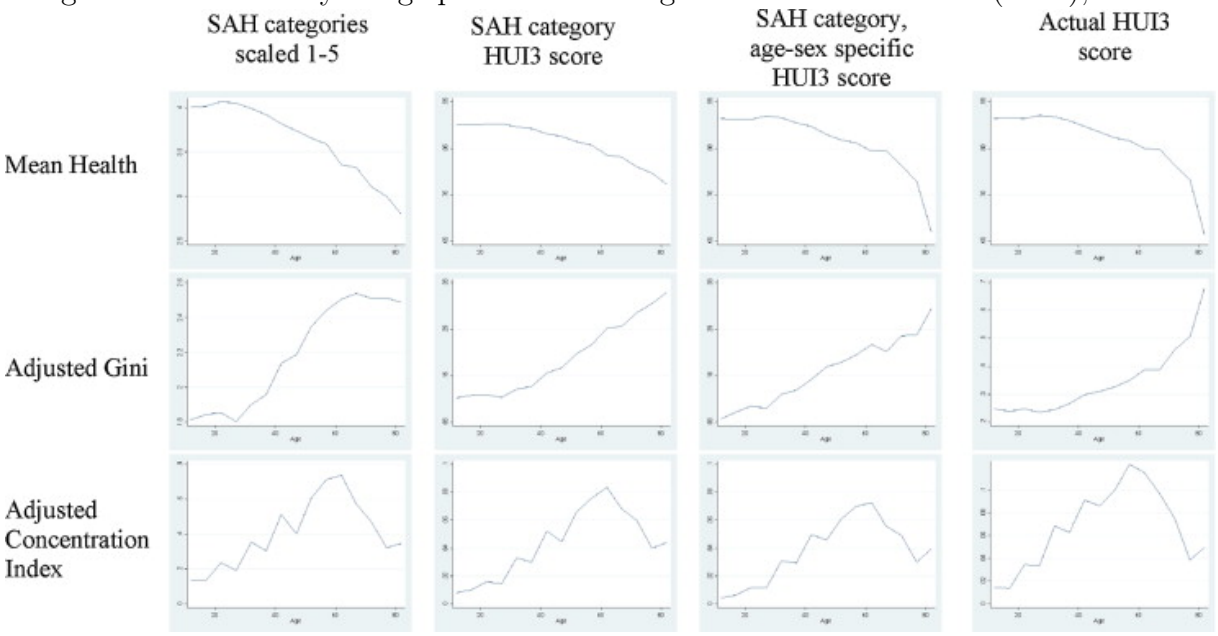
[6]Note that calculation of the concentration index requires a cardinal measure of health and so it cannot be computed while SAH is left on an ordinal scale.

[7]The adjusted concentration index is defined as $4\frac{E(h_i)}{b_h-a_h}CI(h_i)$ where $CI(h_i)$ is the standard concentration index (Wagstaff *et al.*, 1991), and b_h and a_h are the maximum and minimum of the health variable h_i.

[8]Define $\tilde{h}_i = c + dh_i, c > 0, d > 0$, then, of course, $E(\tilde{h}_i) = c + dE(h_i)$. The value of mean health depends on the values of c and d, but its life cycle profile, represented by the difference in mean health between age groups, is invariant up to multiplication by the scalar d. The latter does not hold for the standard concentration and Gini indices (Wagstaff *et al.*, 1991) since $CI(\tilde{h}_i) = \frac{dE(h_i)}{c+dE(h_i)}CI(h_i) \neq c + dCI(h_i)$. For the adjusted concentration and Gini indices (Erreygers, 2009a-b) we have an even stronger result since their values are independent from c and d, i.e. $\frac{4E(\tilde{h}_i)CI(\tilde{h}_i)}{\tilde{b}_h-\tilde{a}_h} = \frac{4E(c+dh_i)CI(c+dh_i)}{c+db_h-c-da_h} = \frac{4[c+dE(h_i)]}{db_h-da_h}\frac{dE(h_i)CI(h_i)}{c+dE(h_i)} = \frac{4E(h_i)CI(h_i)}{b_h-a_h}$.

within the respective SAH category[9]. Figure 2.1 illustrates, using the CCHS data for males, the impact on age profiles of applying age-sex specific HUI3 scores to the SAH categories. Profiles are presented for the health measure's central tendency (mean), total inequality (adjusted Gini) and income-related inequality (adjusted concentration index). For each statistic, the age profile is presented for SAH categories scaled using: (i) 1 (very bad) to 5 (very good), as in Deaton and Paxson (1998), (ii) category means of HUI3, (iii) category and age-sex specific means of HUI3. A fourth graph for each statistic shows the profile using the actual HUI3 score for each observation.

Figure 2.1: Sensitivity of age profiles to scaling of self-assessed health (SAH), Males.



Source: Authors' calculations on basis of the 2001 Canadian Community Household Survey (CCHS).

Using the simple 1-5 scaling of SAH, the profile shows a rather implausible near linear decrease in mean health with age. This is similar to what was found by Deaton and Paxson (1998). Using the SAH category specific HUI3 scores has little impact on the profile for mean health but when age-sex specific scores within each SAH category are used the profile displays greater concavity indicating, as seems plausible, that mean health declines more rapidly at older, than at younger ages[10]. We conclude that SAH alone is too crude a measure to capture the age profile of mean health but scaling it by age-sex

[9]Lindeboom and van Doorslaer (2004) find no evidence of variation by income in the thresholds for reporting SAH in the 1994-95 CCHS data. Consequently, we allow only for variation in HUI3 by age and sex, and not by income, within a given SAH category.

[10]For the mean, the age profile drawn from the SAH category, age-sex specific means of HUI3 is necessarily identical to that drawn from the observation specific HUI3 scores.

specific utility scores yields plausible results. For the adjusted Gini, moving from the 1-5 scaling of SAH to age-sex specific utility scores within each SAH category brings the age profile closer to that obtained from the observation specific utility scores. But, of course, the magnitude of the dispersion can never reach that obtained from the individual specific scores. At any given age, the adjusted Gini using the age-specific HUI3 scores is about three times smaller than that obtained from the actual scores. The age profile of the adjusted concentration index is less sensitive to the method of scaling SAH.

Since the BHPS question instructs respondents to rate their health relative to others of the same age, one would expect there to be less age-related variation in reported SAH and more age-related variation in true health within each SAH category in the UK. Application of age-specific means of HUI3 within each category may therefore be particularly appropriate for the UK data and help to correct for the inconsistency in the wording of the SAH question relative to the other countries.

2.3 Decomposition of age and cohort effects

The longitudinal nature of the ECHP makes it possible to perform a true cohort analysis. Following the same individuals over time identifies the age effect by observing how their health changes as they age. In order to identify the cohort effect, we group observations by birth year intervals and compare, for example, mean health of two cohort groups at the same age across different waves. This is done through cohort level regression analysis, as in Deaton and Paxson (1998). This avoids the imposition of any restrictions on the nature of the age and cohort effects at the individual level, which would be unavoidable if the regression analysis were conducted at that level.

Although, ideally, one would construct a cohort specific to each birth year, this would result in too few observations per cohort in each wave. Instead, we construct 13 cohorts of five-year birth intervals. Since there are very few observations in the oldest cohorts we exclude individuals born before 1912. The youngest cohort was born within the period 1972-1976, with an average age of 20 in the first wave and of 27 in the last wave. The oldest cohort was born between 1912 and 1916, with an average age of 80 in the first wave and 87 in the last.

It is the ageing over time within a cohort that identifies the age effect. The cohort effect is identified through comparison between consecutive cohorts at three overlapping ages across waves. This is illustrated in Table 2.1. The youngest cohort ages on average from 25 to 27 during the last three waves of the panel and can be compared with the second cohort that covers this average age span over the first three waves. Likewise, the

Table 2.1: Average ages of the three youngest cohorts across the panel

Cohort 1	20	21	22	23	24	25	26	27							
Cohort 2						25	26	27	28	29	30	31	32		
Cohort 3											30	31	32	33	34

second and third cohorts both span the 30-32 age range during the last and first three waves respectively.

We compute each of the three statistics — mean, adjusted Gini index and adjusted concentration index — for SAH category, age-sex specific HUI3 scores (hereafter referred to as scaled SAH) for each cohort in each wave and estimate both the age and cohort effects by regressing each statistic on a full set of age and cohort dummies. Because it is likely that health dynamics differ across sexes, all analyses are performed separately for males and females. These regressions are based on 104 observations, i.e. 8 waves times 13 cohorts. We use a separate dummy for each cohort, but the construction of the age dummies is more subtle. We need to ensure that the age range captured by any given age dummy does not span more than one cohort defined by a five-year birth interval. Single year age dummies would, of course, suffice but leave too few degrees of freedom. Instead, we enter a combination of one and two-year age dummies to indicate the average age of each cohort in each wave[11].

We use all observations present in the first wave (1994). The 1994 cross-sections are made representative of the non-institutionalised populations in that year through application of the ECHP first wave sampling weights (Eurostat, 2003). To keep subsequent waves of the unbalanced panel representative of the 1994 populations, and so deal with health related attrition that has been found in these data (Jones *et al.*, 2006), we construct and apply attrition corrected weights. For all wave 1 observations, we use probit regressions to predict the probability of remaining in the sample at each of the subsequent waves as a function of the following wave 1 explanatory variables: sex, income, household composition, thirteen 5-year age dummies, five SAH dummies and an interaction between SAH and age. The wave 1 ECHP provided sampling weights are then multiplied by the

[11]More precisely, dummies are defined as follows: dummy1 (age=20), dummy2 (age=21,22), dummy3 (age=23,24), dummy4 (age=25), dummy5 (age=26,27), etc. Note that these age dummies only cover the average age of the cohort in a particular wave. So there is no dummy for ages 18-19. For the UK this structure is not suitable since the dummies do not correctly embed the overlap in ages between five year birth cohorts given that one wave of the panel is omitted. We therefore used six year birth cohorts and a combination of 1, 2 and 3 year age dummies for the UK. This ensures an overlap of three years for consecutive cohorts, as for the other countries.

inverse of these (wave specific) predicted probabilities and, for each wave, this product is used to weight each observation in the calculation of the three statistics of interest[12]. These statistics are regressed on the set of age and cohort dummies using OLS.

Note that while application of attrition corrected weights helps deal with health related attrition, there may still be some unavoidable bias from pre-sample (i.e. before the first wave) mortality and institutionalisation. If the unhealthiest have already died or been institutionalised by the time a given cohort reaches old age at the beginning of the sample period, the mean health of this cohort will be overestimated. This must be kept in mind when interpreting the estimated cohort effects.

2.4 Results

Before presenting the main results we show how each of the three statistics of interest differ by age without disentangling the life cycle and cohort effects. This serves as a point of reference against which to gauge the effect of making the age-cohort decomposition. In Figure 2.2, we present for each age group the mean, adjusted Gini index, and adjusted concentration index of scaled SAH, separately for males and females. These age profiles derive from straightforward regressions with only age dummies[13]. The point estimates, presented with 95% percent confidence intervals, show differences from the reference category of 20-24 years. Intercepts are not shown in the graphs to facilitate cross-country comparison of the curvature of the age profiles, and not their levels which are potentially more prone to cross-country reporting heterogeneity. Countries are arranged geographically by northern and southern Europe, with Ireland included in the latter group because it bears closer resemblance to those countries with respect to the main results presented below.

For all three statistics, the unadjusted age profiles display a remarkable degree of consistency across countries. In most countries, there is little difference in mean health over the age range of 20-40, between 40 and 70 there is a steady decrease in health and this becomes much steeper above the age of 70. The deterioration of health with age is less steep in the UK, the Netherlands, Belgium, Denmark and Ireland than in other countries. For the UK, it seems likely that this is at least partly due to the different wording of the BHPS SAH question that asks respondents to rate their health relative to others of the same age. The adjusted Gini index of health seems to increase with age in

[12] In order to apply this method, we exclude observations that join the panel after the first wave.

[13] Since the dependent variables are constructed for cohorts of individuals of different sizes, we use robust standard errors to correct for the resulting heteroskedasticity, if any.

all countries and the rise is most pronounced at higher ages. The increase is shallower in Belgium, Denmark, Ireland, the Netherlands and the UK. In most countries, the adjusted concentration index increases with age, reaching its peak in the age range in which most people retire (i.e. 55-65) before decreasing. Another striking observation from these graphs is that the age profiles for males and females show very little difference, especially for the mean and the adjusted Gini index. For the adjusted concentration index, gender differences are somewhat larger. For males there is often a clear peak around retirement age, while for females the pattern is more gradual, and the peak often occurs somewhat earlier.

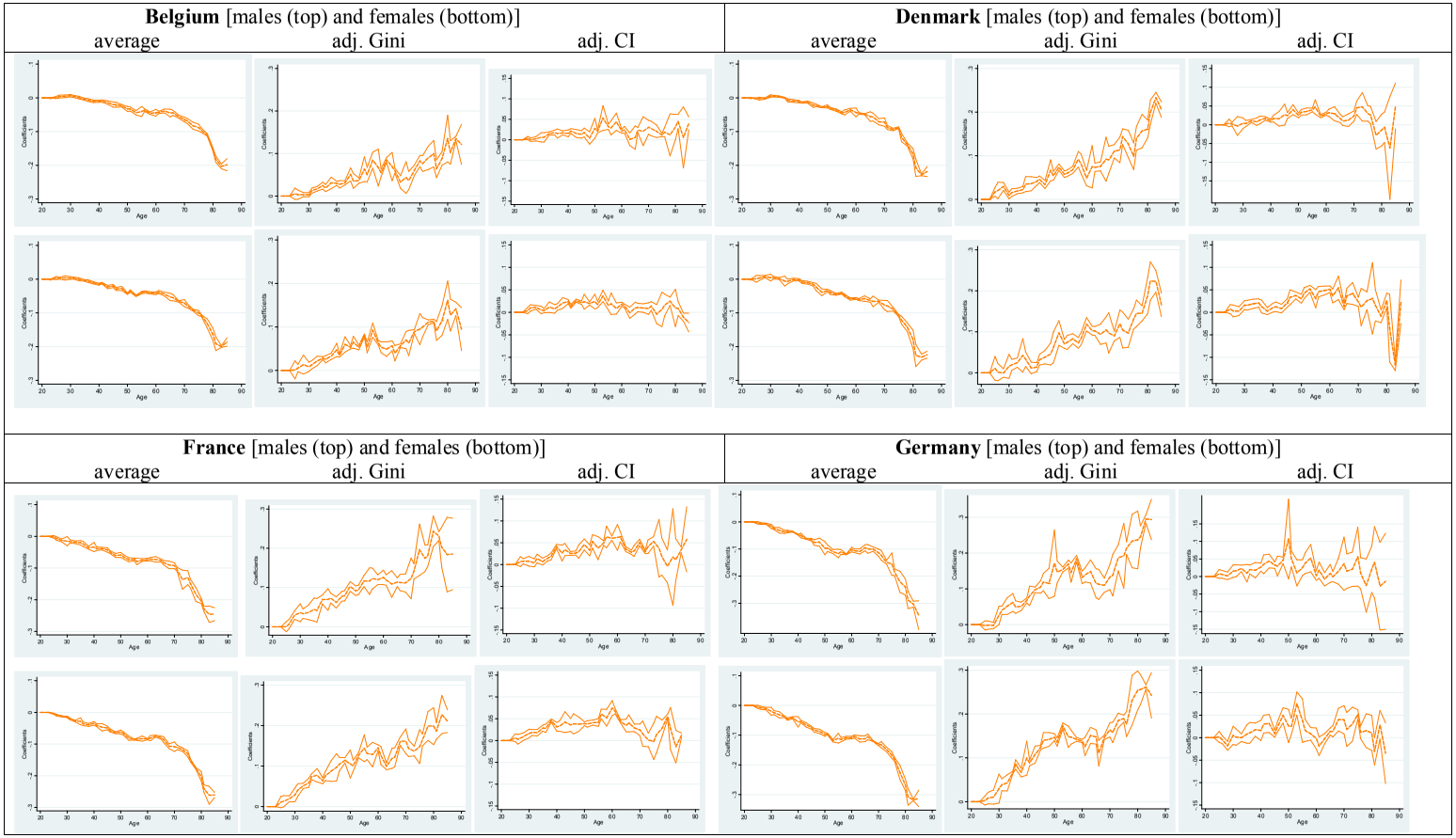
In Figure 2.3, we present the life cycle (light) and cohort (dark) profiles for each of the three statistics that derive from the age-cohort regression analysis described in the previous section. Again we present coefficients, which show deviations from the reference age (20-24) and cohort birth-year (1972-76) groups. For example, when the cohort profile for the mean lies below the horizontal at zero, it implies that earlier born cohorts experience worse health at any given age. For every country, tests strongly confirm the joint significance of all the age dummies and of all the cohort dummies for both the mean and the adjusted Gini index. For the adjusted concentration index, both the age and the cohort dummies are jointly significant, usually at much less than 5%, in every case but for the cohort dummies for females in Greece and Spain, and the age dummies for males in Germany[14].

For mean health, in general, the life cycle profiles hardly differ from the unadjusted age profiles presented in Figure 2.2. Health changes little as individuals age between 20 and 40. From the age of forty health begins to decline but there is often a levelling, or even an improvement in health between 55 and 65. The flattening of the profile around retirement age has been observed in other studies (e.g. Deaton, 2008). Deaton (ibid) conjectures that it could result from the pre-retirement group being particularly intolerant to the onset of health problems. This may be motivated by the incentives created by disability insurance in the pre-retirement age range. More optimistically, it could reflect a positive impact of retirement on health — a 'honeymoon phase' (Atchley, 1976).

Beyond the 65-70 years age range, health begins to deteriorate rapidly. In a number of countries, the degeneration of health appears to be even more compressed into the advanced years once control is made for differences in health across generations. Comparison between Figures 2.2 and 2.3 confirms that this is true for Greece, Ireland, Italy, Portugal and Spain. These differences in the age profiles reflect strong cohort effects in these (mostly) southern European countries, which are observable in Figure 2.3. In each

[14] All test statistics are reported in table A.2 in Van Kippersluis *et al.* (2008).

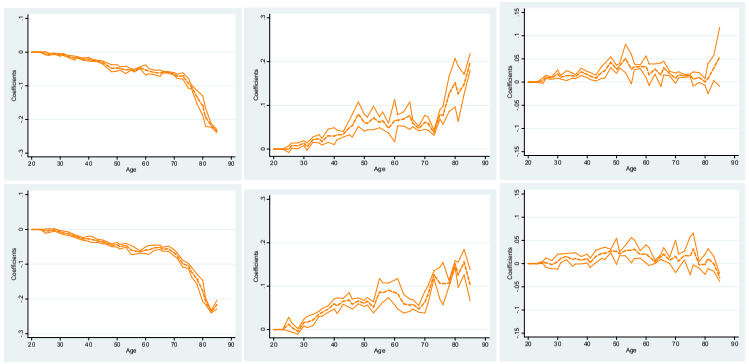
Figure 2.2: Age differences in mean, adjusted Gini, and adjusted Concentration Index of Self Assessed Health



Notes: Self-Assessed Health scaled by age-sex specific HUI3 scores, for males (top) and females (bottom) (ECHP, 1994-2001)

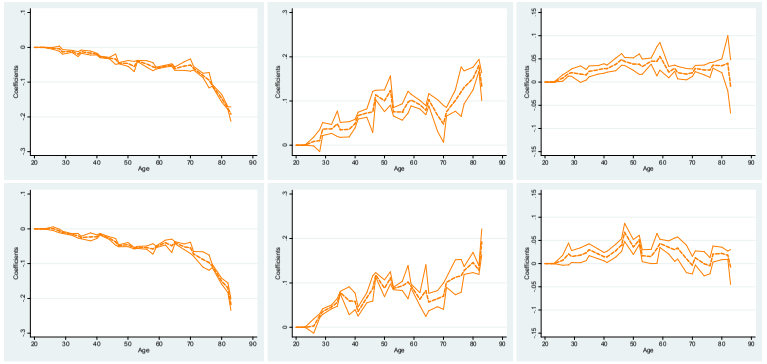
The Netherlands [males (top) and females (bottom)]

average adj. Gini adj. CI



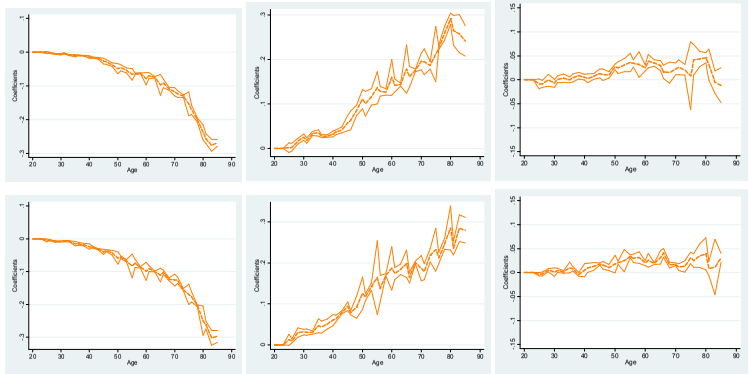
United Kingdom [males (top) and females (bottom)]

average adj. Gini adj. CI



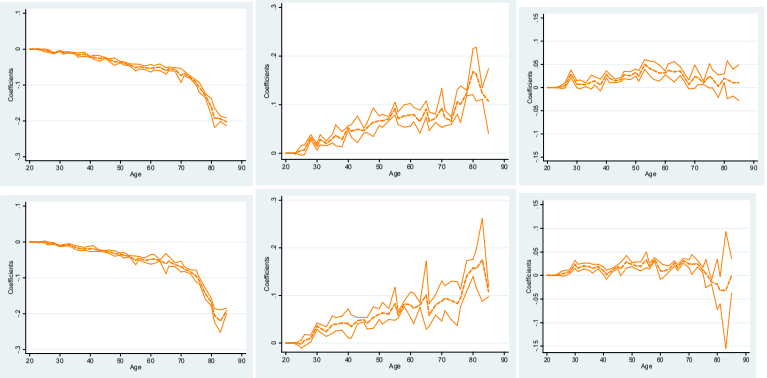
Greece [males (top) and females (bottom)]

average adj. Gini adj. CI



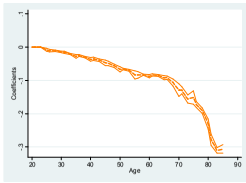
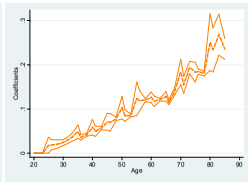
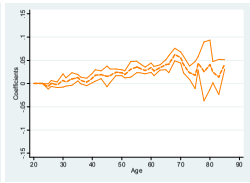
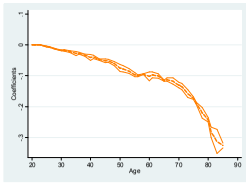
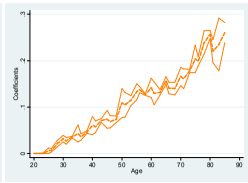
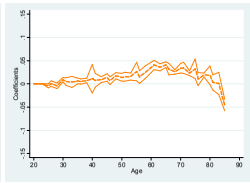
Ireland [males (top) and females (bottom)]

average adj. Gini adj. CI



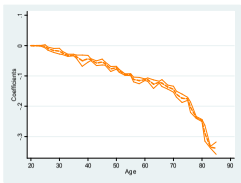
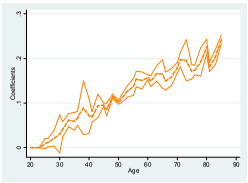
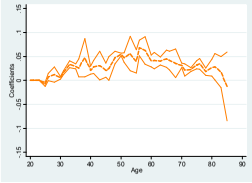
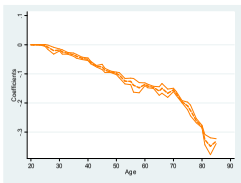
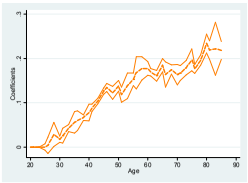
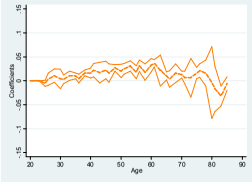
Italy [males (top) and females (bottom)]

average adj. Gini adj. CI

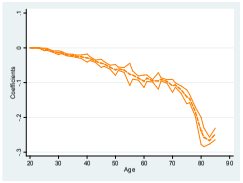
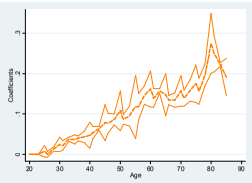
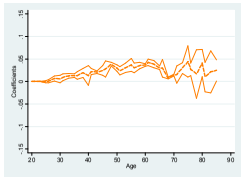
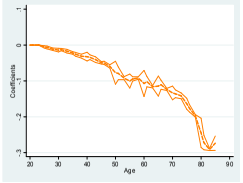
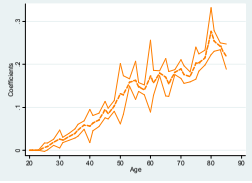
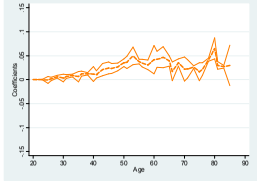
Portugal [males (top) and females (bottom)]

average adj. Gini adj. CI

Spain [males (top) and females (bottom)]

average adj. Gini adj. CI

of these countries, older generations have markedly worse health than their younger counterparts (at a given age). Horsman *et al.* (2003) estimate a difference in HUI3 of 0.05, or more, to represent a clinically important change in health. Usually the health difference between the youngest and oldest cohort is around 0.1, so this is a substantial improvement in health over generations. Notably, four of these countries are southern European and the fifth, Ireland, has also experienced very rapid economic and social development over the period spanned by the lives of the adults represented in these data. It is conjecture, of course, but it seems plausible that the generational differences in health we observe reflect the advances that have been made in these countries over the post-war era. In most of the northern European countries (Belgium, Denmark, France, Germany and the Netherlands) the cohort effects are jointly significant, but there is no evidence of any clear improvement or deterioration in average health across generations. In the UK generations born before 1930 for males and 1940 for females actually report significantly better health, on average, than more recently born cohorts. This is surprising given gains that have been recorded in objective indicators of population health, such as life expectancy. It may reflect generational differences in health expectations that confound the reporting of health, or it could be an artefact stemming from the reference to age in the BHPS SAH question, although one would expect the latter to affect the age profile more than the cohort profile. While different cohort effects for the northern and southern countries (including Ireland) clearly emerge, differential rates of (pre-sample) institutionalisation may have contributed to this. Greater use of residential, long-term care in northern countries will raise the mean health of older generations of the non-institutionalised individuals included in the samples relative to their counterparts in the southern European samples. Generational differences in mean health may then be more apparent in the southern samples.

With few exceptions, there is an upward gradient in the cohort profile for the adjusted Gini index indicating that health is more dispersed among older generations. This is consistent with what Deaton and Paxson (1998) found for the US. The tendency towards reduced inequality in health among younger generations is strongest in the countries that also show an improvement in average health (Greece, Ireland, Italy, Portugal and Spain). In the northern, more industrialized countries, particularly in France, Germany, the Netherlands, the compression of the health distribution among younger generations is less pronounced. The UK is the only country in which there is less health inequality among older cohorts — although only significantly for females — which may be related to the age reference in the SAH question, although, again, one would expect this to impact more on the life cycle profile than the cohort effects.

There is less consistency across countries in the life cycle profiles of the adjusted Gini index. The tendency for inequality to increase with age that was observed for all countries in Figure 2.2 appears to be driven by the cohort effects. Once these are controlled for there is evidence that health becomes significantly more unequally distributed as individuals age only for France, Germany, Dutch females and the UK. For these countries the evidence is even stronger than that found for the US by Deaton and Paxson (1998) in support of their hypothesis that the variability of health increases over the life cycle due the cumulative effect of permanent health shocks. But in the remaining countries, the life cycle profile of the adjusted Gini index displays a variety of patterns after the age of 50 and is actually significantly decreasing with age in Spain and for females in Ireland. There is no emergence of a stylised fact that health consistently becomes more unequally distributed as a cohort ages.

After taking out the cohort effect, the adjusted concentration index continues to show some sign of peaking around retirement age only for British, Dutch, French, and Irish males and Danish females. But the precision is low, indicated by wide confidence intervals, and the profiles could not be described as having a distinct inverted U-shape. In a few cases income-related health inequality is increasing over the life cycle (Greek males and Dutch females) and in Italy it is falling. In most countries, particularly for females, the cohort profile of the adjusted concentration index is flat. Unlike what Deaton and Paxson (1998) found for the US, there is little evidence from Europe that the income gradient in health is becoming steeper among younger generations. There is evidence that this is occurring only among Dutch females. For many of the others, if anything, the trend is in the opposite direction. But it is evident that the confidence intervals for the concentration index are often substantially wider than those for the other statistics and this might reflect a difficulty in separately identifying the age and cohort effects for this statistic.

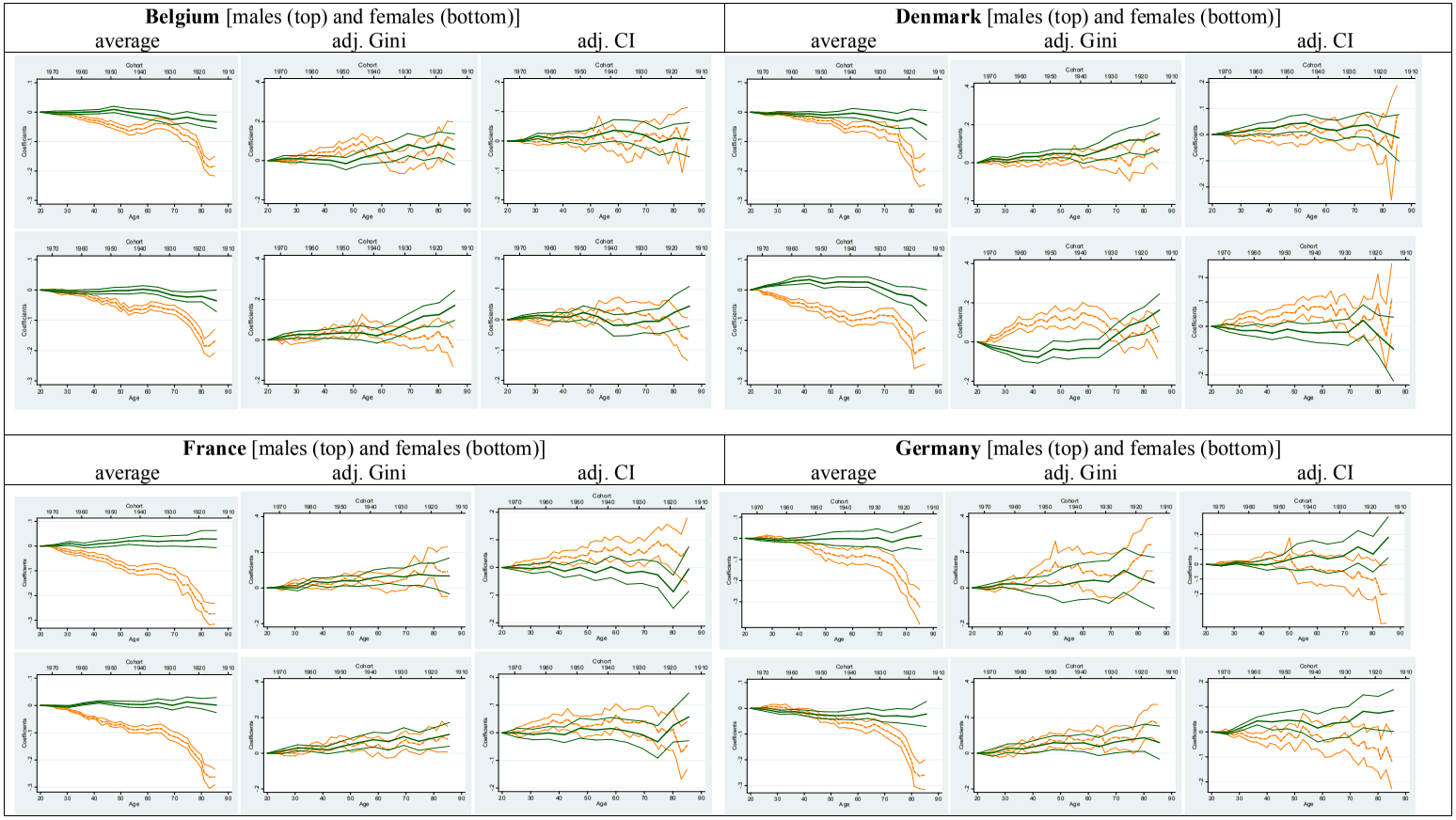
2.5 Robustness checks

In this section we check the robustness of the results to allowing for possible period effects and extending the number of waves to better disentangle age from cohort effects. All results are available on request from the authors.

2.5.1 Allowing period effects

The regressions underlying Figure 2.3 do not allow for any period effects. That is, variation in health from year-to-year that is common to all ages and cohorts. Period effects are not

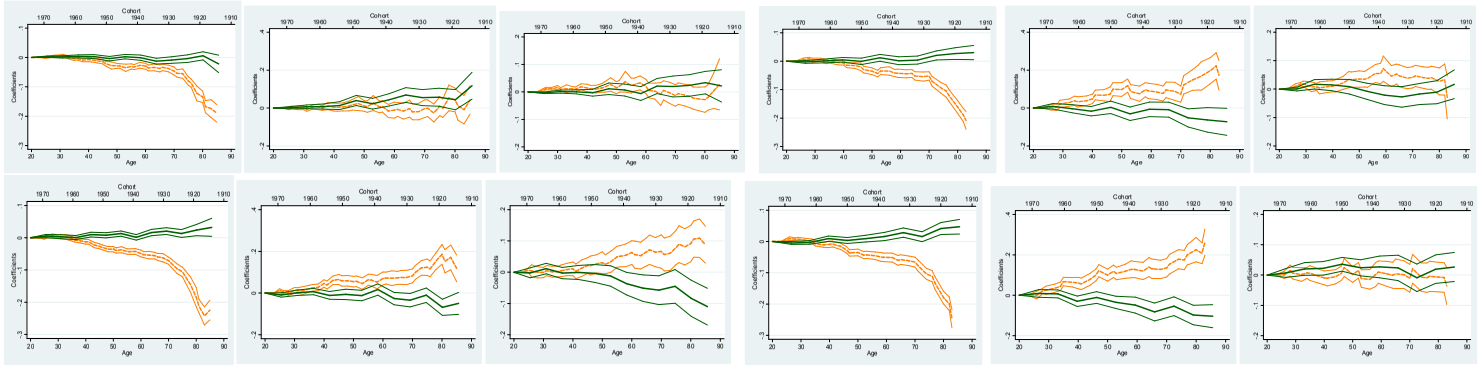
Figure 2.3: Life cycle (light, dotted) and cohort (dark, solid) profiles of mean, adjusted Gini, and adjusted Concentration Index of Self Assessed Health



Notes: Self Assessed Health is scaled by age-sex specific HUI3 scores for males (top) and females (bottom) (ECHP, 1994-2001)

The Netherlands [males (top) and females (bottom)]

average | adj. Gini | adj. CI

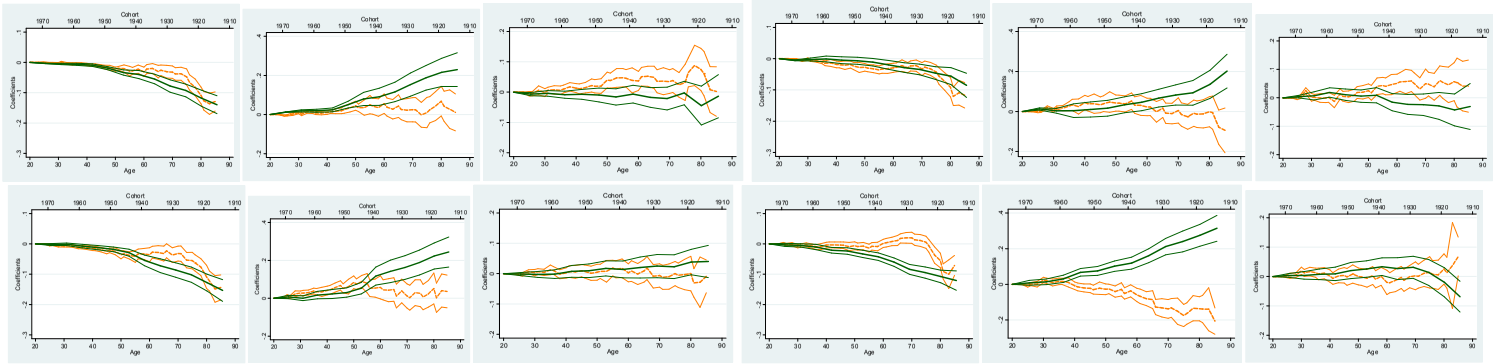


United Kingdom [males (top) and females (bottom)]

average | adj. Gini | adj. CI

Greece [males (top) and females (bottom)]

average | adj. Gini | adj. CI

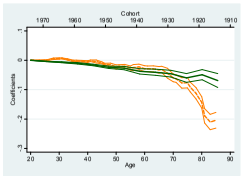
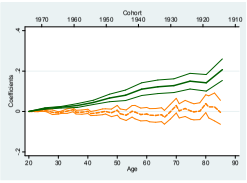
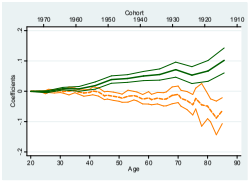
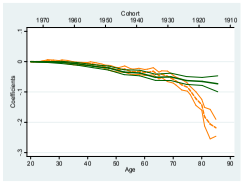
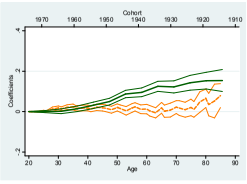
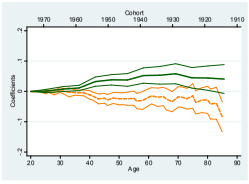


Ireland [males (top) and females (bottom)]

average | adj. Gini | adj. CI

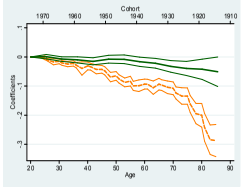
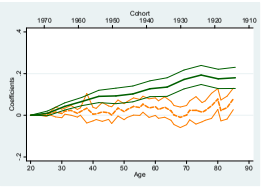
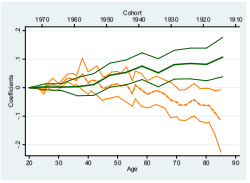
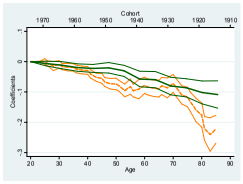
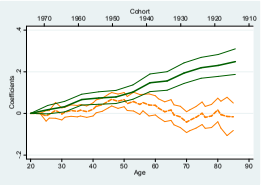
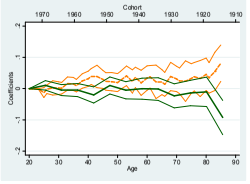
Italy [males (top) and females (bottom)]

average adj. Gini adj. CI

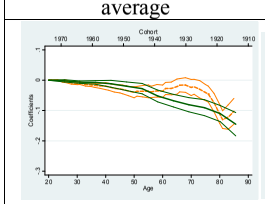
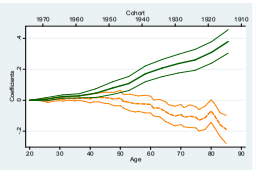
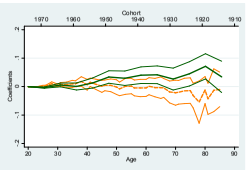
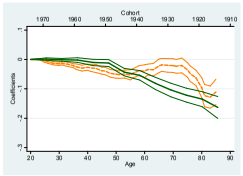
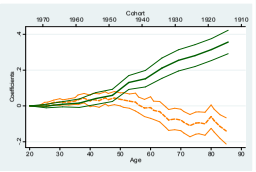
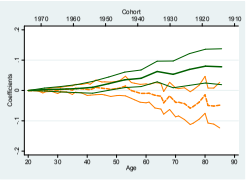
Portugal [males (top) and females (bottom)]

average adj. Gini adj. CI

Spain [males (top) and females (bottom)]

average adj. Gini adj. CI

taken into account in the basic analysis for two reasons. First, it seems likely that any common year-to-year variation in health over the span of eight years is much less important than age and cohort differences. Second, it is well known that unrestricted age, cohort and period effects cannot all be separately identified (Weiss and Lillard, 1978) and, even with restrictions, it might be overly ambitious to attempt to disentangle all three effects from only eight years of data. Yet, period effects on the distribution of health certainly cannot be ruled out entirely a priori. They could derive from a business cycle effect operating through labour market conditions, levels of stress, consumption of alcohol or accidents (Ruhm, 2003; 2005). Any changes in disability insurance eligibility rules may also affect the reporting of health. It is also possible that innovations in medical technology over the period of the study could impact on the distribution of health, although it seems unlikely that our health measure is sufficiently sensitive to pick this up.

Since we are intrinsically interested in the age and cohort effects and we expect the period effects, if any, to mainly derive from business cycle fluctuations, it seems most appropriate to achieve identification by constraining the period dummies to be orthogonal to a linear time trend (Deaton, 1997). This essentially ensures that all trends in the data are attributed to the cohort and age effects. The interpretation is now of a fixed age profile, a cohort-shift in this profile and period effects stochastically moving the data off the profile. Practically, with the orthogonality constraint and restricting the coefficients of the period effects to sum to zero, this boils down to estimating the regression with six period dummies, defined as in Deaton (1997).

For most countries and statistics, the period dummies were not jointly significant (especially for males)[15]. In any case, the inclusion of period effects lowered the precision of the estimates of the age and cohort effects greatly, while the patterns of the age and cohort profiles remained similar.

Since our time frame is rather short, making it difficult to separate cycles from a trend, we also experimented with year specific macroeconomic indicators — GDP growth and the unemployment rate — to capture any business cycle effect. Using Belgium as a test case, the regression analyses were repeated with the addition of each of the two macro indicators separately and jointly. The addition of GDP growth had a similar impact on the point estimates to the inclusion of the restricted period dummies, but without

[15]For males, only Denmark, Italy, Portugal and the UK have significant period effects for the mean. Only Spain has significant period effects for the adjusted Gini index. The effects for the adjusted concentration index are significant for British, German, and Portuguese males. For females, significance of the period effects is found for Denmark, France, Germany, Greece, Italy, Portugal, Spain and UK for the mean. For the Gini index German, Greek, Italian, and Spanish females have significant period effects. Finally, for the concentration index only German females have significant period effects.

inflating the confidence intervals as much. Only for the mean health of females was GDP growth significant. Adding the unemployment rate to the regression had a larger impact on the age coefficients for all statistics, although it was significant only in the regression of the concentration index for females. It appears that this variable is competing with age to pick up the over time variation, leading to a distorted age profile. Including both indicators gave equivalent results as adding the unemployment rate to the regression. Since allowing for period effects, either through dummies or macro indicators, reduces the precision of the estimates while leaving the age and cohort profiles much the same, we prefer the more parsimonious specification.

2.5.2 Extending the number of waves

All the analyses use eight waves of data taken from the ECHP for most countries and from the BHPS and GSOEP in the cases of the UK and Germany. The latter two panels run for more than eight waves, providing an opportunity to examine whether eight waves are sufficient to adequately identify age and cohort effects. For the UK, we repeated the analysis using 13 waves of the BHPS (1991-2004). While with eight waves the cohort effects are identified through the comparison of two adjacent cohorts at common ages across waves, with 13 waves it is possible to compare three different cohorts at the same age. As anticipated, this resulted in more precise estimates of the age and cohort profiles. But the shapes of these profiles hardly changed, except for the age profile of the adjusted concentration index for males which is no longer significant. This supports our warning at the end of section 4 that it appears more difficult to separately identifying the age and cohort effects for the adjusted concentrated index. Overall though, our findings suggest that an eight year time span is sufficient to identify the patterns of the age and cohort effects.

2.6 Conclusion

In the context of ageing European populations, it is vital to identify both how the distribution of health evolves as individuals age and also how it is changing across generations. A cross-sectional analysis of the relationship between health and age does not suffice as this age profile is confounded by generational differences in the distribution of health.

Not surprisingly, we find average health to decline with age. More interesting is the consistency in the pattern of this decline. In most countries, there is a gradual but steady fall in mean health from early adulthood until around the age of 50. The deterioration

in health generally levels off in middle-age before accelerating rapidly beyond the age of 70. Deaton and Paxson (1998) found a less plausible linear decline in health in the US, most probably because they imposed the restriction that differences in health are constant between all adjacent SAH categories. The middle-age plateau in the profile of mean health is rather puzzling. It may simply be an artefact of the insensitivity of SAH to changes in health in this age range, although the use of age-sex specific HUI3 scores within each SAH category makes this less likely. The decline in mean health between the ages of 70 and 80 is substantial and clinically significant. This rapid deterioration suggests that while health might not be a significant physical constraint on raising the retirement age toward 70, currently it would be a substantial constraint beyond that age. Of course, few individuals will wish to work until their health makes it physically impossible to continue and so the expected rapid decline in health after 70 may encourage individuals to retire well before that age while their health still permits them to enjoy their increased leisure time. While these issues are central to the debate currently being conducted in many European countries in response to proposals to raise the retirement age, we should be careful not to overstep the implications that can be drawn from our descriptive analysis. We have identified how health changes on average with age but not why it does so. We do not know what role retirement itself plays in determining the age at which health begins to decline rapidly. It could well be that a rise in the retirement age would shift the age profile of health, although whether it would be most likely to shift outward or inward is difficult to establish from the current evidence (Charles, 2002; Shan *et al.*, 2005; Neuman, 2008).

The retirement age is seldom, if ever, increased for the generation currently approaching it. Assessment of whether population health is likely to be a constraint on this policy therefore requires knowledge of the extent to which the age profile of mean health is shifting out for younger generations. Our results indicate that this is happening in southern Europe and in Ireland but not in northern Europe. The groupings of countries are consistent with anticipated differences in the timing of health benefits from improved nutrition, living conditions and access to medical care over the course of the last century and also with evidence on cohort differences in adult height across Europe (Garcia and Quintana-Domeque, 2007). Garcia and Quintana-Domeque find that the same four southern European countries that are included in the present study, plus Ireland, experienced increases in mean height much later than northern European countries. Bozzoli *et al.* (2007) confirm this finding and demonstrate that the post neonatal mortality rate is negatively and strongly correlated with the mean height of a cohort in adulthood. This is consistent with a strong life-long impact of health conditions experienced in infanthood

(Barker, 1995; Case *et al.*, 2005; Van den Berg *et al.*, 2006). It suggests that the gains in adult health for younger generations that are observed in the ECHP data for the four southern European countries and Ireland, but not for the northern countries, are due to the more marked improvements over the course of the last century in the childhood health conditions registered in southern Europe. It is tempting to relate this to the later economic development of the five mentioned countries but Bozzoli *et al.* (2007) do not find any role for national income in explaining increases in adult height.

The absence of any evidence of health gains to younger cohorts in northern European countries is perhaps surprising given continuing declines in age-specific mortality rates (Lafortune and Balestat, 2007) but it does not contradict two other European studies that used EQ-5D scores to scale SAH in a way that resembles what we did using the HUI3. Sutton (2004) actually found reported health to be worse among younger cohorts in the UK. Burström *et al.* (2005) found the same result in Sweden up to the age of 45, but beyond this age reported health was better for younger cohorts at any given age. Moreover, Crimmins (2004) showed that morbidity and disability show different trends for the elderly and it is possible that generational differences in health expectations are obscuring gains in health when measured by SAH. There is good evidence from the US of substantial year-on-year declines in age specific disability rates (Manton *et al.*, 2007). The evidence on disability trends for other OECD countries is less consistent (Lafortune and Balestat, 2007).

We find a dramatic fall in health inequality over time in all 11 European countries studied except France, Germany, the Netherlands and the UK, where it only slightly decreased or remained stable. This finding of lower variability in health among younger cohorts is consistent with Deaton and Paxson (1998) and implies that both the US and Europe have succeeded in decreasing the degree of inequality in population health over the past century. But, consistent with other European studies (e.g. Kunst *et al.*, 2005), there is no evidence of falling socioeconomic inequality in health across generations. While this may be considered a failure, it should be judged alongside the US evidence of increasing socioeconomic inequality (Deaton and Paxson, 1998).

We do not find consistent empirical support for Deaton and Paxson's (1998) hypothesis of rising variability in health over the life cycle. But the prediction is confirmed for the three largest countries — France, Germany and the United Kingdom — and more emphatically so than is true for the US (Deaton and Paxson, 1998). These are not the three health systems that one would immediately think of as being least likely to provide effective health care such that health shocks accumulate and dispersion increases over the life cycle. Rather than attempt to reconcile the evidence with the hypothesis, it is perhaps

more appropriate to reconsider the theory. It may well be that a model of permanent, additive and uncorrelated health shocks does not best describe the evolution of health and so cannot predict the course of its variability over the life cycle. The issue deserves further study.

Without controlling for cohort differences, we confirm US evidence (Elo and Preston, 1996; Smith and Kington, 1997; Deaton and Paxson, 1998; Smith, 2005a) of an inverse U-shaped income gradient in health that peaks around the retirement age. But after taking account of cohort effects, the inverse U-shape prevails only for British, Dutch, French, and Irish males and Danish females. In fact, while the dummies for age and cohort effects are still jointly significant in most countries, for males as well as for females, there is no single dominant pattern appearing for the majority of countries. In general, it simply seems much harder to disentangle age and cohort effects with sufficient precision for the adjusted concentration index than for the mean and adjusted Gini index of SAH.

An important potential limitation of our study is that we identify the cohort effects from only three overlapping ages in only two adjacent cohorts. The strong assumption we have to impose is that health differences between cohorts at these three ages are representative of the differences over the whole life-span. While we have confirmed the robustness of our results for the UK using 13, rather than 8, waves, a pseudo-panel approach of a long series of cross-sections would provide more identifying information and may allow testing of whether the cohort effects merely shift and do not tilt the age profiles. It would, however, lose the cross-country dimension of the present study. A second limitation is the absence of institutionalised individuals in the ECHP. Our attrition corrected sampling weights correct for the onset of institutionalisations during the time period of our panel, but not for any pre-sample mortality and pre-sample institutionalisation. We thus cannot exclude that mean health of the very old is overestimated and that the cohort effects would be steeper in the absence of such pre-sample biases. Given differential rates on institutionalisation in the north and south of Europe, this may be confounding the north-south comparison of cohort effects. A third, unavoidable limitation is that we can only correct for selective attrition based on observable, and not unobservable, characteristics. Despite these limitations, our findings appear relatively robust to a number of methodological choices and restrictions imposed by the data. Future work should concentrate on whether age and cohort profiles in health differ by socio-demographic groups within countries. This would allow consideration not only of how individuals age on average, but how the health of certain socioeconomic groups deteriorates relative to that of other SES groups (Case and Deaton, 2005).

Appendix

Tables

Table 2.2: Descriptive Statistics

	FRA	GER	NL	UK	GR	ITA	POR	SPA	BE	DK	IRE
Obs. Wave 1	13166	11829	9048	7809	11859	16657	10936	16883	6427	5652	9472
Obs. Wave 2	11623	11041	8140	6841	10506	15488	10208	14006	5749	4919	7572
Obs. Wave 3	11039	10518	7786	6618	9640	14967	9671	13066	5316	4385	6405
Obs. Wave 4	9926	10053	7347	6359	8877	13672	9281	11754	4812	3956	5678
Obs. Wave 5	9156	9413	6769	6105	8837	14239	10019	11835	4809	3851	5462
Obs. Wave 6	8523	8948	6200	5550	8331	13553	9762	11198	4443	3623	4621
Obs. Wave 7	7902	8480	5731	5287	8104	12692	9461	10424	4150	3425	3720
Obs. Wave 8	7601	8070	5106	5021	8012	11516	9237	10044	3736	3368	3248
Mean HUI3	0.87	0.84	0.90	0.88	0.89	0.87	0.83	0.87	0.89	0.90	0.91
Mean income	184928	57515	55377	19841	4295866	39169	2339057	2876558	1224194	249015	21579
Mean age	47.55	46.16	47.43	47.31	49.54	46.30	49.11	47.66	47.35	46.86	46.15
Proportion of males	0.47	0.48	0.47	0.47	0.47	0.49	0.48	0.48	0.47	0.49	0.49

Notes: HUI3 refers to the Health Utility Index Mark 3. Income is measured in national currency units.

Chapter 3

Socioeconomic Differences in Health over the Life Cycle in an Egalitarian Country

A strong cross-sectional relationship between health and socioeconomic status is firmly established. This paper adopts a life cycle perspective to investigate whether the socioeconomically disadvantaged, on top of a lower health level, experience a sharper deterioration of health over time. Data are drawn from the Dutch Health Interview Surveys covering the period 1983 - 2000. The analysis focuses on the self-rated health and disability of persons aged 16 – 80. We show that in the Netherlands, as in the US, the socioeconomic gradient in health widens until late-middle age and narrows thereafter. The analysis and the available evidence suggests that the widening gradient is attributable both to health-related withdrawal from the labor force, resulting in lower incomes, and the cumulative protective effect of education on health outcomes. The less educated appear to suffer a double health penalty in that they begin adult life with a slightly lower health level, which subsequently declines at a faster rate.

This chapter is based upon:

Van Kippersluis, H., O. O'Donnell, E. van Doorslaer, and T. Van Ourti, (2010), "Socioeconomic Differences in Health over the Life Cycle in an Egalitarian Country", *Social Science and Medicine*, 70(3): 428-438.

3.1 Introduction

Health differs by socioeconomic status. The socially and economically advantaged enjoy better health, irrespective of whether it is measured by morbidity, disability or mortality. This strong socioeconomic gradient in health is firmly established in evidence from both the developed and the developing world (Marmot, 1999; Smith, 1999; Mackenbach *et al.*, 2002; CSDH, 2008). Despite an abundance of literature, the causal mechanisms and pathways responsible for the association are still poorly understood. In part, this stems from the static nature of much of the analysis. Both health and socioeconomic status (SES) are multidimensional and dynamic, and the relationship between them may reflect different effects and feedbacks over the course of life. Does low economic status in early life lead to the development of health problems in middle and later life? Or, does poor health interfere with the acquisition of education and, subsequently, the chances of securing, or holding onto, a well-paid job? The difficulty of answering such questions led Cutler *et al.* (2008) in a recent review to remark that *"(. . .) differential patterns of causality make a single theory of socioeconomic gradients in health difficult to imagine. We suspect, though, that the right theory will emphasize the life cycle."*

The advantage of bringing a life cycle perspective on the gradient has proved extremely useful already in the United States, where Smith (1999, 2005a) and Case and Deaton (2005) have challenged the common view that the socioeconomic gradient in health reflects the effect of socioeconomic status, in particular income, on health, arguing instead that a large part of the gradient derives from a feedback effect of health on income through labor force participation. Banks *et al.* (2009) made a first attempt to unravel these mechanisms in the UK and their preliminary conclusion is that the same mechanism is able to explain an important part of the gradient there as well.

This paper does not aim to present evidence on the causal impact of SES on health, or vice versa, but rather to motivate such analysis by describing how socioeconomic differences in health evolve over the life cycle. More specifically, we investigate whether the stylized facts emerging from the US literature are also apparent in Dutch data. The Netherlands differs markedly from the US in relation to social structure, income inequality, health and disability insurance, social protection and health care organization (see e.g. Hurd and Kapteyn, 2003). For example, the gross replacement rate for disability insurance benefits in 1993 was 63 percent in the Netherlands, as opposed to only 30 percent in the US (MacFarlan and Oxley, 1996). Similar large differences exist for unemployment benefits and social assistance, where also the maximum period of entitlement is much longer in the Netherlands than it is in the US (MacFarlan and Oxley, 1996). No

doubt as a consequence of these differences, socioeconomic inequality in health is much lower in the Netherlands than it is in the US (Van Doorslaer *et al.*, 1997). But little or nothing is currently known about whether and how the countries differ in the way in which the socioeconomic gradient in health changes over the life course. If there were no such differences, it would suggest that the observed patterns result from the fundamental relationships between education, occupation, work and health over the life course, and are not responsive to the social, health and economic policy environment.

A life cycle perspective additionally provides information on how much more rapidly health declines for some groups than others. While it is clear that, at any given age, the socially disadvantaged experience a lower level of health, there is no consensus over whether they can also expect their health to deteriorate more rapidly. On the one hand, proponents of the cumulative-advantage hypothesis maintain that differences in health by SES are established early in life and subsequently widen as the economic and health disadvantages of the less privileged interact and accumulate (House *et al.*, 1994; Ross and Wu, 1996; Lynch, 2003; Willson *et al.*, 2007). The competing view — the age-as-leveler hypothesis — maintains that deterioration in health is an inevitable part of the process of ageing irrespective of economic means or social position, with the result that the SES-health gradient narrows at advanced ages (Beckett, 2000; Herd, 2006). If there is cumulative advantage, then interventions that were effective in breaking this process by expanding opportunities to socially disadvantaged groups earlier in life would have large pay-offs in terms of improved health and labor-market prospects over the life cycle.

As in the US, we find that socioeconomic differences in health first diverge, reach a peak around late middle-age, and then converge in old age. It appears that a large part of the socioeconomic gradient in health is governed by labor force participation. Once one restrict attention to working individuals, health differences across income groups are greatly reduced and do not widen with age up to middle-age. The convergence in old age may partly stem from selective mortality — only the most robust of the lower income groups survive — although it appears also to reflect the reduced dependence of income on health after retirement.

The paper is organized as follows. Section 3.2 summarizes existing evidence on the SES health gradient over the life cycle. Although we will often refer rather loosely to the 'life cycle', since we are particularly interested in the interactions between health, work and income, we restrict attention to the years of adultood. For this reason, we do not survey either the ever-expanding literature on socioeconomic differences in the evolution of child health (e.g. Case *et al.*, 2002; Currie and Stabile, 2003; Currie *et al.*, 2007; Murasko, 2008), or that on the impact of early-life conditions on health in adulthood

(Barker, 1995; van den Berg *et al.*, 2006; van den Berg and Lindeboom, 2007). In Section 3.3, we introduce the data and methods, and in Section 3.4 evidence on the SES-health gradient over the adult life course in the Netherlands is presented. In Section 3.5 we consider the consequences of the evidence presented for health inequalities and policies to tackle them.

3.2 Related literature

Case and Deaton (2005) report large differences in average self-reported health (SRH) in the US by income quartile that increase up to age 50-55, before narrowing particularly after age 60 until they disappear by age 80. A similar life cycle pattern has been observed for other indicators of SES and health (Smith 2005a; Case and Deaton, 2005) and in one European (Belgian) study (Deboosere and Neels, 2008).

Interpretations of the observed widening and then narrowing of the SES-health gradient with age differ in the extent to which the observed pattern is presumed to reflect substantive changes in the relationship between SES and health over the life course, as opposed to simply being the product of methodological limitations. According to the cumulative-advantage hypothesis, the SES-health gradient increases over the life course possibly due to gestation of the effects. For example, the health effects of socioeconomic differences in smoking become apparent only in middle age (Lynch, 2003). Alternatively, social advantages in factors that affect health (social capital, networks and information) may accumulate across the life cycle (Lynch 2003). The latter argument is also a prominent feature of economic theory (Becker, 1964; Grossman, 1972b). Heckman (2000) and Cunha and Heckman (2007) note that a higher level of health leads to a higher level of health in the next period (*self-productivity*), but additionally investments in health at later ages are more beneficial if an individual has lived a healthy life (*dynamic complementarity*). Empirical support for this hypothesis is provided by Ross and Wu (1996), Lynch (2003), Kim and Durden (2007), and Willson *et al.* (2007), among others.

The competing view, the age-as-leveler hypothesis, proposes that biological determinants increase relative to socioeconomic determinants at older ages (Herd 2006). A large literature documents the fact that at young ages much of the variation in health is associated with SES (see e.g. Case *et al.*, 2002). But biology exerts a stronger influence as we age, and there is less room for SES to play a role. Generally, the evidence is consistent with a cumulative-advantage process operating until middle age, with age indeed acting as a leveler over the years of retirement (Kitagawa and Hauser, 1973; Kunst and

Mackenbach, 1994; House *et al.*, 1994; Elo and Preston, 1996; Deaton and Paxson, 1998; Beckett, 2000; Case and Deaton, 2005; Smith, 2005a; Herd, 2006).

Smith (2005a, 2005b, 2007) uses various US panel data sets to probe the SES-health relationship in an effort to unravel the causal mechanisms that may be responsible for it. His argument is that there is a strong causal impact (mediated through work participation) of health on financial indicators of SES, and a causal impact only of education, among the SES indicators, on health. This is founded on the timing of the onset of chronic conditions (assumed to correspond to a health shock) in relation to measures of SES. A substantial impact of a health shock on the probability of future work and the income of older Americans is taken as evidence of a strong causal mechanism running from health to SES (Smith, 2007). The effects on employment, income and wealth are strongest at ages 51-61. This offers quite a different interpretation of the quadratic relationship of the SES-health gradient with age than that of a cumulative-advantage process eventually being overturned by the leveling effect of age. Smith (2007) consistently finds education to be a strong predictor of the onset of new chronic conditions, having controlled for initial health and a multitude of background factors including employment and smoking behavior, but finds no effect of financial measures of SES on changes in health.

While Case and Deaton (2005) agree that health-related interruptions to work contribute strongly to the income-health gradient and its variation with age, they also emphasize the impact of work on health. Pooling data from the 1986-2001 US National Health Interview Surveys (NHIS), they show that the initial divergence followed by convergence of the age profile in self-reported health is also apparent by labor-force status, even at the same income percentiles. The association of health with income is swamped by the association with employment. There is evidence of health selection out of the labor force among manual but not among non-manual workers, and the manual workers that remain in employment experience a more rapid deterioration in health. This all suggests a strong role for the nature of work in explaining adult health trajectories in the US.

3.3 Data and Methodology

We examine socioeconomic differences over the adult life cycle in self-reported health, disability and mortality.

3.3.1 Analyses of self-reported health and disability

The self-reported health and disability data are obtained from the Dutch Central Bureau of Statistics (CBS) Health Interview Surveys covering the period 1983-2000 (Gezondheidsenquête 1983-1996 and Permanent Onderzoek Leef-Situatie (POLS) 1997-2000). These are annual, nationally representative cross-section surveys of the non-institutionalised population including children. We restrict attention to adults aged 16 to 80 containing between 5,500 and 8,500 observations per year and a total of 121,232 observations across the 18 pooled cross-sections we analyse. In addition to health indicators, the survey provides data on income, education and employment status. Frequencies of all the variables used in the analysis for the pooled dataset are given in Table 3.1 in the appendix.

Self-reported health (SRH) is obtained from the question: "How is your health in general?" Responses are on a five-point scale: "very good","good","fair", "sometimes good / sometimes bad", and "bad". We dichotomize and focus on individuals reporting the last two categories, which we refer to as "bad health".

In addition to SRH, we examine disability both because it is a different dimension of health with a potentially distinct life cycle profile (Crimmins, 2004; Freedman *et al.*, 2007) and because it is a more objective measure that is less prone to reporting differences by age or SES that may distort the health profiles (Bound, 1991; Kapteyn *et al.*, 2007). Disability is measured using the OECD Long-Term Disability Questions on the eight domains listed in Table 3.1 (McWhinnie, 1982; Gudex and Lafortune, 2000). We create a binary indicator of whether the individual reports at least one severe disability, defined as only being able to carry out an activity with great effort or not at all, in any of the three mobility-related domains since these, rather than loss of eyesight or hearing, better correspond to work-related disability associated with poor health. While this is still a self-reported measure, it is more specific than the SRH assessment of health in general, which will obscure an age effect if individuals report their health relative to others of the same age.

The CBS Health Interview Surveys measure income as annual net household income, which for the years 1983-96 is recorded in 11 fixed categories, and for the years 1997-2000 is available in intervals corresponding to deciles. Since we use only rank in the income distribution, the inconsistency of categories is not a major limitation, which is confirmed by a sensitivity analysis excluding the last four waves. Within each income category living standards will vary partly due to variation in actual incomes and partly because of variation in needs. We have no information on the former, but we can partially account for the latter by ranking inversely by household size within each category. The resulting ranks are then used to identify age-gender-wave specific quartiles in the (per

capita) income distribution, which are used to compare the health of those at the top and bottom of the income distribution at each age.

The educational variable used is the highest education degree ever obtained, while for students we use the current educational level. This is used to distinguish between a low education group (not completed more than primary school) and a high education group (university and college graduates).

Employment status is identified up to the age of 70 and for individuals who are retired, their last occupation is coded, but this is not true for the last four cross-sections, which were excluded from the analysis by occupation. In order to compare the health of individuals in the most and least physically demanding jobs, we assign those working in industrial, craft and transport occupations to a "manual" category, and executives and specialists to a category we label "non-manual".

Our analysis is graphical, showing how the life cycle profile of bad health (disability) differs between those at the bottom and at the top of the distribution of each socioeconomic indicator — income, education and occupation — and also by employment status. We pool the 18 cross-sectional CBS surveys and compute the percentage of individuals in bad health (disability) in two-year age intervals, separately for females and males. A kernel-weighted local polynomial is estimated through the age-specific prevalence rates. Attention is restricted to adults between the ages of 18 and 80, the upper cut-off being chosen to maintain sufficient cell sizes.

Variation of health by age does not identify the life cycle health profile in the presence of cohort differences in health. We test for the importance of cohort effects by grouping observations into five-year birth intervals and then identifying the life cycle health profile for each SES group from the variation in health as each cohort ages across the repeated cross-sections (see Van Kippersluis *et al.* (2009)).

3.3.2 Analysis of Mortality

Differential rates of mortality and institutionalization by SES will bias estimates of SES differences in the life cycle health profile derived from surveys of the non-institutionalised (surviving) population. For example, 'survival of the fittest' could explain why socioeconomic differences in health appear to narrow at older ages (Lynch, 2003). To gauge the magnitude of any such effect, we examine mortality rates by income over the life cycle corrected for cohort effects. This analysis is performed using administrative data that links tax records with the mortality register for the period 1998-2005.

From the tax records, we observe incomes of a random sample of around one-third of the Dutch population, corresponding to about 5 million individuals per year, for the periods 1998-2000 and 2002-2004. Age-specific income quartiles are calculated for each year using annual household income calculated after taxes and social benefits and adjusted for household size and age structure (see Van Doorslaer *et al.* (2008) for details). These tax records are linked to the subsequent year's Municipality Register, which records each death, its timing, and the gender and date of birth of the deceased. This dataset is then linked to the Cause-of-Death Register to identify any additional deaths that are not reported in the Municipality Register. This allows us to compute the age-gender specific mortality rate for each income quartile.

3.4 Socioeconomic Differences in Health over the Life Cycle

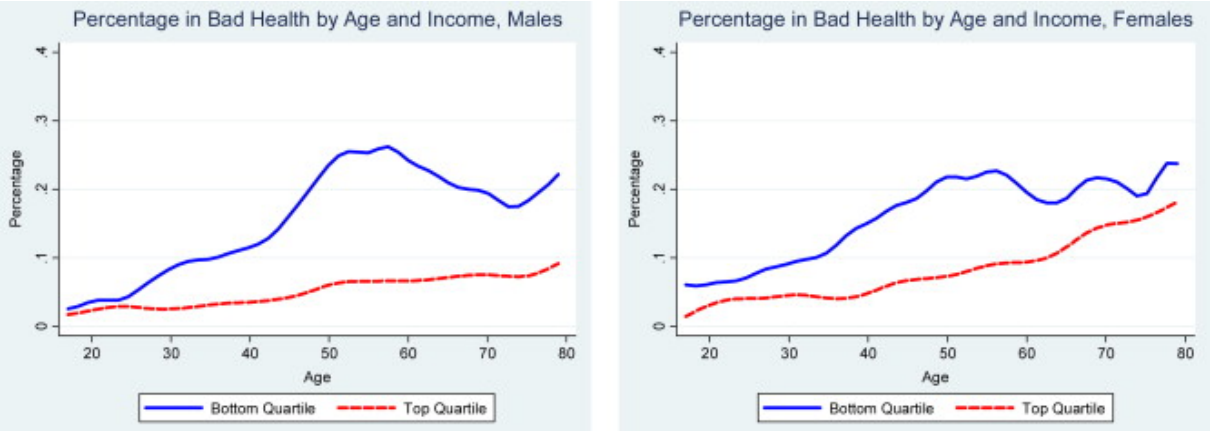
3.4.1 The life cycle profile of self-reported health by income

Figure 3.1 shows a clear income gradient in reported bad health at all but the youngest ages. For example, there is a 10% prevalence of morbidity in the bottom income quartile of females already at age 35; while in the top quartile this prevalence is only reached at 65. The age profile of the gradient displays striking consistency with the evidence reviewed in section 3.2. For both genders, but particularly for males, income differences in health diverge until 55 or so, at which point almost 30% of males in the bottom quartile report bad health, compared with only 5% in the top quartile, before converging in old age. The differences between top and bottom quartiles are statistically significant at the 5% level from age 30 on for both males and females. The increase in the income gradient until the age of 55 is due to a very sharp increase in reported poor health in the bottom quartile, which contrasts with the much more steady increase in the top quartile. This is consistent with the patterns in the US data identified by Smith (2005a). The decline in reported ill-health in the age range of 55-65 is very marked at the bottom of the income distribution, but is absent at the top. That health is apparently improving in this age range at the bottom of the income distribution is rather surprising. We return to possible explanations of this phenomenon below.

Although not shown here, when all income quartiles are examined, rather than just the top and bottom as in Figure 3.1, a health gradient emerges between the middle quartiles and the top one as age increases. Like the difference between the bottom and top, the gradient between the middle and top peaks in middle-age. But there is always a very

marked upward jump in the prevalence of bad health at the bottom quartile. Up to the age of 55, the difference in prevalence between the bottom and second bottom quartile is greater than that between the latter and the top quartile. It is particularly among the poorest individuals that health deteriorates rapidly until middle-age.

Figure 3.1: Self-reported health by age, split according to income quartile and gender



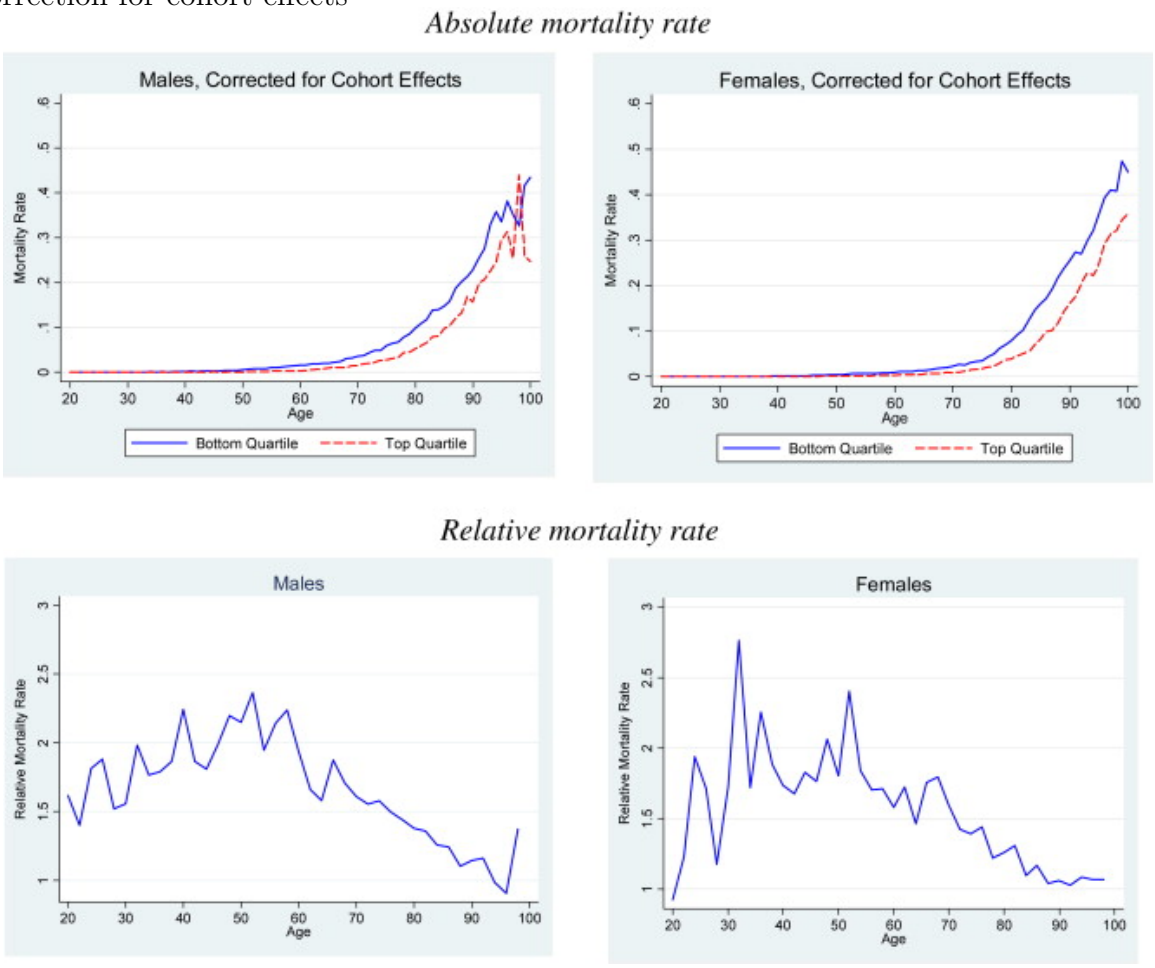
Source: Authors' calculations from CBS Health Interview Surveys, 1983-2000. Sample weights applied.

Figure 3.10 in the Appendix is the counterpart of Figure 3.1 after correction for cohort effects. Comparing the two figures it would appear that (particularly for males) any cohort effects are small and that life cycle profiles can be approximated fairly well by raw age variation. This is consistent with previous findings for a number of northern European countries (Van Kippersluis *et al.*, 2009).

Figure 3.2 (top panel) shows mortality rates by income quartile corrected for cohort effects. Beyond the age of 60, one can observe large differences in mortality rates by income. Cumulative differences are even more pronounced. Below 60, differences are difficult to discern since the absolute rates are so small. But the graphs in the bottom panel of figure 3.2 reveal that relative differences in mortality rates are largest between 50 and 60, around the age at which differences in reported health peak. These differences in mortality suggest that part of the convergence in the income-health gradient at older ages is due to the higher survival of the most healthy among the lowest income quartile.

Comparison of Figure 3.3 with Figure 3.1 reveals income differences in the life cycle profile of disability are similar to those in general health, with the income gradient again peaking, particularly for males, in middle-age. This suggests the pattern is not simply attributable to differential reporting of health. For low income women and high-income men, the prevalence of disability continues to rise beyond the age of 60, suggesting that the flattening-off observed in the respective profiles of bad health (Figure 3.1) is due to

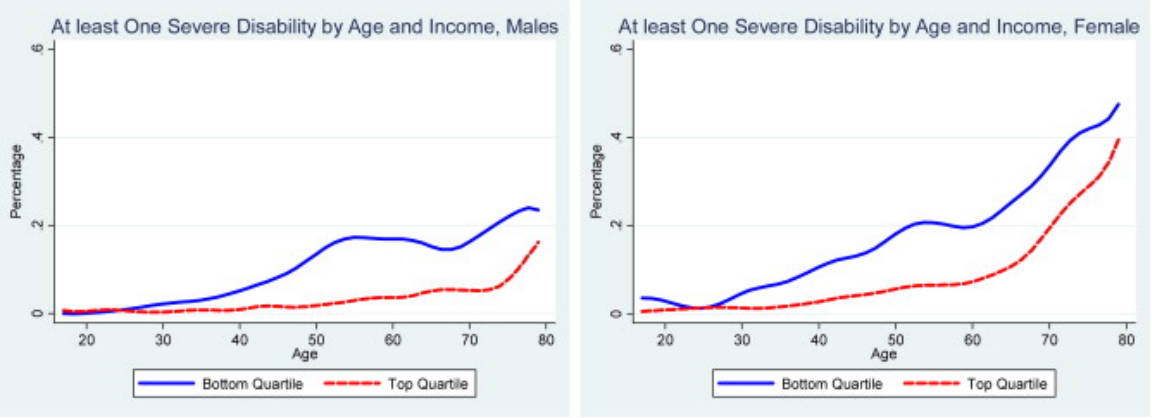
Figure 3.2: Mortality rate over the life cycle, split according to income and gender with correction for cohort effects



Source: Authors' calculations from linked Dutch administrative data, 1999-2001 and 2003-05. The vertical axis in the bottom panel shows the ratio of the mortality rate for the bottom quartile and the top quartile.

a tendency among these groups to report health relative to others of the same age. As a result, there is less difference between males and females in the convergence at older ages in the income gradient in disability than there is in the gradient in bad health.

Figure 3.3: Prevalence of disability by age, split according to income and gender

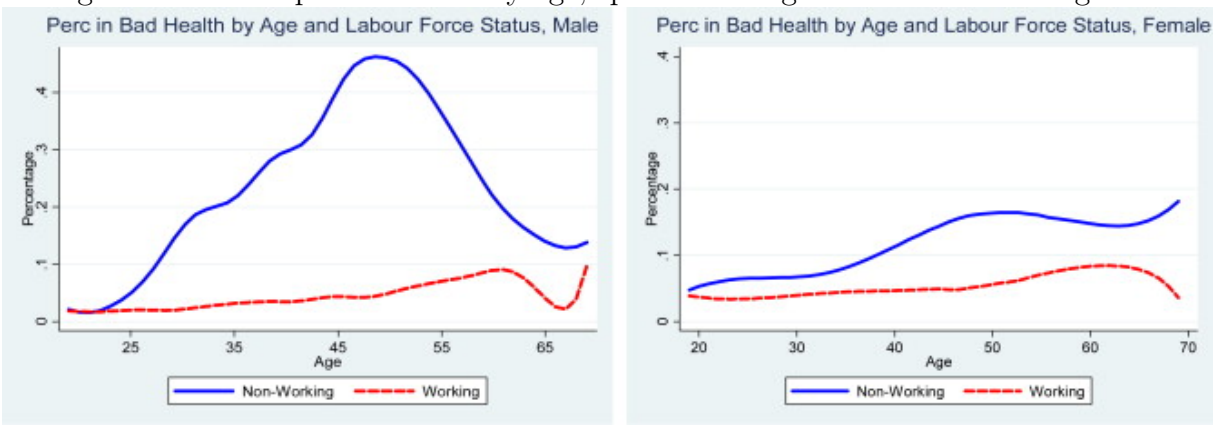


Source: Authors' calculations for CBS Health Interview Surveys, 1989-2000. Sample weights applied.

3.4.2 How much does labor-force status matter?

As discussed earlier, Case and Deaton (2005) and Smith (1999, 2005a) suggest that part of the trend observed in the income-health gradient is due to the increasing impact of health on income through employment. To gain further insight into the plausibility of this hypothesis, we now compare directly the evolution of reported health with age across workers and non-workers (Figure 3.4).

Figure 3.4: Self-reported health by age, split according to work status and gender



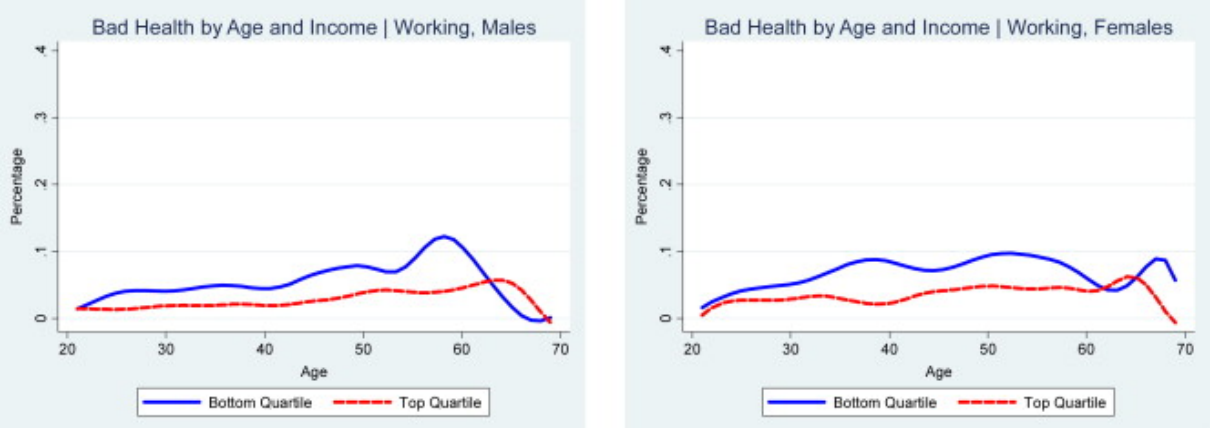
Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

The first striking observation is the rather flat profile of reported bad health for those working at each age. Practically no working males in early adulthood report bad health; the proportion begins to increase from age 30, but approaches 10% only by age 60. In contrast, already before the age of 30 around 15% of non-working males report bad health, and the proportion rises very steeply with age, such that 45% of non-working men report bad health at the age of 50. Non-working men are always in poorer health, and the widening differential suggests that health progressively becomes a more important reason for not working until age 50. The decline after age 50 in the prevalence of reported ill health amongst non-workers must be due to the growing importance of non-health reasons for not working — principally, voluntary retirement. The fact that ill health continues to rise among those that carry on working over this age range is consistent with this explanation. To the extent that income is determined by labor-force status, these graphs suggest that the observed widening and then narrowing of the income gradient in health over the life cycle may be driven by the varying effect of health on employment. Figure 3.11, presented in the Appendix, confirms that income is strongly correlated with work status; the non-participation rate is highest in the lowest income quartile, and this correlation peaks in middle age and falls substantially at older ages.

The pattern in the profile for non-working women is quite different. The prevalence of ill health among non-working women does not rise steeply with age, and never differs so markedly from the prevalence among working women. These differences reflect the lower participation rate of women, and the relatively lower importance of ill health as a reason for non-participation. This adds further support to the hypothesis that the effect of health on employment is an important determinant of the life cycle variation in the income gradient of health, which is more pronounced for males than for females.

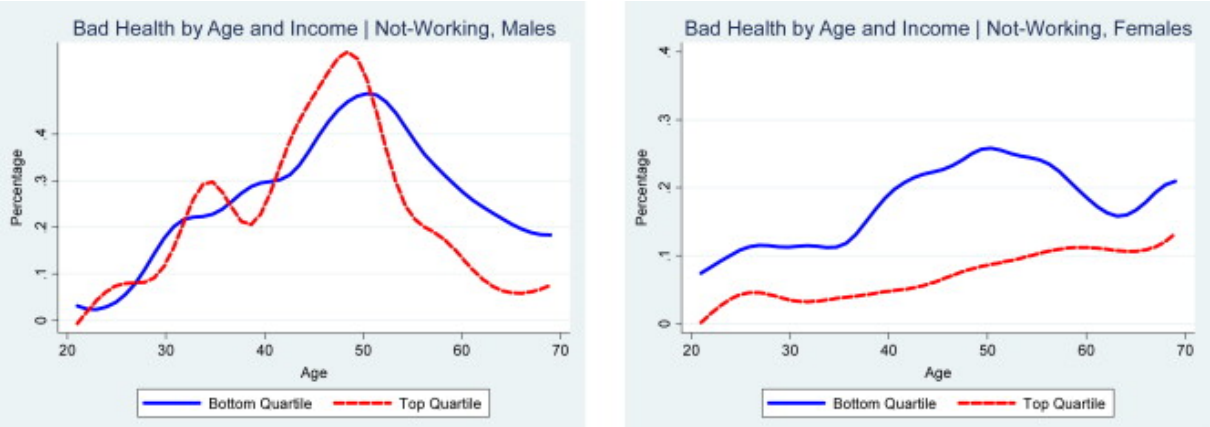
We now plot the health profiles by income group separately for workers and non-workers. Health differences by income are much smaller among workers (Figure 3.5) and the life cycle profiles much flatter than they are for the whole sample (3.1). Apparently, labor force participation accounts for a large part of the health-income association and its variation with age. For non-working females (Figure 3.6), the differences are still apparent, while for males they are not. Given that there are very few non-working males in the top income quartile, this figure should be interpreted with caution. While this all is consistent with some impact of income (or a correlated socioeconomic characteristic) on health after controlling for employment status, it may also be true that low-income non-workers are more likely to be inactive for health reasons.

Figure 3.5: Self-reported health among workers by age, split by income and gender



Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

Figure 3.6: Self-reported health among non-workers by age, split by income and gender

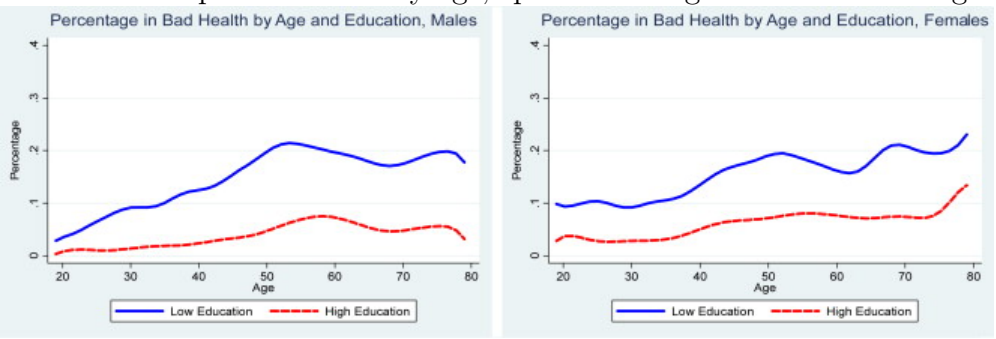


Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

3.4.3 Is the picture different by education or occupation?

Figure 3.7 presents the prevalence of reported bad health for the low and high education categories. For males, a widening of the education gradient up to late middle age is immediately apparent. Compared to the income gradient in Figure 3.1, the magnitude of the education gradient is larger in early adulthood (probably because education is a better indicator of SES in this age range) and smaller in late middle age. Hence, the widening of the education gradient over the years of normal working age is less pronounced than that for the income gradient. Further, while there is some narrowing of the education gradient beyond the age of 60, this is much less evident than is the case for the income gradient. These differences are consistent with the evidence from the US (Smith, 2005a). A plausible explanation is that, unlike income, education is not responsive to health changes in adulthood. The income gradient may strengthen with age, as health shocks increasingly lead to labor market withdrawal and a drop in income. But there is no such mechanism to drive the dynamics of the education-health relationship. For females, the picture is quite different. Already at young ages there is a very strong education gradient in reported health, which remains stable throughout the years of younger adulthood before increasing moderately in middle age.

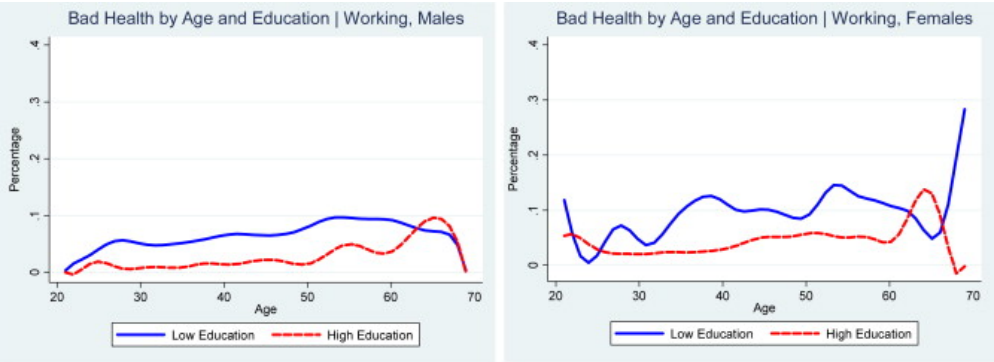
Figure 3.7: Self-reported health by age, split according to education and gender



Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

To better gauge the relative importance of work status versus socioeconomic status in determining the life cycle profile of health, Figure 3.8 presents the prevalence of bad health by education for those currently working. This removes most of the life cycle variation in health, and about half of the difference across education categories, suggesting that the increasing gradient observed in Figure 3.7 may be attributable to the less well educated being more vulnerable to health conditions that interfere with employment as they age. It could also be that the low educated are more likely to withdraw from the labor force as they age, and report poor health as justification for this.

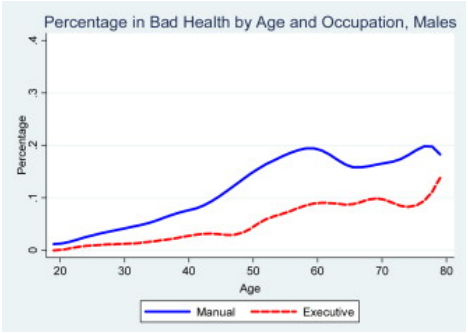
Figure 3.8: Self-reported health among workers by age, split by education and gender



Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

Occupation is less predetermined than education, but is more so than income, offering another opportunity to examine the life cycle pattern of the SES-health gradient with limited potential for health to impact on SES. We restrict our attention to males, for whom the manual vs. non-manual distinction is likely to be more crucial in identifying work of different levels of physical effort. There is a widening of the occupational gradient in health up to late middle age (Figure 3.9), which is comparable to that observed for education but, as anticipated, is less stark than that observed for the income gradient. In early adulthood, differences in health by occupation are evident but not marked. Given that little time has passed at this stage of the life cycle for occupation to exert an influence on health, the observed differences presumably derive from earlier childhood experiences that impact on both occupational choice and health. But the steeper health trajectories experienced by manual workers mean that, by age 45, 10% of them report bad health- whereas non-manual workers reached this prevalence of ill health only at the age of 60.

Figure 3.9: Self-reported health by age, split according to occupation



Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

3.5 Discussion

Our analysis demonstrates that in the Netherlands socioeconomic differences in health widen until late middle-age before narrowing in old-age. This life cycle profile in the socioeconomic-health gradient is remarkably similar to that observed in the US despite the stark differences that exist between the countries in characteristics such as health system finance and organisation, income inequality, social protection and disability insurance, which would be expected to influence socioeconomic inequalities in health and their life cycle evolution. It seems that countries with very different health and welfare systems end up displaying not quite identical, but nonetheless remarkably similar dynamic associations between health and socioeconomic status. This suggests that fundamental mechanisms linking age, health, income and education that are relatively unresponsive to policy parameters, at least within the range observed in Western, high-income economies, may be responsible for these relationships.

On the basis of our analysis and the more robust causal evidence available in the literature, two mechanisms are beginning to emerge as particularly important in understanding socioeconomic differences in health. First, there is a large effect of health on income that operates through employment and grows with age until voluntary retirement begins to dominate health as the main reason for labor force withdrawal (Case and Deaton, 2005; Smith, 2005a). This is not to say that all socioeconomic-related health inequality is a reflection of the impact of health on SES. Far from it. The second mechanism is a non-negligible life-long effect of education on health. While ill-health can impinge on investments in education, studies that take account of this and other sources of endogeneity still find that education exerts a causal impact on health (Lynch, 2003; Lleras-Muney, 2005; Oreopoulos, 2006; Smith, 2007; Silles, 2009). In contrast, although large health differences across income groups are visible, these do not derive from a causal impact of income on health (Adams *et al.*, 2003; Contoyannis *et al.*, 2004; Smith, 2005a; Frijters *et al.*, 2005).

It is hardly surprising that health-related exit from the labor force is an important causal mechanism linking health and socioeconomic status at middle age in the US, a country with relatively limited social protection mechanisms. It is perhaps more surprising that the same holds for a country like the Netherlands, known for its generous social protection, disability insurance and retirement schemes. But this generosity can itself increase the observed correlation between income and health, if individuals with chronic health problems are encouraged to withdraw from the labor force, and given that replacement rates, although generous, are less than 100%. The strong impact of ill-health

on employment and consequently income in pre-retirement years and the absence of such an effect post-retirement would seem to be an important explanation for the widening of observed health differences by income in middle age and their narrowing in old age. It is not clear that this pattern should provoke any policy response given that a replacement rate of less than 100% is desirable in the presence of moral hazard.

We should emphasize that we examined differences in health by income quartile and these tell us nothing of the magnitude of any impact of health on income, which we would expect to be relatively small in the Netherlands due the generosity of disability insurance. Kapteyn *et al.* (2007) demonstrated that more than half of the much higher rate of reported disability in the Dutch working population compared to that of the US could be explained by differential reporting of specific health conditions. This suggests that the generosity of Dutch disability insurance and the readiness for this to be used as a route out of work into retirement has exerted a considerable bias on the income-health gradient and its life cycle profile. The apparent improvement in health beyond age 55 among the poorest individuals (Figure 3.1) may simply reflect the reduced incentive to report poor health once individuals reach an age at which they can legitimately withdraw from the labor force for non-health reasons.

The observed decline in health inequalities in old age is partly an inevitable consequence of the ageing process with biological determinants of health dominating socioeconomic ones (House *et al.*, 1994), but it probably also reflects the survival of only the fittest members of the more disadvantaged socioeconomic groups. The narrowing of observed health inequalities in old age is partly the result of socioeconomic disparities in mortality at younger ages. However, the fact that the education disparity in health narrows much less in old age than that for income suggests that selective mortality is not the most important explanation of this trend.

While the direction of causality in the education-health relationship appears, on balance, to be the reverse of that in the income-health gradient, the two relationships can be integrated into a unified theory of cumulative advantage over the life course. This would start with early-life conditions and parental background affecting education. We have not reviewed the literature on the lifetime economic and health consequences of early childhood conditions here, but the evidence base to support causal effects is growing (Barker, 1995; Case *et al.*, 2002; 2005; Currie and Stabile, 2003; Currie *et al.*, 2007; van den Berg *et al.*, 2006; van den Berg and Lindeboom, 2007). Education strongly influences the choice of occupation, and thereby the extent to which an individual's physical health determines his productivity, but also the health consequences of work itself. Someone with little or no education often has no option but to enter into heavy manual labor, which is likely

to exert a greater toll on his health; that person is also more likely to be laid-off once health problems begin to impede productivity. So, the health-related earnings losses that are partly responsible for the income-health gradient are themselves, to some extent, the result of socioeconomic differences in lifetime opportunities and their impact on health. It is unlikely that there is a single explanation for the observed life cycle pattern of the socioeconomic gradient in health. It is not simply the consequence of income reductions from health-related work loss, nor moral hazard and justification bias responses to social protection against these losses. Nor is it all a the result of cumulative disadvantage. The reality is likely to be one of many mechanisms interacting.

Overturning these processes of cumulative advantage represents a major challenge even to countries as egalitarian as the Netherlands. The potential rewards are, however, immense. Getting closer to a goal of equality of opportunity could raise national wealth by breaking the connection between low education, poor health and early labor force exit. There could also be substantial savings on age-specific health care expenditures to offset those arising from extended longevity. Of course, breaking the cycle requires identifying the specific causal mechanisms that link education with later life health outcomes but, at present, there is very little robust evidence on this. Further, not all educational disparities in health derive from inequality of opportunity in the sphere of health choices. Some must reflect ill-health in childhood and adolescence constraining choices over education. Besides ensuring adequate access to schooling for less healthy children, education policy can do little to address this source of health disparities.

Policies aimed at improving health conditions at work (particularly for those in low-skilled occupations) are potentially important in preventing health related labor force exits, thereby extending working lives and maintaining the financial sustainability of pension systems. Given the moral hazard effects that are inevitably created by financial protection against health related loss of employment, which has been a particular problem in the Netherlands with as much as 10% of the labor force on disability insurance at the end of the 1980s, it seems crucial to enact policies in the workplace that can both prevent the development of health problems and reduce the impact of those problems on work capacity by providing appropriate support to partially disabled workers and those with a diagnosed condition. But there are obvious limitations on the extent to which workplace interventions can reduce health inequalities. They can do little to halt health deterioration that derives from early life experiences. Nor can they do much to correct differences in health by income that derive from education related differences in life style.

Appendix

Tables

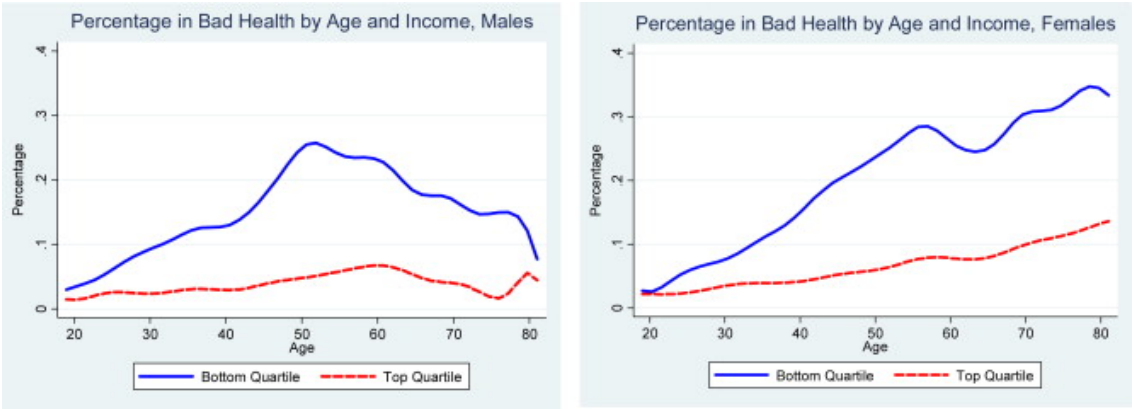
Table 3.1: Descriptive Statistics

Self-Reported Health	**Very Good**	**Good**	**Fair**	**Sometimes good/ Sometimes bad**	**Bad**						
N=121,232	31,455	64,845	15,230	6,720	2,982						
	(25.95%)	(53.49%)	(12.56%)	(5.54%)	(2.46%)						
Disability	**Hearing (1)**	**Hearing (2)**	**Reading (3)**	**Vision (4)**	**Chewing (5)**	**Carrying (6)**	**Bending (7)**	**Walking (8)**			
N=82,067	1,953	259	3,111	931	3,776	4,668	2,937	2,922			
	(2.38%)	(0.32%)	(3.79%)	(1.13%)	(4.60%)	(5.71%)	(3.58%)	(3.56%)			
Severe Work Disability	**No**	**Yes**									
N=85,680	79,244	6,436									
	(92.49%)	(7.51%)									
Income	**<8**	**8-10**	**10-12**	**12-13**	**13-14**	**14-16**	**16-18**	**18-20**	**20-25**	**25-30**	**>30**
In thousands Euro	8,338	7,268	6,630	5,932	5,644	7,470	6,524	5,738	7,663	4,763	7,592
N=73,562	(11.33%)	(9.88%)	(9.01%)	(8.06%)	(7.67%)	(10.15%)	(8.87%)	(7.80%)	(10.42%)	(6.47%)	(10.32%)
Labor Force Status	**Working**	**Not-Working**									
N=81,534	41,490	40,044									
	(50.89%)	(49.11%)									
Education	**Primary**	**Secondary**	**Higher**								
N=116,497	24,186	71,699	20,612								
	(20.76%)	(61.55%)	(17.69%)								
Occupation	**Manual**	**Executive**	**Other**								
N=57,805	14,918	15,308	27,579								
	(25.81%)	(26.48%)	(47.71%)								

Notes: The OECD disability questions that are asked include (1) Hear normal conversation with three or four other persons, (2) Hear normal conversation with another, (3) Read ordinary newsprint, (4) See the face of someone from four meters, (5) Bite and chew on hard foods, (6) Carry an object of five kilos for ten meters, (7) Bend down (when standing) and pick up shoe, (8) Walk for 400 meters without resting. The work disability variable we use concerns only the last three OECD disability questions and indicates whether an individual has at least one severe disability in these three domains. Labor Force Status and Occupation only measured until the 1997 wave and until age 70. The "Other" category among the occupations includes administrative, service, and commercial occupations.

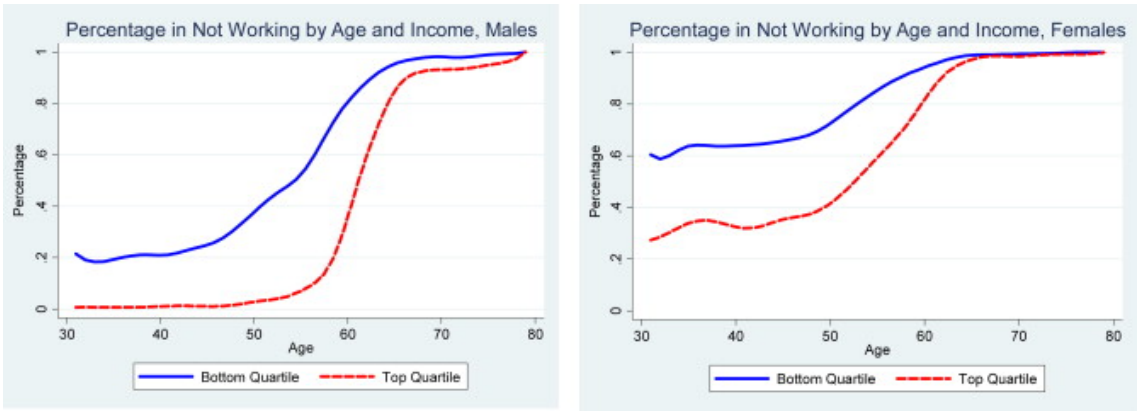
Figures

Figure 3.10: Self-reported health over the life cycle, split according to income and gender (cohort corrected)



Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

Figure 3.11: Prevalence of population working over the life cycle, by gender and income quartile



Source: Authors' calculations for CBS Health Interview Surveys, 1983-2000. Sample weights applied.

Part II

Causal Pathways

Chapter 4

Long Run Returns to Education: Does Schooling Lead to an Extended Old Age?

While there is no doubt that health is strongly correlated with education, whether schooling exerts a causal impact on health is not yet firmly established. We exploit Dutch compulsory schooling laws in a Regression Discontinuity Design applied to linked data from health surveys, tax files and the mortality register to estimate the causal effect of education on mortality. The reform provides a powerful instrument, significantly raising years of schooling, which, in turn, has a large and significant effect on mortality even in old age. An extra year of schooling is estimated to reduce the probability of dying between ages of 81 and 88 by 2-3 percentage points relative to a baseline of 50 percent.

This chapter is based upon:

Van Kippersluis, H., O. O'Donnell, and E. van Doorslaer (2009), "Long Run Returns to Education: Does Schooling lead to an Extended Old Age?", Tinbergen Institute Discussion Paper 09-037/3.

4.1 Introduction

Inequalities in health and life expectancy by education are striking. For example, in the Netherlands individuals with a university or college degree live, on average, 6 to 7 years longer than those who finished only primary school. The difference in life expectancy in good health is as much as 16 to 19 years (CBS, 2008). Such disparities in health and mortality by education have been documented for many countries (Grossman and Kaestner, 1997; Smith and Kington, 1997; Mackenbach *et al.*, 1997; Cutler and Lleras-Muney, 2008). Indeed, Michael Grossman has claimed that *"years of formal schooling completed is the most important correlate of good health"* (Grossman, 2003, p.32). Yet, very few studies have found robust evidence of a causal impact of education on health. The correlation could also stem from childhood ill-health constraining educational attainment (Perri, 1984; Behrman and Rosenzweig, 2004; Case *et al.*, 2005) and confounding factors such as ability and time preference (Fuchs, 1982; Auld and Sidhu, 2005; Deary, 2008). Establishing whether education causally impacts on health is essential to the formation and evaluation of education and health policies.

The purpose of this paper is to establish whether education has a causal impact on mortality using exogenous variation in education that comes from a compulsory schooling law in the Netherlands. This reform, which increased the educational attainment of the population suddenly and strongly, provides a valuable instrument within a regression discontinuity design. We observe a very large sample of individuals between the ages of 80 and 88. This makes it possible to investigate whether education has long run effects on health that result in an extended life. If education effects are present at this age, then it is highly likely that they exist at younger ages, such that our estimate can be interpreted as a lower bound on the total impact of education on mortality.

Lleras-Muney (2005), exploiting changes in compulsory schooling laws in more than 30 US states, estimates that an additional year of schooling reduces the 10 year probability of dying by at least 3.6 percentage points. However, the analysis is based on only an approximation to mortality derived from the change in cohort size between subsequent censuses. Further, Mazumder (2008) demonstrates that the magnitude and the significance of the effect are not robust to the inclusion of state-specific cohort trends. Clark and Royer (2007) measure mortality somewhat more directly than Lleras-Muney (2005) and Mazumder (2008) from the number of deaths divided by the number of births of a cohort. They find that the 1947 extension to the school leaving age in the UK had a strong impact on educational attainment, but no significant effect on mortality. Albouy and Lequien (2008) exploit two compulsory schooling reforms in France in 1923 and 1953.

They do not find a significant effect of years of schooling on mortality, which is observed directly from micro data, but neither did the reform have a very strong impact on years of schooling.

While the weight of existing evidence does not support an effect of education on mortality (Clark and Royer, 2007; Mazumder, 2008; Albouy and Lequien, 2008), some studies (but not all) using compulsory schooling laws as a source of exogenous variation in education find an effect on self reported health outcomes (Spasojevic, 2003; Oreopoulos, 2006; Mazumder, 2008; Silles, 2009) and hospitalizations (Arendt, 2008). An obvious weakness of evidence based on self-reported health is that reporting thresholds may vary with education (e.g. Bago d'Uva *et al.*, 2008) and, if this is the case, the resulting bias would not be corrected by instrumenting education[1].

The main weakness of most of the analyses of mortality, with the notable exception of Albouy and Lequien (2008), is that they rely on approximate measures of mortality at the cohort level rather than directly observed survival at the individual level. We observe a very large number of individuals from linked Dutch administrative and survey data for which a mortality follow-up is available. Given the strong instrument, provided by the compulsory schooling reform, and tremendous power, generated by the sample size and micro data, it is quite unlikely that we will fail to detect an effect of education on mortality should one exist in the population. A potential limitation of the previous literature is that it is concerned with the health returns to additional years of schooling, ignoring the nature and quality of education (Feinstein *et al.*, 2006; Cutler *et al.*, 2008). In addition to estimating the impact of an extra year of schooling on mortality, we estimate the health returns to completion of (the Dutch equivalent to) high school and investigate whether this has a discrete beneficial effect beyond that predicted linearly from the equivalent additional years of schooling.

We find that education significantly lowers mortality. An additional year of schooling reduces the probability of dying between the ages of 81 and 88 (inclusive) by more than 2 percentage points, relative to a baseline probability of 50 percent The instrumental variable (IV) estimate is larger than the ordinary least squares (OLS) estimate, indicating downward bias in estimates that fail to take account of endogeneity. A significant impact of education on mortality is found only with the larger one of our samples, suggesting that lack of statistical power is a potential explanation for the failure of most previous studies to find an effect. Finally, high school graduation reduces the probability of dying between 81 and 88 by as much as 17 to 26 percentage points but this does not appear to

[1]If individuals with more schooling tend to understate their health, as some of the evidence suggests (Bago d'Uva *et al.*, 2008), then impact of education on health will be underestimated.

due to any sheepskin effects of finishing high school on mortality beyond that predicted linearly by additional years of schooling.

The paper is organized as follows. Section 4.2 provides background information on the Dutch reforms that enable us to estimate the causal impact of education on mortality. The data and methods are discussed in section 4.3. The main results are presented in section 4.4. Section 4.5 investigates whether there are additional gains in survival from high school graduation over and above those implied by the years of schooling this entails. Section 4.6 considers the lessons learnt and the limitations of our study.

4.2 Compulsory Schooling in the Netherlands

The first compulsory schooling law, mandating 6 years of education, was introduced in the Netherlands in 1900 (Dodde, 2000). In the years before World War I several attempts were made to increase the compulsory years of schooling to 7, and even 8. Due to the war, a law raising the minimum school leaving age came into force only from 1st of January 1922. The law required children to be enrolled in school at least from their 7th birthday until they had completed at least 7 years of schooling, or had reached the age of 14 (Hentzen, 1928, p. 4). However, since some schools did not have the resources to offer the 7th year of schooling, the law was never enforced (Hentzen, 1932). In 1924, the number of years of compulsory schooling was officially reversed to 6 and the increase postponed until 1930. But an improvement in the economy prompted parliament to ask for the rise of the minimum school leaving age to be brought forward. The government agreed in 1927 and 7 years of schooling became compulsory starting from the 1st of July 1928 (Hentzen, 1932; De Graaf, 2000).

Yet, since the law only had an effect at the transition of the school year (Hentzen, 1932, p. 185), only at the start of new school year the new compulsory schooling law came fully into effect for all schools (CBS, 1931, p. 19; Mandemakers, 1996, p. 55). In the cities schools usually started in August or September and on the countryside in March or April, while on the countryside it was more convenient to have children free from duty already before the summer (CBS, 1931, p. 36).

Despite the fact that the law required children to be enrolled in school only from their 7th birthday on, the large majority of children started school long before that age. After 1920 the age at which children start school became more and more between $5\frac{1}{2}$ and 6, because in many cases it was not allowed to enter the class during the year (Sterringa, 1934, p. 88). Children born in 1916 who started school at age 6 will have completed their duties under the old regime before the new law was enforced (Sterringa, 1934, p.

113). This cohort was not affected by the reform. In contrast, children born in 1917 were at most 11 on the 1st of July 1928, and could not have met the conditions to leave school under the pre-1928 legal regime. Therefore, the cohort born in 1917 was the first to be affected by the 1928 reform and to be forced to complete 7, rather than 6, years of schooling[2]. We put the threshold therefore at the 1st of January 1917[3]. Importantly, since most primary schools had only 6 grades, the 1928 reform induced — as we will show later — a large proportion of pupils to enter secondary school.

Compulsory schooling was raised further to eight years in 1942 under the German occupation with the aim of promoting the German language. A lack of competent teachers and materials, and resistence from the population, meant that both the quality of the extra education and compliance with the law were very questionable (Meijsen, 1976; HSG[4], 1946/1947). After the war, a new law confirmed the increase in the minimum years of schooling to eight, but deferred its enforcement to January 1st 1950 given high rates of school absence and the unpopularity of the war-time increase, circumvention of which had become almost a heroic deed (HTK[5], 1946/1947). Later increases in the minimum school leaving age in 1969, 1975 and 1985 are too recent to investigate their potential mortality effects.

The introduction of the seventh and the eighth year of compulsory schooling was a major breakthrough in the Dutch educational system, since it forced pupils to go beyond primary school and start secondary schooling. However, since between 1938 and 1949 the percentage that attended an additional level of education after primary school increased from 50 to 73% (CBS, 1951), the target population for the 1950 reform was much smaller than for the initial reform in 1928. We have confirmed that the 1950 reform did induce some individuals to obtain more schooling, but the effect is not large and so does not provide a strong instrument for education. In the analysis we restrict attention to the 1928 reform, which, as will become apparent, had a strong impact on educational attainment.

[2]It is possible that some individuals born in 1916 that started school at age 7 were affected by the reform, and some individuals born in 1917 that started school at age 5 were allowed to drop out, but this is plausibly a small fraction.

[3]As noted, some schools started their school years in April and some in September, such that it is impossible to sharply determine the threshold. Yet, with a Fuzzy Regression Discontinuity Design (RDD) this is not vital since the only requirement is that the probability of receiving treatment jumps discontinuously at the threshold, which is obviously the case as we will show later.

[4]HSG refers to notes of both the Lower House (House of Representatives) and the Upper House (Senate)

[5]HTK refers to notes of the Lower House (House of Representatives)

4.3 Data and Methods

4.3.1 Data

Our data are linked survey and administrative records from Statistics Netherlands. We use the annual cross-sectional general household survey (POLS) 1997-2005, the tax records (RIO) for 1998, and the Cause-of-Death register for 1998 until 2005 inclusive. All these files are linked to the Dutch Municipality Register (GBA), which covers, inter alia, year of birth, sex, province, and ethnicity.

The POLS samples a representative cross-section of the non-institutionalised Dutch population ranging from around 10,000 to 80,000 respondents per year[6]. It collects extensive information on demographic and socioeconomic characteristics. The respondent's education is recorded by two variables: the highest level followed and that finished on the standard Dutch categorization (Standaard Onderwijs Indeling (SOI) 1998)[7]. The SOI is very close to the International Classification of Education (ISCED) and is easily converted into years of schooling following standard guidelines (SHARE, 2007)[8]. In the relatively few cases that individuals reported to have followed a higher level than they finished, we take the average of the corresponding years[9]. Through linkage with the Cause-of-Death register we are able to observe death and its cause for all POLS respondents who died between 1998 and 2005.

The RIO is a huge administrative tax-register covering one third of the Dutch population, i.e. around five million observations per year. Apart from detailed income information, it also contains demographics. By linking the RIO to the Cause-of-Death register we observe, again, mortality for all individuals. Unfortunately, education information is not available in the RIO, but, as will be explained below, it is still possible to combine estimates from the linked RIO-death register with those from the POLS to obtain IV estimates of the impact of education on mortality.

[6]Specifically 34,439 in 1997, 80,789 in 1998, 42,605 in 1999, 37,482 in 2000, 24,231 in 2001, 22,259 in 2002, 25,163 in 2003, 21,706 in 2004, and 10,378 in 2005.

[7]SOI is missing in the 2003 wave of the POLS. However, there is a highly similar educational variable available in that year, which is redefined into SOI and ISCED.

[8]The Dutch SOI consists of 7 levels of education: toddler school, primary education (6 years), lower vocational secondary education (10 years), higher general secondary education (13 years, similar to high school), first phase higher education (15 years, intermediary vocational education), second phase higher education (16 years, higher vocational education) and third phase higher education (17 years, university education).

[9]We drop a very small number of individuals who report to have followed a lower level than the one they report finishing. Also, a few individuals who claimed to have followed a level to which they do not have access given their highest level completed are disregarded.

The CBS Linked Data are unique in the context of data that have been used previously to estimate the impact of education on health in the sense that they provide a mortality follow-up of both a sample survey (POLS, 1997-2005) and a very large administrative database (RIO, 1998). So, individual mortality is observed, and with the POLS we observe both education and mortality for the same sample observations. The mortality record provides 8 years of follow-up by cause of death. Given that the cohorts affected by the reform we study were born around 1917, we observe these individuals in their 80s. Pre-sample selective mortality might be important, especially if education does reduce the risk of an early death. But in this case, our estimates will provide a lower bound on the impact of education on life expectancy. We estimate the effect of years of schooling on the probability of dying between 80 and 88 (inclusive), conditional on being alive at age 80[10]. This allows us to examine whether there are long-term returns to education on life expectancy.

4.3.2 Identification strategy and estimation

We exploit the 1928 compulsory schooling law as an instrument for education within the framework of a Regression Discontinuity Design (RDD) (Thistlethwaite and Campbell, 1960; Trochim, 1984; Hahn *et al.*, 2001; Lee and Card, 2008; Imbens and Lemieux, 2008; Van der Klaauw, 2008; Lee and Lemieux, 2009). All analyses are done separately for males and females. Year of birth (cohort) determines whether the individual is exposed to the reform, so sorting around the threshold is absent. Furthermore, not all individuals exposed to the reform are induced by it to change their education. We therefore have a discrete running-variable Fuzzy Regression Discontinuity set-up (Lee and Card, 2008) and the appropriate estimator is parametric Two Stage Least Squares (2SLS), in which years of education is instrumented by the reform. Under the standard assumptions of RDD (Hahn *et al.* 2001; Van der Klaauw, 2002), this provides an estimate of the Local Average Treatment Effect (LATE) of an additional year of education on mortality for individuals in the cohort exposed to the reform that were induced to stay at school.

[10] An estimated 27 percent of Dutch males born in 1917 and 51 percent of females is still alive at age 80 and, for this cohort, life expectancy at birth was around 60 for males and 68 for females (Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) www.mortality.org)

In the first stage, we regress years of education on a flexible polynomial in cohort[11], the indicator of whether the individual is exposed to the reform, and wave dummies[12]. The usual concern about the strength of the instrument applies here. If the instrument is weakly correlated with the possibly endogenous variable, then the IV estimator will be biased in the same direction as OLS (Bound *et al.*, 1995; Staiger and Stock, 1997). F-tests on the strength of the instrument are reported. Moreover, the flexibility and goodness-of-fit of the polynomials are tested using the G-test suggested by Lee and Card (2008). We allow for non-random specification error in the polynomial by computing robust standard errors clustered at the cohort level (Lee and Card, 2008).

A reduced form linear probability model of the binary mortality variable on the same variables as in the first stage is also estimated. Again a G-test on the flexibility and goodness-of-fit of the polynomial is performed (Lee and Card, 2008). For the models that pass the relevant tests, the full 2SLS is estimated.

A drawback of using the POLS in this analysis is the potential lack of power for an outcome such as mortality. With around five million observations, this is not a problem with the RIO data but, unfortunately, this dataset does not include education. However, since in the exactly identified case the coefficient of interest in 2SLS estimation can be written as the ratio of two reduced form coefficients, it is possible to perform these two reduced form regressions using separate samples from the same population. This technique has been labeled Two Sample Two Stage Least Squares (TS2SLS) (Angrist and Krueger, 1992; Arellano and Meghir, 1992; Inoue and Solon, 2009). The standard error of the TS2SLS coefficient is obtained by the delta method (Devereux and Hart, 2008).

[11]Following Lee and Lemieux (2009), we allow the polynomials to differ on either side of the threshold since otherwise the information on one side of the threshold is used in the estimation of the trend on the other side, which is against the spirit of a RDD. Eight models with different degrees of flexibility of the polynomial are estimated — linear, quadratic, cubic and quartic in cohort with and without interactions with the reform dummy. We present results only for the linear and quadratic models (with and without interactions) as these prove to be sufficiently flexible.

[12]Wave dummies are entered in the analysis of the POLS data only to correct for potential bias due to the differential mortality probabilities within the observation period across the cross-sections. That is, someone observed in the 1997 POLS is more likely to have died by 2005 than someone in the 2004 POLS. Since the wave dummies are included in the mortality regression, they are also included in the years of schooling regression.

4.4 Main Results

4.4.1 OLS estimates

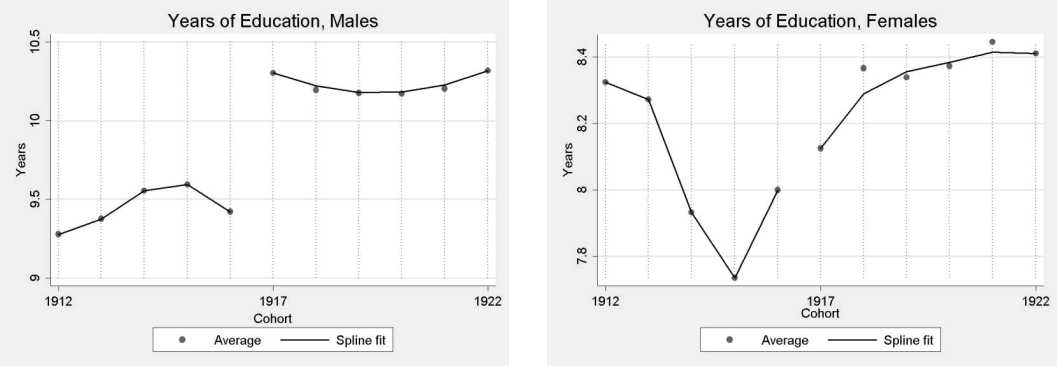
Table 4.1 gives OLS estimates of the impact of an additional year of schooling on the probability of dying in the 1998-2005 period. For males, the reduction in the probability of dying is 1.4 percentage points per additional year of schooling, and this falls to 1.1 points, but still remains strongly significant, when both a cohort trend (quartic polynomial) and covariates (marital status, ethnicity, province, and city size) are controlled for. For females, there is also a significant but slightly smaller effect of 0.8-0.9 percentage points.

There is clearly a negative association between education and mortality in this population. Our aim in the remainder of the analysis is to establish whether there is a causal component in this correlation.

4.4.2 First stage results

Figure 4.1 show years of education completed by cohort for males (left) and females (right), where a non-parametric lowess smoother is estimated on both sides of the reform threshold. For males, there is a very clear and large discontinuity in educational attainment at the first cohort that is fully affected by the 1928 reform — the 1917 birth-year cohort. The average years of schooling increases by more than 0.8 at the threshold. For females, there is no such discontinuity, suggesting that the 1928 reform did not have a strong impact on the schooling of females and does not provide a useful instrument. This was already recognized by Statistics Netherlands in 1931 who observed that *"For boys apparently secondary education was considered more necessary than for girls"* (CBS, 1931, p. 50). Given this, we do not pursue the estimation of a causal effect for women.

Figure 4.1: Years of education by cohort and gender



Notes: 1912-1922 birth-year cohorts, males (left) and females (right), POLS 1997-2005.

Table 4.2 shows linear regression estimates of variation in years of education in relation to the reform indicator (1 if 1917 cohort or later), controlling for different specifications of cohort trends and for wave dummies. In all specifications the magnitude of the coefficient of the reform indicator lies in the range of 0.60-1, indicating that the 1928 reform raised the average years of schooling by between seven months and one year. The reform indicator is strongly statistically significant in all models, which is confirmed by the extremely strong robust F-tests of the instrument, passing all criteria proposed in the literature (Bound *et al.*, 1995; Staiger and Stock, 1997; Stock and Yogo, 2002)[13]. Finally, the G-test indicates that all polynomials are sufficiently flexible. Given the spirit of RDD (Lee and Lemieux, 2009) and the lower AIC values of the more flexible models 3 and 4, these will be the focus of our discussion.

We now check robustness of the first-stage results to the choice of bandwidth around the threshold (see e.g. Lee and Lemieux, 2009). The base case uses five birth-year cohorts before and after the reform to estimate a reliable polynomial through cohorts. Identification relies on the cohorts being interchangeable, and this will only hold if the cohorts are sufficiently close to each other. Yet, focusing attention on just a few cohorts around the threshold reduces power. Therefore, the estimation is repeated using ten and three cohorts on either side of the 1917 threshold — see table 4.3. Overall, using ten cohorts, the coefficient of the reform is very similar to our base case, except for model 4 where the coefficient sharply declines and becomes more consistent with estimates from the other models. The reform indicator is still highly significant, with the F-tests showing no evidence of a weak instrument (Bound *et al.*, 1995; Staiger and Stock, 1997; Stock and Yogo, 2002). G-tests are not presented but they give no indication that the polynomials are insufficiently flexible. The reform dummy remains strongly significant when using only three cohorts on either side of the threshold, and although the magnitude of the coefficient increases somewhat, it remains in the same range. With only three cohorts the G-tests (not shown) are less satisfactory, but this is to be expected given the limited number of observations from which to fit the polynomial.

Overall, changing the bandwidth does not affect the significance of the instrument and the magnitude of its effect is quite robust. This supports our use of observations further away from the threshold, but as an additional check on whether this introduces bias,

[13]The Stock and Yogo critical values (Stock and Yogo, 2002) can only give approximate confirmation of the strength of the instruments since they are computed for an iid error model and here we have made the F-tests robust to heterokesdasticity and clustering at the cohort level. Baum et al. (2007) suggest relying on the Staiger and Stock (1997) rule of thumb that the (robust) F-statistic should exceed 10 for weak instruments not to be considered a problem. This is satisfied for all models.

we examine robustness to controlling for covariates (marital status, ethnicity, province and city size) that may possibly differ across the five-year span on either side of the reform threshold. While covariates can safely be excluded with a valid RDD, when using observations further away from the threshold covariates might be used to correct for possible differences in underlying characteristics (Lee, 2006). The results given in the third panel of Table 4.3 confirm that the coefficient of the reform dummy hardly changes relative to the baseline for each specification, and statistical significance remains strong.

A final robustness check examines whether potential always-takers bias the estimated effect of the reform on educational attainment. The reform affected the individuals who would have dropped out at the age of 12 prior to the 1928 law but became legally required to stay in school until at least their 13th birthday. Individuals who would have continued school irrespective of the reform, the so-called 'always-takers', are of no use in our analysis, and only cause downward bias in the estimated effects. Since it might be argued that individuals who finished university or higher vocational education are likely to have continued school anyway, irrespective of the law change, we check robustness to the exclusion of these individuals. Comparison of the estimates in the fourth panel of table 4.3 with those in table 4.2 reveals that the magnitude and statistical significance of the coefficients hardly changes, so that potential always-takers do not appear to bias the results[14].

We conclude that the 1928 reform had a sudden and strong positive impact on years of schooling for males and so provides a strong instrument for education. The magnitude and statistical significance of the effect is robust to several checks, so that we can safely move on to the second stage.

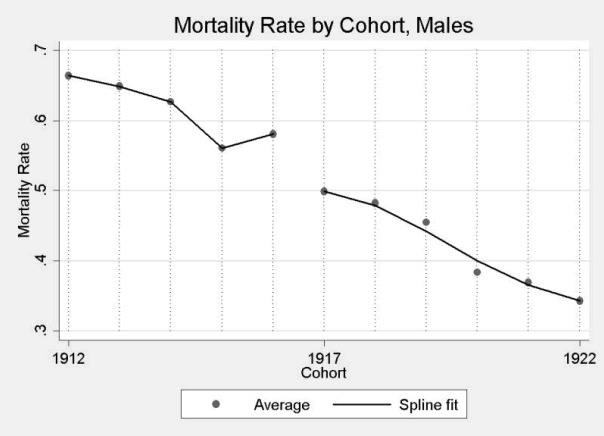
4.4.3 Two Stage Least Squares (2SLS) estimates

Figure 4.2, showing the probability of dying between 1998 and 2005 for male cohorts born between 1912 and 1922, reveals a small downward discontinuity at the 1917 threshold. The reduced form estimate is negative in three of the four models presented in table 4.4, but the effect is never significant. The G-tests indicate that the models are sufficiently flexible. Controlling for covariates, changing the bandwidth, and excluding potential

[14]A necessary condition for the validity of excluding these individuals as always-takers is that the reform did not have an impact on the probability of attaining a university degree or higher vocational qualification. It turns out that indeed we cannot reject a zero effect of the reform on these levels of educational (results available upon request).

always-takers changes the magnitude of the coefficient on the reform dummy marginally in the models, but it always remains insignificant[15].

Figure 4.2: Mortality rate, percentage that died in the period 1998-2005, by cohort (POLS)



Notes: Cohorts 1912-1922, Males, POLS 1997-2005.

Since the first stage results for education support the strength of the instrument, and the reduced form results for mortality confirm that the polynomials provide a sufficiently flexible specification, we proceed to Two- Stage Least Squares estimates, which are given in table 4.5. The point estimates of the effect of years of education on mortality are all negative and in the order of magnitude of 0.02-0.03, except for model 1. However, the effect is never close to statistical significance and, on the basis of these estimates, we cannot reject the hypothesis that education has no impact on mortality.

4.4.4 Two Sample Two Stage Least Squares (TS2SLS) estimates

Despite the large discontinuity in educational attainment produced by the 1928 reform, the POLS data provide no evidence that years of schooling has a significant causal impact on mortality. However, this could simply be due to a lack of power if there are an insufficient number of observations in the relevant cohorts. Using the much larger RIO sample to estimate the mortality effect allows investigation of this possibility.

TS2SLS is consistent under the assumption that the two samples are drawn from the same underlying population (Angrist and Krueger, 1992; Arellano and Meghir, 1992). A potential concern in our case is that the first stage is estimated using the POLS cross-sections for 1997-2005, while the second stage is estimated using the RIO 1998 with a

[15]Results available upon request.

mortality follow-up, so that, strictly speaking, the underlying populations are not identical. More specifically, the RIO linked mortality register data provides information on mortality between 1998 and 2005 for a sample of individuals alive in 1998, while individuals who have died or been institutionalized subsequent to 1998 are not available for the POLS cross-section samples in the period 1999-2005. The first stage estimate may therefore be biased if survivors are affected differently by the reform than decedents. To investigate this, the first stage is estimated separately for survivors and decedents in the period 1998-2005. Moreover, an additional check is performed by estimating the first stage using only the POLS 1998 observations. Results are shown in table 4.3, where we also present estimates using all 1997-2005 POLS observations (as in table 4.2) without wave dummies since these are not necessary in the reduced form for mortality using the follow-up to the 1998 RIO (see footnote 10). Excluding the wave dummies (panel 5) has little or no impact on the estimates. Although significance is greater for decedents, the point estimates of the effect of the reform are of the same order of magnitude for survivors (panel 6) and decedents (panel 7), except for model 4. Furthermore, point estimates of the effect of the reform obtained using just the 1998 POLS (panel 8) are in the same range as using all waves. We conclude that the point estimates obtained using POLS data from 1997-2005 show little or no bias and can safely be used to produce the first stage estimate for TS2SLS.

Figure 4.3 shows the probability of dying in the period 1998-2005 for cohorts born between 1912 and 1922 using the 1998 RIO. The trend is smoother than in the corresponding graph produced from the POLS data[16] (figure 4.2) because of the larger sample size. Although the discontinuity at the reform threshold seems small, there does appear to be a small downward shift in the trend at that point. The corresponding reduced form estimates presented in the second panel of table 4.4 confirm that the 1928 reform induced changes that reduced the probability of dying between the ages of 81 and 88 (inclusive) by 1 to 3 percentage points, an effect that is significant in all specifications except model 1. Controlling for covariates does not change the results, while widening the bandwidth slightly increases the magnitude of the effect[17].

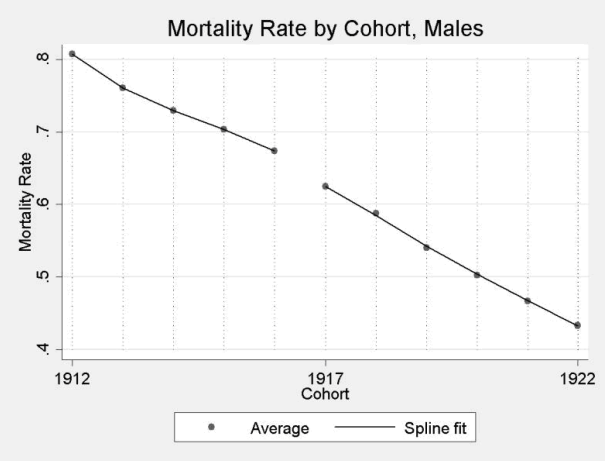
The TS2SLS estimates presented in table 4.5 indicate that, except for the linear model, an additional year of schooling reduces the probability of dying between the ages of 81 and 88 by between 2 and 3 percentage points. The estimates are statistically significant

[16]Note that the average mortality rate in the 1998 RIO is higher than in the POLS data, which is due to the fact that the POLS data samples a new cross-section of survivors every year between 1998 and 2005.

[17]Results available upon request.

at the 10 percent level or less. Given that around 50 percent of males who completed 6 years of schooling, corresponding to primary school completion, died between the ages of 81 and 88, our estimates suggest that one extra year of schooling reduced the probability of dying by 4-6 percent compared to the baseline.

Figure 4.3: Mortality rate, percentage that died in the period 1998-2005, by cohort (RIO)



Cohorts 1912-1922, males, RIO 1998.

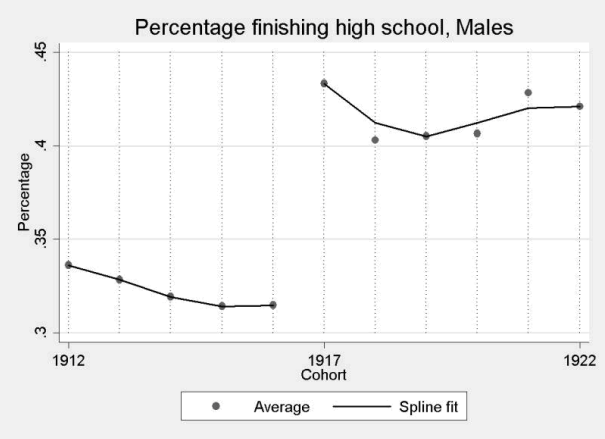
4.5 Are there additional returns to high school graduation?

While it is conventional to study the health (and other) returns to additional years of education, one might question whether health-related knowledge is acquired linearly irrespective of the stage of education completed. An additional year in lower vocational education is likely to constitute quite a different learning experience from an additional year in high school, such that the impact of each on health outcomes may be expected to differ (Feinstein *et al.*, 2006). Moreover, linearity imposes the potentially restrictive assumption that the health gains obtained from moving from lower vocational school (10 years) to high school (13 years) are only 3/4 of the gains following graduation from high school (13 years) to university (17 years). While the evidence for 'sheepskin effects' of education certificates on health is mixed (Cowell, 2006; Chevalier and Feinstein, 2007; Cutler and Lleras-Muney, 2008), and a priori it is difficult to imagine why a certificate should bring health returns over and above those of the years of schooling necessary to obtain the certificate, estimates of the health impact of years of schooling overlooks potentially

useful information on how the returns vary with the nature of schooling, as represented by the level of education completed.

In order to establish whether there is evidence of non-linearity and heterogeneity in the impact of education on health, in this section we estimate the effect of high school completion on mortality and compare this with the effect predicted from the linear specification. The 1928 reform increased the compulsory years of schooling from 6 to 7. Since primary school in the Netherlands consists of 6 years, this reform induced many individuals to enter secondary school. Indeed, as shown in figure 4.4, the percentage that finished high school appears to have increased dramatically due to the reform.

Figure 4.4: Percentage finishing high school by cohort



Notes: Cohorts 1912-1922, Males, POLS 1997-2005.

If, after controlling for the level of education, years of schooling have no additional impact, then the causal effect of the level of schooling on mortality can be identified. Yet, clearly with just one instrument we cannot identify the causal effect of all levels of schooling. It turns out, however, that the reform only had a significant impact on the propensity to complete (i) high school, and (ii) higher vocational education[18]. The causal effect of high school completion on mortality can then be identified if higher vocational education does not have any additional health benefits on top of those acquired by completing high school. Alternatively, if the reform induced pupils to complete high school but not higher vocational school conditional on finishing high school then we can identify the average effect of high school and higher vocational education completion on mortality. While debatable, these assumptions are not necessarily more restrictive than those necessary to identify the effect of years of schooling.

[18]Results available upon request.

The first stage results are presented in table 4.6. It can be seen that the 1928 reform strongly and significantly impacted on the propensity to finish high school. Around 10 percent of the individuals born in these cohorts were induced to complete high school by the reform. Combining these estimates with the reduced form mortality estimates from table 4.4, gives the TS2SLS estimate of the effect of high school graduation on mortality presented in table 4.7, which is between 17 and 26 percentage points in our preferred, more flexible models. Coupling this result with the fact that around 50 percent of males who did not complete high school died between the ages of 81 and 88, the probability of dying in this age range is reduced by a remarkable 34-52 percent for those that finished high school.

Given that high school graduation amounts to seven additional years of schooling for the compliers who, before the reform, would have ended their education after primary school, estimates from the models linear in years of education imply that high school graduation reduces the probability of dying between the ages of 81 and 88 by 14-19 percentage points (i.e. 7×2 to 7×2.7). This is below the estimate of a 17-26 point decrease generated by the binary model of high school completion, but not substantially so. Moreover, as explained above, it might be that the 17-26 percent partly stems from the completion of higher vocational education if this level had any additional health benefits on top of those provided by high school completion. Consequently, while there may be some 'sheepskin effects' of high school graduation, beyond what is predicted by the years of schooling it entails, there magnitude would appear to be rather modest.

4.6 Discussion

Education significantly and substantially reduces mortality even in old age. Our analysis reveals that, for Dutch males surviving to the age of 81, an additional year of schooling reduces the probability that they will die before reaching 89 by 2-3 percentage points, or 4-6% relative to the baseline probability. This suggests that the well-documented large correlation between education and health outcomes (Grossman and Kaestner, 1997; Smith and Kington, 1997; Mackenbach *et al.*, 1997; Cutler and Lleras-Muney, 2008) is not spurious but stems (at least partly) from a causal effect of education on health, and consequently mortality.

Our study exploits a compulsory schooling reform introduced in the Netherlands in 1928 which provides a strong and exogenous instrument for the educational attainment of Dutch males. The average years of schooling increased by 0.6-1.0 due to the reform, a result that is robust to several specification checks. In spite of the strong instrument,

the treatment effect estimated from a smaller sample (n=3650) is not significant. But by exploiting mortality data in a larger sample (n=66891), and combining this with education information from the smaller sample, we do find a significant impact of education on mortality. Our analysis is based on mortality observed at the individual level, rather than that approximated from changes in cohort sizes, or from a comparison of birth and death rates. The power provided by our sample size and individual level data is a great advantage over previous studies (Lleras-Muney, 2005; Clark and Royer, 2007; Mazumder, 2008), while compared to Albouy and Lequien (2008) the strength of the instrument and the age at which our affected cohort is observed might explain why we find a significant impact whereas they do not.

A distinguishing feature of our study is that we estimate the impact of schooling on mortality for individuals aged above 80. There is obviously pre-sample selective mortality and our estimates cannot be taken as indicative of the mortality effect of education at all ages. However, provided that education has a non-positive effect on mortality at all ages, then our estimates constitute a lower bound on the total effect of education on life expectancy.

Our IV estimates of the impact of schooling on mortality are larger than those obtained from OLS. This is consistent with measurement error exerting a greater downward bias on the OLS estimate than any upward bias arising from unobservables, such as ability and time preference that increase investments in both education and health. Indeed, IV estimates of the impact of education on earnings are often greater than those obtained from OLS (e.g. Card, 1999) and this has also been observed in estimates of the health returns to education (Arkes, 2003; Arendt, 2005; Lleras-Muney, 2005; Oreopoulos, 2006; Silles, 2009). But one should be cautious in comparing the IV and OLS estimates since they estimate different parameters. OLS seeks to estimate the average treatment effect (ATE) across the population, while the RDD IV identifies the LATE among compliers at the threshold of the 1928 reform (Hahn *et al.*, 2001; Van der Klaauw, 2002; Imbens and Lemieux, 2008). It is quite plausible that the treatment effect is larger at a lower level of education (Auld and Sidhu, 2005) such as those forced to stay on at school by the reform. From a policy perspective, this is clearly an interesting group. We wish to know the impact of education reforms on individuals whose behaviour is changed by them.

The fact that we find an effect implies that there are very long run returns to education. Not only does education raise earnings over the life cycle, it also extends the horizon of the lifetime. This is an important finding in the context of rising education levels and the ageing of populations worldwide. As more and better educated individuals reach old age, we can anticipate that mortality rates among the elderly will fall further and

populations will become even more 'grey'. Of course, this need not mean that health and social care needs rise since falling mortality rates will reflect improving levels of health, but it does mean that pensions will be stretched further to meet the consumption needs of an extended old age. The labour market returns to education need to be invested to provide for the health returns in the form of extended life.

The other side of the coin is that poorly educated individuals die earlier, enjoying a less extended period of retirement. There is a double injustice here. Not only does a lack of education lead to a deprivation of life itself, but it implies a lower return on investments in pensions made over the working life. On equity grounds, a case could be made for varying the retirement age with education[19], although the moral hazard effects induced by such a policy would probably render it undesirable.

Our results imply that education policies can be important instruments for tackling health inequalities. Design of effective policies requires knowledge of the causal mechanisms responsible for the impact of education on mortality in old age that we find. Two broad hypotheses have been advanced to explain an impact of education on health. Grossman (1972b) argues that, through information acquisition and processing skills, education raises the productivity of investments in health. This hypothesis is consistent with US evidence on educational disparities in the adoption of new medical technology (Goldman and Lakdawalla, 2001; Lleras-Muney and Lichtenberg, 2002; Glied and Lleras-Muney, 2003; Cutler *et al.*, 2008) and the management of illness (Goldman and Smith, 2002). In the context of the Dutch universal system of health insurance, inequality in access to medical care is a less plausible reason for education to impact on health, although differences in the management of disease could not be so easily dismissed.

A second hypothesis is that education operates through health behaviour — diet, exercise, smoking, drinking etc. (Muurinen, 1982). Empirical work has confirmed that the lower educated do indeed indulge in less healthy behaviour (Feinstein *et al.*, 2006; Cutler and Lleras-Muney, 2008) and that lifestyle acts as a mediator between education and health (Contoyannis and Jones, 2004; Balia and Jones, 2008). The cohorts studied in this paper were 42-52 when the US Surgeon General's Report on Smoking and Health was published in 1964. US studies report that the better educated were more likely to quit smoking, or not take it up, after the publication of evidence on its risks (De Walque, 2007; Grimard and Parent, 2007). Although most smoking-related deaths occur before the age of 80, Peto *et al.* (1992) estimate that 39% of deaths of Dutch males in 1995 above the age

[19]In the Netherlands this argument was made earlier by Bovenberg *et al.* (2006), who proposed to let the legal retirement age vary with life expectancy in order to make the pension system fairer and more robust to ageing and intergenerational tensions.

of 70 are smoking related. We investigated whether there was any evidence of a mechanism through smoking behaviour by comparing the estimated treatment effects of education on different causes of death. While we did find a significant effect on deaths from respiratory diseases and on all deaths categorised as smoking attributable mortality (US Department of Health and Social Service, 1989; Peto *et al.*, 1992), there was no consistent evidence of a larger effect on these than on other causes of death (results available upon request). Further research seeking to unravel the pathways responsible for the causal impact of education on mortality is clearly warranted.

Tables

Table 4.1: OLS estimates of the effect of years of education on the probability of dying between 1998 and 2005

Males			
Years of Education	-0.015^{***}	-0.013^{***}	-0.011^{***}
	(0.002)	(0.002)	(0.002)
Cohort	NO	YES	YES
Covariates	NO	NO	YES
Females			
Years of Education	-0.009^{***}	-0.008^{***}	-0.008^{***}
	(0.002)	(0.002)	(0.002)
Cohort	NO	YES	YES
Covariates	NO	NO	YES

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Estimates based on cohorts born between 1912 and 1922. Data are from the 1997-2005 POLS. Cohort refers to a quartic polynomial in cohort. Covariates include wave dummies, marital status, province, city size and ethnicity. Standard errors are given in parenthesis.

Table 4.2: First Stage OLS estimates of impact of 1928 Compulsory Schooling Law on years of education

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Reform	0.685***	0.669***	0.636***	1.039***
	(0.099)	(0.109)	(0.112)	(0.050)
Cohort	0.013	0.016	0.039	−0.241***
Cohort^2		−0.001		−0.049***
Reform*Cohort			−0.032	0.121**
Reform*Cohort^2				0.073***
Wave98	−0.243	−0.243	−0.242	−0.244
Wave99	−0.189	−0.189	−0.19	−0.194
Wave00	−0.660**	−0.660**	−0.661**	−0.661**
Wave01	0.128	0.128	0.128	0.125
Wave02	−0.098	−0.099	−0.099	−0.106
Wave03	−0.322	−0.322	−0.323	−0.328
Wave04	0.738**	0.737**	0.737**	0.734**
Wave05	−0.165	−0.166	−0.167	−0.165
Constant	9.676***	9.695***	9.743***	9.443***
AIC	19566.78	19566.76	19566.66	19565.42
F-statistic	47.55	37.94	32.21	436.08
G-statistic	0.17	0.17	0.16	0.02

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: The estimates are on male cohorts born between 1912 and 1922, controlling for cohort and wave dummies. Data are from the 1997-2005 POLS (N=3650). Models 1-4 refer to linear and quadratic polynomials (models 1 and 2, respectively) which are allowed to differ on either side of the threshold (models 3 and 4, respectively). "Reform" is 1 if 1917 cohort or later. The "Cohort" variable is centered on the 1917 cohort. AIC is the Akaike Information Criterion. F-statistic is for test of significance of 'reform', and is robust to clustering at cohort level and heteroskedasticity. G-statistic is the test statistic of the flexibility of the cohort polynomial, which follows a F(J-K,N-J) distribution, where J is the number of cohorts used in the estimation, K is the number of parameters, and N is the number of observations (Lee and Card, 2008). Standard errors (in parenthesis for "Reform") are robust to heteroskedasticity and clustering at the cohort level.

Table 4.3: Robustness checks on the impact of the 1928 Compulsory Schooling Law on years of education, males.

Variable	Model 1	Model 2	Model 3	Model 4
10 birth year cohorts (1907-1927), N=7845.				
Reform	0.690***	0.751***	0.787***	0.619***
	(0.078)	(0.080)	(0.075)	(0.139)
F-statistic	77.65	88.68	110.90	19.76
3 birth year cohorts (1914-1920), N=2218.				
Reform	0.800***	0.825***	0.819***	1.197***
	(0.058)	(0.055)	(0.058)	(0.053)
F-statistic	192.58	228.68	197.9	512.77
5 birth year cohorts with control for covariates, N=3650.				
Reform	0.629***	0.665***	0.642***	0.926***
	(0.099)	(0.092)	(0.084)	(0.038)
F-statistic	40.45	52.75	58.38	590.30
5 birth year cohorts excluding always-takers, N=3448.				
Reform	0.659***	0.572***	0.522**	1.029***
	(0.139)	(0.175)	(0.204)	(0.203)
F-statistic	22.59	10.74	6.54	25.79
5 birth year cohorts without wave dummies, N=3650.				
Reform	0.692***	0.681***	0.651***	1.082***
	(0.105)	(0.117)	(0.124)	(0.053)
F-statistic	43.40	33.80	27.44	423.00
5 birth year cohorts among Survivors, N=1991.				
Reform	0.681*	0.864**	0.757**	1.384***
5 birth year cohorts among Decedents, N=1659.				
Reform	0.698***	0.575***	0.567***	0.689***
5 birth year cohorts, just 1998 sample, N=833.				
Reform	0.911**	0.806**	0.760**	0.917*

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Data are from the 1997-2005 POLS. "Reform" is 1 if 1917 cohort or later. Always-takers refer to those that finished university/college or PhD. Models 1-4 and F-statistic as in Table 4.2. Standard errors (in parenthesis for "Reform") are robust to heteroskedasticity and clustering at the cohort level.

Table 4.4: Reduced Form OLS estimates of the impact of the 1928 Compulsory Schooling Law on the probability of dying between 1998 and 2005 inclusive

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Reduced Form estimates, POLS 1997-2005, N=3650.				
Reform	0.000	−0.013	−0.020	−0.029
	(0.023)	(0.021)	(0.020)	(0.027)
Cohort	−0.031***	−0.028***	−0.020***	−0.005
Cohort^2		−0.001		0.003
Reform*Cohort			−0.013**	−0.036
Reform*Cohort^2				−0.001
Wave98	−0.055**	−0.055**	−0.055**	−0.055**
Wave99	−0.092***	−0.092***	−0.092***	−0.092***
Wave00	−0.148***	−0.148***	−0.148***	−0.148***
Wave01	−0.224***	−0.224***	−0.224***	−0.224***
Wave02	−0.306***	−0.307***	−0.307***	−0.307***
Wave03	−0.403***	−0.403***	−0.403***	−0.404***
Wave04	−0.453***	−0.453***	−0.453***	−0.453***
Wave05	−0.562***	−0.563***	−0.563***	−0.563***
Constant	0.649***	0.663***	0.675***	0.692***
AIC	4830.1	4829.5	4829.1	4828.8
G-Statistic	0.47	0.40	0.36	0.33
Reduced Form estimates, RIO 1998 follow-up, N=66891.				
Reform	−0.008	−0.014*	−0.018***	−0.029***
	(0.009)	(0.007)	(0.005)	(0.004)
Cohort	−0.037***	−0.036***	−0.032***	−0.018***
Cohort^2		−0.000*		−0.002**
Reform*Cohort			−0.007***	−0.027***
Reform*Cohort^2				−0.001*
Constant	0.625***	0.633***	0.640***	0.655***
AIC	92278.0	92277.0	92274.5	92275.6
G-statistic	1.21	0.95	0.60	0.25

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Estimates based on cohorts 1912-1922, males. "Reform" is 1 if 1917 cohort or later. AIC and G-test as in table 4.2. Standard errors (in parenthesis for "Reform") are robust to heteroskedasticity and clustering at the cohort level.

Table 4.5: RDD estimates of the impact of years of education on the probability of dying between the ages of 80-88 (2SLS) or 81-88 (TS2SLS)

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
2SLS estimates				
Years of Education	0.000	−0.019	−0.031	−0.028
	(0.034)	(0.030)	(0.027)	(0.025)
TS2SLS estimates				
Years of Education	−0.011	−0.020*	−0.027***	−0.026***
	(0.013)	(0.011)	(0.009)	(0.004)

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Estimates based on cohorts 1912-1922, males. Both the POLS and RIO data are linked to the Cause-of-Death Register (DO). Years of Education is instrumented by the 1928 compulsory schooling reform. 2SLS estimates based upon POLS 1997-2005, N=3650. TS2SLS estimates based upon POLS 1997-2005, N=3650 & RIO 1998 follow-up, N=66891. Models 1-4 as in table 4.2 for 2SLS and as this without wave dummies for TS2SLS. Standard errors (in parenthesis) are robust to heteroskedasticity and clustering at the cohort level and are obtained by the Delta method for the TS2SLS estimation.

Table 4.6: First Stage OLS estimates of impact of 1928 Compulsory Schooling Law on the probability of completing high school.

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Reform	0.096***	0.108***	0.105***	0.105***
	(0.016)	(0.015)	(0.013)	(0.010)
Cohort	0.000	−0.002	−0.005***	0.007
Cohort^2		0.001***		0.002***
Reform*Cohort			0.006	−0.021**
Reform*Cohort^2				0.001
Wave98	−0.030	−0.030	−0.030	−0.030
Wave99	−0.027	−0.027	−0.027	−0.027
Wave00	−0.081**	−0.081**	−0.081**	−0.081**
Wave01	−0.002	−0.002	−0.002	−0.002
Wave02	−0.034	−0.034	−0.034	−0.035
Wave03	−0.030	−0.031	−0.030	−0.031
Wave04	0.077*	0.077*	0.077*	0.078*
Wave05	−0.065	−0.064	−0.064	−0.064
Constant	0.347***	0.334***	0.334***	0.347***
AIC	5102.6	5102.1	5102.4	5101.8
F-statistic	36.51	54.86	65.97	111.40
G-statistic	0.18	0.13	0.16	0.09

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Estimates based on cohorts 1912-1922, males. Data are from the 1997-2005 POLS (N=3650). "Reform" is 1 if 1917 cohort or later. AIC, F-test and G-test as in table 4.2. Standard errors (in parenthesis for "Reform") are robust to heteroskedasticity and clustering at the cohort level.

Table 4.7: RDD estimates of the impact of high school completion on the probability of dying between the ages of 81 and 88.

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
High school completion	-0.078	-0.128^{*}	-0.168^{***}	-0.264^{***}
	(0.096)	(0.067)	(0.049)	(0.044)

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Estimates based on cohorts 1912-1922, males. Data are from the 1997-2005 POLS linked to the Cause-of-Death Register (DO) (N=3650) and the RIO 1998 linked to DO (N=66891). Models 1-4 as in table 4.2. Standard errors (in parenthesis) are robust to heteroskedasticity and clustering at the cohort level, and are obtained by the Delta method.

Chapter 5

Whose Income is Least Protected against Ill-Health?

We estimate the causal effects of health shocks on labour force participation and personal and household disposable income in the Netherlands using exogenous variation in health provided by acute hospitalizations. An acute hospital admission significantly and substantially lowers the probability of remaining at work and, on average, reduces annual personal disposable income by about 1,600 Euro three years after a health shock, and the loss does not diminish over time. The impact on household income is much larger, around 3,500 Euro, which indicates that disability insurance mechanisms insure against most of the loss in personal income, but cannot fully smooth the income of the household. The most vulnerable groups are the lower socioeconomic groups — those in the lowest income quartile, and the lower educated. Individuals working in the public sector are best protected against income losses.

This chapter is based upon:

García-Gómez, P., H. van Kippersluis, O. O'Donnell, and E. van Doorslaer (2010), "Whose Income is Least Protected against Ill-Health?", Mimeo Erasmus University Rotterdam.

5.1 Introduction

On average, OECD countries spend two percent of their GDP on disability and sickness benefits — almost 2.5 times what they spend on unemployment benefits. Prompted by such large, and increasing, expenditures, research has concentrated on quantifying the extent to which they are attributable to moral hazard, while policy reforms have attempted to constrain the inflow to disability rolls. Less attention has been given to the intended benefit of disability and sickness insurance in the form of income protection. The magnitude of this benefit is likely to vary across individuals with their demographics, severity of disability, socioeconomic status, occupation and employment sector.

Charles (2003) argues that the earnings losses from disability are likely to be greater for individuals who are older at the onset of disability, both because more of their accumulated human capital is destroyed and because they have less incentive to invest in acquiring disability specific human capital. He finds empirical support for this proposition and both mechanisms. This implies that older individuals benefit more from having insurance against disability-related income losses. The benefit may also differ by education since highly educated professionals have more human capital that is not contingent on physical health. Lower educated unskilled workers rely more on their physical health. Should they lose this, their relative earnings capacity will fall by more. More highly educated individuals may also be better at managing their disease, for example, through adherence to medication (Goldman and Smith, 2002), such that it impinges less on their earning capacity. Differences by employment sector may also be anticipated. In the public sector, where there is less exposure to competition and consequently salaries are more regulated and less tied to productivity, employment and earnings should be less responsive to ill-health. The employer, rather than the disability insurance scheme, protects against the economic consequences of ill-health.

Motivated by these hypotheses, this paper identifies the causal effects of ill-health on employment, disability benefit receipt and income, and examines the extent to which these effects vary with demographic, socioeconomic and job characteristics. It identifies the population groups whose employment is least protected against health deterioration and those most vulnerable to a large fall in income as a result of ill-health. We identify these effects by tracing the tax records of Dutch individuals for up to seven years after they experience an acute admission to hospital. This provides exogenous variation in health that allows its causal effects on employment, benefit receipt and income to be identified. This circumvents an endogeneity problem that has plagued attempts to identify the employment impact of ill-health from reported data on health that is potentially subject to

a 'justification bias' (Bound, 1991). That is, ill-health is reported as a justification for not working and, possibly, claiming disability benefits. The seven-year follow-up provides a rare opportunity to observe how employment and income evolve both in the period immediately following the health shock and over the longer term. This makes it possible to examine the extent to which an individual is able to recover, at least in terms of labour market performance, from a significant deterioration in health, perhaps by investing in human capital that is more consistent with a reduced health status. Any permanent reduction in disposable income will require a downward adjustment of consumption and so welfare (Deaton, 2002b; Attanasio and Davis, 1996; Blundell *et al.*, 2008).

After moral hazard, concern about the efficiency of public insurance against disability, health or unemployment risks tends to focus on the crowd-out of private and informal insurance. In relation to disability, an example of the latter is substitution of the earnings of the disabled person with those of the spouse. Formal insurance is more efficient and should therefore be preferable to such an arrangement, not least because the need of the disabled person for care may constrain the extent to which the spouse can replace the lost earnings. Depending upon the severity and nature of the disability, the need for care may be so great that the spouse, and not only the disabled person, is forced to reduce his, or her, earnings. There is little evidence on this issue. Some studies look at the effect on wives' labour supply when husbands turn ill. Typically the effects are a very small increase in hours worked or no effect at all (Parsons, 1997; Berger, 1983; Berger and Fleisher, 1984; Haurin, 1989). Charles (1999) presents US evidence using the Health and Retirement Study (HRS) which indicates that men are more likely to reduce their labour supply if their wives fall ill, while women more often try to compensate the lost household income by increasing their labour supply. He attempts to explain this finding by noting that there is a significant degree of specialisation within the household, where husbands are often the principal breadwinners, while wives the principal home and caretakers. When one of the two falls ill, the spouse has to compensate. This implies that when a spouse falls ill, women begin to work, or work more hours, while husbands cut back on work to help around the home (Charles, 1999). The finding of a modest increase in the wife's labour supply when the husband falls ill has been partly replicated using US and European data, yet only if the husband is out of work (Blau, 1998; Jiménez-Martín *et al.*, 1999). But one study finds a modest increase in wife's labour force participation if the husband remains at work (Blau and Riphahn, 1999). Hence, the literature is inconclusive regarding the spillover effect of illness to the earnings of other household members (Suhrcke *et al.*, 2006). We add to this evidence by identifying the impact of a

health shock, in the form of an acute hospitalisation, not only on personal income, but also on household income, separately for males and females.

Studies of the 50+ population that rely on self-reported indicators of health generally find a strong negative impact of ill-health, or disability, on earnings and income that operates through employment and work hours rather than wages (e.g. Bound *et al.*, 1999; Smith, 1999; Riphahn, 1999; Charles, 2003; Wu, 2003; Au *et al.*, 2005; Disney *et al.*, 2006; Hagan *et al.*, 2006; Jiménez-Martin *et al.*, 2006). Fewer studies include individuals below the age of 50 but those that do generally find similar effects on employment and income (Lechner and Vazquez-Alvarez, 2004; García-Gómez and López-Nicolás, 2006; García-Gómez, 2008). While the consistency of this evidence is persuasive, its validity is weakened by the possibility that ill-health is reported, perhaps subconsciously, as a justification for labour force withdrawal and benefit receipt. Moller Dano (2005) avoids this problem by using road accidents to provide exogenous variation in health. The study, which uses administrative data on 10 percent of the Danish population, finds negative effects on disposable income only for older individuals and for those with lower initial incomes. There is a significant effect on employment only for males, for whom the employment rate decreases by around 10% after the accident and does not recover in the following six years. Lindeboom *et al.* (2009) also exploit data on accidents, albeit self-reported, recorded in a British cohort study. They find that an accident raises the risk of disability at age 25 by 170% and, in turn, this reduces the probability of employment at age 40 by 0.14.

Evidence on the causal effects of ill-health on employment and income presented in this paper is strengthened by particularly rich data that link demographic, health and socioeconomic information from a survey with hospital admission records and income tax files in the Netherlands for the period 1998 to 2005. Employment and income effects on an initially healthy group of working individuals between the ages of 17 and 64 are identified from exogenous variation in health in the form of acute hospitalisations. An acute hospitalisation is defined as an unscheduled hospitalisation, which cannot be postponed since immediate treatment is deemed necessary. These hospitalisations cover a far wider range of conditions than those caused by road accidents on which Moller Dano (2005) relies for identification. Our results should therefore have greater external validity. In comparison with Lindeboom *et al.* (2007), our health measure is clearly more objective and our data cover the full working aged population, rather than the sub-population below the age of 40 that has a much higher rate of employment. Use of tax files also minimises measurement error in income.

While we argue that acute hospitalisations are exogenously determined, they are clearly not randomly distributed across the population. We take account of observable differences by matching individuals who experience acute hospitalisations with non-hospitalised individuals using propensity scores and combine this with difference-in-differences to correct for time invariant unobservable differences between the two groups.

The Netherlands provides an interesting context in which to examine the employment and income effects of disability. One study finds it to be among the European countries in which the effect of ill-health on employment is largest (García-Gómez, 2008). This has been attributed to the generosity of its Disability Insurance system (Aarts *et al.* 1996; Bound and Burkhauser, 1999), which has become more stringent in recent years (OECD, 2009). The focus of policy has recently shifted from income protection toward protection of employment (De Jong, 2008). We examine the extent to which maintenance of employment following a health shock is sufficient to protect income and how this varies systematically across the population. We also compare actual replacement rates for those that move onto disability benefits with the reduction in income of those that remain at work. These replacement rates are calculated on the basis of income, and not earnings, and so need not correspond to the legally defined rates. Divergence between actual and defined replacement rates will depend on the extent to which non-earnings sources of income are affected by the onset of disability. Additional sources of income, and their sensitivity to disability, are likely to vary across population groups. Consequently, actual replacement rates may vary in quite a different way from those defined in the DI system. In addition to examining the rate of replacement of personal income for individuals moving onto disability benefits, or into retirement, we also compute and compare replacement rates for total household income, which depend upon whether disability leads to an increase or decrease in the earnings of a spouse.

In section 5.2 we describe features of the Dutch Disability Insurance system and labour market that can be expected to influence the extent to which ill-health impacts on labour force status and income. Section 5.3 describes the data and explains the empirical strategy. Section 5.4 presents the results after which a discussion follows in section 5.5.

5.2 Work and disability in the Netherlands: institutional background

In the Netherlands, as in most other developed countries, there are three major cash transfer programs for individuals of working age that do not work — Disability Insurance

(DI), Unemployment Insurance (UI), and social assistance (welfare). Despite a dramatic decrease in the share of GDP spent on DI and sickness benefits from 7.6% in 1990 to 4.6% in 2005, it still accounts for 22% of public social spending — three times the amount spent on UI (OECD, 2009). Individuals are entitled to DI benefits if they have a degree of disability based on residual earning capacity beyond 15 percent. Entitlement does not depend on whether illness or injury is contracted through work and, more peculiarly, is independent of contributions history. DI benefits are paid after a waiting-period of one year[1]. Until then, the employer is responsible for financing the sick pay, which is equal to 70 percent of the gross wage. However, collective bargaining agreements usually ensure that sick employees receive up to 100 percent of their net salary[2]. The replacement rates are defined in terms of previous net salary excluding overtime or bonuses, so actual replacement rates on disposable income might be below 100%.

DI pays benefits in two phases. In the first phase, which lasts for 0 to 6 years depending on age at onset of disability[3], the recipient receives a percentage of the previous wage. The percentage is based on the severity of the illness or injury up to a maximum of 75 percent if the individual is assessed as 80% disabled or more. The partially disabled, who represent around 20% of recipients (OECD, 2009), receive pro rata benefits and are allowed to work to close the earnings gap. Two thirds of those awarded partial benefits work. For them the benefit acts as a wage subsidy (García-Gómez *et al.*, 2009).

After this first phase, the benefit is no longer set only in relation to previous wage and the severity of disability but is equal to the minimum wage plus an addition increasing in age and previous wage[4]. In all cases this follow-up benefit is constrained to be lower than that paid in the first phase, but it can be paid until the individual reaches the age of 65. Individuals can choose whether to insure against the difference between initial and follow-up benefits, and in most cases this is part of the collective bargaining agreement (De Jong, 2008).

In the period we consider (1998-2005) there was a separate disability insurance program for the self-employed. It awarded minimum benefits at a lower replacement rate than DI for employees, and the contribution rate was uniform.

[1] After 2004 the waiting period was extended to two years, but this is outside our observation period as individuals in our sample are expected to enter into sickness benefits because of a health shock in 2002 at the latest.

[2] Civil servants always receive 100 percent of their net salary, and regular employees get this in 90% of cases (Burkhauser *et al.*, 2008).

[3] More specifically, the entitlement period ranges from 0 years for those under 33 to 6 years for those over 58 years of age.

[4] The exact formula is: the minimum wage + (age - 15)*(previous wage - minimum wage)

Individuals who are not awarded partial disability status, but who become unemployed are entitled to UI. During our observation period, the UI replacement rate was 70 percent and the benefit period ranged from a minimum of 6 months to a maximum period of 5 years depending on employment history and age. Individuals awarded partial disability status who cannot find gainful employment are entitled to a partial UI benefit of up to 70 percent of lost earnings (De Jong, 2008). If the individual cannot qualify for either DI or UI, then s/he can resort to social assistance, which pays lower benefits unrelated to previous earnings.

The labor market institutions in the Netherlands are considered intermediate in terms of strictness in an international perspective (Freeman, 2007). The regulations for hiring and firing are quite strict, and a high minimum wage applies. The Netherlands has intermediate levels of taxation and generous social benefits, which could reduce labor force participation, hours worked, and employment (see Heckman and Jacobs, 2010 and references therein). Also, in the years considered, generous early retirement and pension schemes were present which made older people retire at young ages.

5.3 Data and Empirical Strategy

5.3.1 Data

Our data are linked survey and administrative records from Statistics Netherlands. We use the annual cross-sectional general household survey (POLS) 1998-2001, the tax records (RIO), the hospital discharge register (LMR), the Cause-of-Death register (DO), and the Municipality Register (GBA) all in the years 1998 until 2005 inclusive. The LMR provides exogenous variation in health in the form of acute hospitalisations. The RIO provides the outcome measures - employment, DI/UI/pension receipt and disposable incomes. POLS provides baseline health and socioeconomic information used for matching. Demographics (year of birth, sex, marital status and nationality), province of residence and size of the city are obtained from the GBA. The death register is used to identify individuals that drop out during the follow up period due to death.

The POLS samples a representative cross-section of the non-institutionalised Dutch population ranging from around 25,000 to 80,000 respondents per year[5]. It collects extensive information on health and socioeconomic characteristics. We retrieve information at baseline on level of education, number of hours worked, job characteristics and home ownership. Health and health-related behaviour are represented by self-assessed health (very

[5]Specifically 80,789 in 1998, 42,605 in 1999, 37,482 in 2000, and 24,231 in 2001.

good – poor), a binary indicator of whether ill-health hampers daily activities, number of general practitioner (GP) visits, smoker status (yes/no), and frequency of engagement in sports.

The RIO is a longitudinal administrative tax-register covering one third of the Dutch population, i.e. around five million observations per year from 1998 to 2005. It consists of a personal file and a household file. The personal file provides personal disposable income, which is gross income from wage, profit and wealth earnings plus transfers less taxes and premiums. Income by source is not available but the main source of income is used to identify labour market status. This can be income from work (subdivided into civil servants, regular employees, executives, and self-employed), income from DI, income from UI, income from old-age pensions, income from other social transfers or no income[6]. As mentioned above, during the one-year waiting period for DI, sickness benefits are paid by the employer. In the tax files, this will appear as income from an employer and individuals receiving sickness benefits will therefore be classified as being employed. This is an unavoidable limitation of using the tax files, which impedes our ability to identify the impact of a health shock on employment in the first year after hospitalisation. The household level tax file provides total household disposable income and the number of household members.

The hospital discharge register contains data on both inpatient and day care patients of all general and university hospitals and most of the specialised hospitals[7] in the Netherlands from 1998 to 2005. Each year there are around two million hospitalisations of around 1.6 million distinct individuals. For the entire Dutch population we observe (i) whether the individual entered the hospital, (ii) whether it was an acute admission, (iii) the admission and discharge date, and (iv) the main diagnosis based on the International Classification of Diseases ICD-9CM. We compute the number of nights in the hospital using the admission and discharge dates.

5.3.2 Empirical Strategy

We compare the labour market status, benefit receipt and income of individuals who have experienced an acute hospitalisation with those that have not. By identifying from variation provided by hospitalisations, our results are not subject to the justification bias that is suspected when variation in self-reported health is relied on. The comparison of outcomes is conducted within a propensity score matching approach (Rosenbaum and Rubin,

[6]The category 'no income' includes individuals with solely income from capital, and those with only bounded transfers such as allowances for renting and children.

[7]The average coverage of the specialised hospitals is 97% (de Bruin *et al.*, 2004).

1983). In addition to matching on a detailed set of health, demographic, socioeconomic and job characteristics, we control for fixed unobservable determinants by conditioning on pre-treatment outcomes, either by including them in the propensity score or by restricting the sample of controls to individuals identical to the treated in terms of pre-treatment outcomes.

The Average Treatment Effect on the Treated ($ATET$) is identified under the assumption that the observable controls and the pre-treatment outcomes include all the factors that determine both whether an individual experiences an acute hospitalisation and his potential outcome in the absence of this event. By combining matching with differences-in-differences (MDID) we weaken this identifying assumption to the requirement that, conditional on observables, in the absence of hospitalisation the evolution of outcomes between the pre-treatment to post-treatment period would have been the same for the group that was hospitalised and the controls who were not (Heckman *et al.*, 1997; Blundell and Costa-Dias, 2009). Essentially, matching ensures that on average treatment and controls are similar in their pre-treatment outcomes, while at an individual basis there still might be (large) differences between a treated individual and his/her matched control. If matching is combined with differences-in-differences this possibility is controlled for under the assumption that the trend in outcomes would have been the same had treatment not happened. In general, this method performs better than matching alone (Smith and Todd, 2005).

We pool the POLS, RIO and LMR data from 1998 to 2001. For each year 1999-2001 we split the sample according to whether or not the individual has an acute hospitalisation[8] (excluding those related to pregnancy and child birth). Treatment and control groups are formed from these two samples by selecting in each case individuals who in the previous year were: i) aged 17-64, ii) working[9] and, iii) not admitted to hospital. Both groups are followed for up to three years after the hospitalisation of the treatment group[10]. This approach provides a sufficient number of treatment observations to examine systematic heterogeneity in effects, but it limits the follow-up period to three years. To examine longer run effects, up to six years after hospitalisation, we link the 1998 POLS, RIO and

[8]Individuals in the control group are allowed to experience non-acute hospitalisations.

[9]Note that since we restrict our sample to individuals working the period before the health shock takes place, matching is technically identical to matching combined with differences-in-differences for employment status. Note however that in the case of income it is impossible to restrict the sample to individuals with the same level of income, such that we decided to apply matching combined with differences-in-differences in that case.

[10]Note that individuals in the treatment group are allowed to have multiple acute hospitalisations over the follow-up period.

LMR and again restrict the sample to individuals aged 17-64, working, and not admitted to the hospital in this year. Treatment and control groups are formed from this sample according to whether or not there is an acute hospitalisation in 1999. Outcomes are observed in the 1999-2005 RIO.

Table 5.1 shows that the percentage of both treatment and control groups that remains at work decreases over time but, particularly two years after hospitalisation, the rate of decrease is greater for the treatment group such that after three years the raw difference in the proportion at work is more than 8 percentage points[11]. As expected, the percentage entering DI is higher in the treatment group — the difference reaching five percentage points three years after hospitalisation. At this time the proportion of the treatment group that has retired is 2.7 points greater than that of the control group. The proportion on UI and on social assistance is much lower but in both cases it is more than twice as high in the treatment group three years after hospitalisation. Personal disposable income increases gradually in the control group but stagnates two years after hospitalisation in the treatment group and decreases after three years. At baseline, household disposable income is very similar in the two groups but while it falls slightly in the treatment group, it is more than 3,000 Euro higher in the control group after three years.

The propensity score for hospitalisation is estimated by a probit including: (1) demographics — age, gender, marital status, province, city size, nationality, and household size, (2) health indicators — self-assessed health, hampered in daily activities, GP visits, smoker status, and frequency of exercise, (3) socioeconomic indicators — home ownership, education, equivalent household income and ratio of personal to household income, and (4) job characteristics — hours worked, occupation and sector of employment (public/private).

Table 5.2 lists the variables included in the propensity score, their mean values for treated and control individuals, and from which dataset they were obtained. There are no substantial differences between the two groups, and none of the small differences is statistically significant at any conventional level. In the sample, the percentage of (middle-aged) men, self-employed, lower educated, and divorced individuals is higher in the treatment group, while this group has worse initial health and exercises less. The probit coefficients from estimation of the propensity score given in the third column of the table confirm male, divorced and individuals hampered in daily activities are significantly more likely to be hospitalised, while the higher educated and those working in the private sector are less likely to enter hospital for an urgent reason.

[11] As acknowledged above, we identify employment status from main source of income. We use 'in work' as short-hand for employee or self-employed earnings being the main source of income.

We restrict the sample to observations within the common support, which results in only very slight reductions in the sizes of the treatment and control groups (from 398 to 395 and 32,085 to 31,487 respectively). The procedure recommended by Dehejia and Wahba (1999, 2002) is followed to ensure satisfaction of the balancing hypothesis.

We employ a kernel matching approach to estimate the $ATET^{MDID}$ on the outcomes of interest — employment status and personal and household disposable incomes — as follows:

$$ATET_t^{MDID} = \frac{1}{N^T} \sum_{i \in T} \left\{ \left[Y_i^t - Y_i^0 \right] - \sum_{j \in C} w_{i,j} \left[Y_j^t - Y_j^0 \right] \right\} \tag{5.1}$$

where $w_{i,j}$ denotes the weight attributed to control individual j when comparing with treated individual i. T refers to the treatment group, and C to the control group. We employ an Epanechnikov kernel with bandwidth of 0.06, and use only individuals under the common support of the propensity score.

The $ATET^{MDID}$ measures the effect in absolute values. Relative effects can be more meaningful for comparison across groups. To obtain the Relative $ATET$ ($RATET$), we divide the $ATET^{MDID}$ in year t by the counterfactual outcome obtained using the common trends assumption for the matched controls[12].

$$RATET_t^{MDID} = \frac{ATET_t^{MDID}}{E(Y_0|T) + [E(Y_t|C) - E(Y_0|C)]} = \frac{[E(Y_t|T) - E(Y_0|T)] - [E(Y_t|C) - E(Y_0|C)]}{E(Y_0|T) + [E(Y_t|C) - E(Y_0|C)]} \tag{5.2}$$

5.4 Results

5.4.1 What is the effect of a health shock on employment and income?

The $ATET$ of acute hospitalisation on employment is negative but small and insignificant in the year of hospitalisation and the following year (Table 5.3, column 1). Since individuals received sickness benefits paid by the employer during the first year of sickness and we are defining labour force status according to the main source of income, we are not able to identify any effect on work status during this first year. Two years after hospitalisation, the probability of remaining at work is significantly lower for the treatment group,

[12]Note that for employment outcomes, initial conditions are the same and so the relative treatment effect is simply the $ATET$ divided by the mean for the control group in the relevant time period.

and is almost eight percentage points lower after three years, which represents a relative decrease of more than 10% (Table 5.9, row 1). The vast majority of individuals leaving employment enter DI. An acute hospitalisation raises the probability of being in receipt of DI by 3.5 percentage points after one year and by 5.2 points after three years. The latter represents a relative increase of over 300%. Hospitalisation raises the probability of being retired three years later by 2.6 percentage points and of receiving social assistance by 1 point, which represent relative increases of almost 100% and 250% respectively on the probability for the control group.

Personal disposable income falls by around 800 Euro in the first year after hospitalisation, and this drop doubles after three years (Table 5.3), which is an average relative decrease of 7% (Table 5.9). The absolute effect on household income is more than twice as large. On average, three years after the event, a household with a member than has been admitted to hospital has an income 3,600 Euro lower than it would otherwise have had. A relative drop of more than 8%. This indicates substantial negative spillover effects on the incomes of other household members. The most likely mechanism seems to be a fall in the earnings of the spouse as a result of meeting the care needs of the disabled person.

If we stratify the effects by gender, it turns out that the results are quite different across males and females. While the effects on remaining at work are very similar to the population average, the probability of entering DI increases much more for males than it does for females — 6 versus 3.6 percentage points (see Table 5.4). The absolute drop in personal disposable income is again quite similar in absolute terms across males and females at around 1,600 to 1,800 Euro. However in relative terms females lose double the amount of men given their lower average incomes (12 versus 6 percent, see Table 5.9). The most striking finding is that our results indicate that if a male falls ill, the drop in household income is three times the size of the drop in personal income, whereas if a female falls ill, the drop in household income is statistically indistinguishable from zero. This suggests that when the husband falls ill, the wife is more likely to reduce her working hours, or stop working completely, to take care of her husband, while if the wife becomes disabled, the husband is ready to compensate for the earnings loss if necessary. This finding challenges existing evidence by Blau (1998), Blau and Riphahn (1999), Charles (1999), and Jiménez-Martín *et al.* (1999) who claim that females are more likely to compensate for the earnings loss if their husbands fall ill, while husbands are more likely to stop working in case their wives become disabled.

We partly confirm the evidence by Charles (2003) that the earnings losses from disability are greater for individuals who are older at the onset of disability. Table 5.4 shows that older individuals are far more likely to stop working after a health shock, a

15 percentage points decrease in the probability of remaining at work, which constitutes a relative decrease of almost 24 percent (Table 5.9). In contrast, younger individual are not more likely to stop working, possibly because few accumulated human capital is destroyed and because they have a strong incentive in acquiring disability specific human capital. The labour market institutions in the Netherlands probably also contribute to a high permanent dropout rate among older employees. It has to be taken into account that older individuals receive far more severe health shocks, such that definitely not all results are entirely driven by behavioural choices. The results on personal and household income do not unambiguously support the theory of Charles (2003). Younger individuals on average lose an amount of 1,400 Euro three years after the shock, while older individuals do not lose anything. This might be explained by the forgone career prospects of younger individuals and the strong contingency of replacement rates on age. Yet, the effects on household income are again larger for older individuals, either reflecting the stronger need for a caring spouse because of a more severe health shock on average for older individuals incurring a shock to their health, or joint retirement (e.g. Hurd, 1990; Jiménez-Martín *et al.*, 1999).

The drop in annual personal income of 1,600 Euro three years after hospitalisation is an average of a much smaller effect on those that remain in work and a large effect on those leaving work. There is no significant fall in income for those that remain in work up to two years after hospitalisation and a fall of 1,000 Euro after three years that is significant at only 10% (Table 5.5) and represents a 4% relative decline (Table 5.9). This is consistent with the repeated finding in the literature that ill-health reduces earnings by constraining employment rather than by depressing wages. While the point estimates are mostly negative and larger than those for personal income, there is no significant effect on household income for individuals who remain in work (Table 5.5). Since these individuals will have minimal or no caring needs, this is consistent with our hypothesis that the negative spillover effect on the income of other household members observed in the full sample is due to meeting caring needs of the hospitalised person.

The effects on the personal incomes of individuals who leave work and enter DI (in any of the follow-up years) are much larger. In the year of hospitalisation, income already falls by almost 1,500 Euro. Since, for the vast majority, the replacement rate of basic earnings is 100% during this period, this fall may be due to the loss of overtime pay and bonuses. Alternatively, individuals might have been on sickness pay already at baseline[13] (i.e. the year before hospitalisation), which means that in the year of hospitalisation they

[13]Recall that we identify labour force status from main source of income such that those on sickness pay are classified as employed.

could be receiving DI benefits if the waiting-period has elapsed. Over the next three years, the loss of income gradually increases and reaches 6,500 Euro, on average, in the third year. This corresponds to a 30% fall in relative terms (Table 5.9) coinciding exactly with the 70% replacement rate offered by DI after the first phase. The smaller losses in years $t+1$ and $t+2$ are probably due to higher benefits in the first phase, or may reflect individuals claiming partial disability benefits supplemented by part-time work. It is very interesting that among those that move onto DI following hospitalisation, the loss of household income is much greater than that of personal income. In the year of hospitalisation, other household members lose as much income as the hospitalised person. This indicates that sudden ill-health has a substantial disruptive on household finances. Time must be taken off work to visit the hospitalised person, to care for them after discharge and, perhaps, to replace child care the sick person can no longer provide. The loss of household income increases by almost 2,000 Euro in each year reaching 8,700 Euro after three years, which is 2,000 Euro more than the loss of personal income (Table 5.5). So, while DI succeeds, on average, in preserving personal income at 70% of what it would have been without illness, there is an additional, not insubstantial, loss of income of other household members to consider when weighing the impact on household finances.

An acute hospitalisation is not necessarily severe as people with minor health problems (like twisted ankles) also enter the emergency room. In order to identify an exogenous and more severe health shock we redefined treatment in two different ways. A first new definition is an acute hospitalisation with a length of stay of at least 3 nights. The arbitrary cut-off at 3 nights derives from a trade-off between severity of shock and sufficient power: we test for the effect of a smaller number of probably more severe shocks. For a second alternative definition of a health shock we exploit the disease information in the same spirit as Smith (1999) with the onset of chronic conditions. We define the treatment group as individuals hospitalised with a primary diagnosis of a cancer, a circulatory disease or a respiratory disease. Since individuals that were admitted to the hospital in the year before the health shock are excluded, this definition of a health shock is to be interpreted as the onset of a new chronic condition.

The results show that moving from any acute hospitalisation to a more restricted measure with a minimal length of stay of 3 nights clearly suggests an increasing degree of severity of the health problem as the probability of remaining at work declines while the probability of entering disability insurance benefits and retirement rises (Table 5.3, column 2). Interestingly, the drop in personal and household income is not necessarily larger among more severe health shocks and in some cases even smaller, although the differences are not statistically different here. The fact that the difference between the

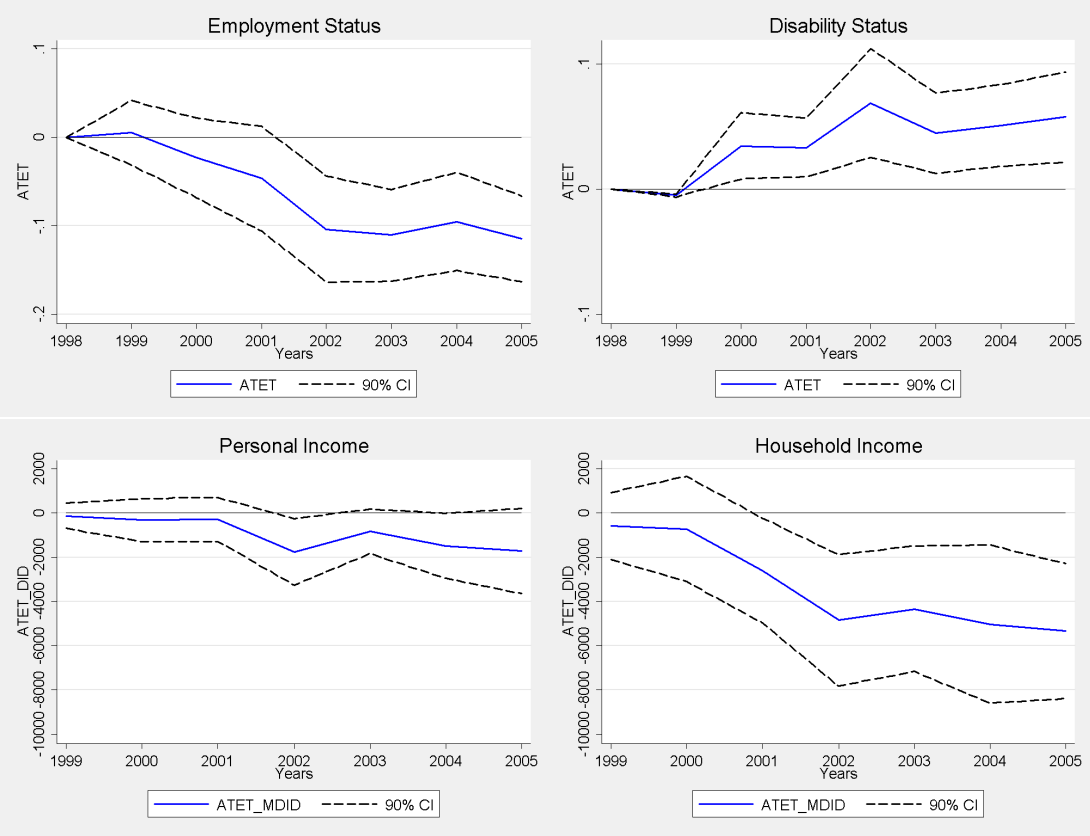
drop in personal and household income is larger for those experiencing more severe shocks is consistent with the story that other household members have to reduce their labour supply to care for the disabled individual. The results for the onset of a severe disease like cancer, circulatory or respiratory disease show that the effects on employment outcomes are very similar to our baseline measure, yet surprisingly the effects on personal and household disposable income are not significant (Table 5.3, column 3).

5.4.2 Are employment and income permanently affected by a health shock?

A permanent drop in income obviously has more serious consequences for household welfare than a transitory impact, necessitating downward revision of life time consumption plans (Deaton, 2002b). It is therefore important to establish to what extent employment and earnings can recover from the effect of a health shock. Using the restricted sample of individuals hospitalised in 1999, we see a significant 10 percentage point fall in the probability of being at work three years later and the probability remains at this lower level until for a further three years (Table 5.6 and Figure 5.1). This provides no evidence of a recovery of employment. The probability of being on DI peaks three years after hospitalisation at almost seven percentage points but the fall from this peak seems to be due to people moving onto social assistance or into retirement rather than returning to work. The loss of personal income also peaks after three years. There is evidence of a recovery in incomes after this point although by the sixth year post hospitalisation the point estimate of the loss of income is almost as large as that in the third year. The effects on household income show no evidence of a recovery. Rather, there is a significant permanent reduction in household income of more than 5,000 Euro six years after hospitalisation. This is much larger than the effect on personal income, again indicating substantial negative spillover effects.

These results are quite different from those presented by Charles (2003) for the US. He found that earnings (not income) fall sharply at the onset of (self-reported) disability and subsequently partially recover. He argues that the recovery can be due to adaptation — investment in human capital that is productive despite the limitations imposed by the health condition. An alternative explanation is that there is recovery of health status itself. In the Netherlands there is little evidence of recovery that might be driven by either mechanism. The most likely explanation for this is the generosity of DI notwithstanding the rationalisation of recent years. Once on DI it is possible to remain there

Figure 5.1: Long Run Average Treatment Effects on the Treated on Labour Force Status and Income



Notes: Authors' calculations on basis of POLS 1998-2001 and RIO 1998-2005. Treatment effects for employment status (top left), disability status (top right), personal income (bottom left), and household income (bottom right)

until retirement and high replacement ratios, which we confirm at 70% of what income would otherwise have been, give little reason to retrain for a new occupation.

5.4.3 Whose income and employment are least protected against health shocks?

The magnitude of the effect of a health shock on employment and income is likely to be heterogeneous across individuals. Most notably, the magnitude of the effect will reflect the severity of the shock, yet in this paper the interest lies more in how the effect differs with socioeconomic status, occupation and employment sector. Still, for a reasonable comparison among these subgroups, severity of the shock needs to be accounted for. We do so by exploiting the Cause-of-Death Register (DO) to discern subsequent mortality

after the shock, and the hospital discharge register (LMR) to observe subsequent hospitalisations, and the primary diagnosis pertaining to the health shock. In case the effect of the health shock on subsequent mortality and hospitalisations, and the primary diagnoses, are similar across subgroups we can be more secure that the effects on employment, benefit receipt, and income are due to genuine differences in preferences and constraints rather than simply reflecting differences in the severity of the health shock.

The poor appear to be affected most by a health shock. Two and three years after hospitalisation, the probability of remaining at work is smaller for those in the bottom quartile of the personal income distribution than it is for those in the top quartile (Table 5.7). The effects are generally not significant, which is probably attributable to the limited sample sizes. Despite being more likely to exit from work, the poor are not more likely to enter DI. In fact, two and three years after hospitalisation, the point estimate of the effect on the transition to DI is larger for those at the top of the income distribution and it is only significant for this group. As a consequence of being less likely to remain in work, yet also less likely to move onto DI, the loss of income, even in absolute terms, is larger for the poor. Three years after hospitalisation their loss of personal income is 2,500 Euro, on average, which is a 21% relative decline in comparison with a 4% decline for those at the top of the distribution (Table 5.9). The relative loss of household income is also larger for the poor at around 10% compared with 4% for the richest members of society, although the average treatment effect is not significant.

Effects of hospitalisation on employment and DI receipt are larger and more significant for the lower educated (no more than lower vocational secondary school) than the higher educated. After three years, hospitalisation reduces the probability of remaining at work by 11 percentage points for those with the lowest education, while there is no significant effect for the higher educated (Table 5.7). A low educated person is eight points more likely to enter DI three years after hospitalisation while the probability is raised by less than four points among the better educated. The absolute drop in personal income shows no consistent difference by education, reflecting the extent to which DI helps cushion the consequences of the greater loss of employment in the low educated group, although in the third year after hospitalisation it is 300 Euro larger for the lower educated and the relative loss of income is 3% higher for this group. The point estimates of the absolute effect on household income tend to be larger for the higher educated and they are only significant for this group. In relative terms, the effects are comparable. From the effects on mortality, future hospitalisations and an inspection of diagnoses (not shown), we conclude that the severity of the shock does not differ much across educational groups.

Both the employment and the incomes of public sector workers are more protected against health shocks than those of their private sector counterparts and the self-employed. Effects on employment are generally not significant, a reflection of the small sample sizes, but three years after hospitalisation the point estimate for public sector workers is half that of private sector employees and a quarter of that of the self-employed (Table 5.8). Hospitalisation does not significantly raise the likelihood of claiming DI for public sector workers, but it clearly does so in the private sector, where the probability rises by 5.6 percentage points after three years. Private sector workers are also significantly more likely to enter retirement as a result of a health shock, whereas this is not true in the public sector. These differences feed through into the effects on incomes. The point estimates suggest that the self-employed incur the largest losses in both personal and household incomes. After three years, the loss of personal income is 2,000 Euro for the average self-employed person with an acute admission to hospital — a relative decrease of 10% (Table 5.9). The loss in household income for this group is 5,000 Euro. Losses are next highest for private sector workers, who, on average, lose 1,600 Euro of personal income and 3,400 Euro of household income (Table 5.8). For public sector workers the point estimates indicate a negligible loss of personal income and even a gain in household income. No effect seems a safe conclusion.

These differential effects do appear to be driven, in part but not wholly, by differences in the severity of the health condition. Public sector workers have lower subsequent hospitalisation rates, yet slightly higher mortality rates. The diagnoses of public and private sector workers are quite different, but one is not necessarily more severe than the other.

5.4.4 Robustness checks

Adjusting the kernel bandwidth from 0.06 to a higher (0.1) and lower bandwidth (0.01) and using a higher number of bootstrap replications (200) gives qualitatively similar results for employment outcomes, but somewhat smaller effects for the treatment effects on both personal and household income (Table 5.10 in the Appendix). Second, we checked sensitivity to the specification of the propensity score, since the effects might be highly sensitive to the set of variables included in the propensity score (Smith and Todd, 2005). Three different specifications of the propensity score were estimated without (sets of) insignificant variables (without occupational characteristics, without income information, and without information on exercising), and the results we obtained were extremely close

to the baseline specification[14]. Finally, average treatment effects on the treated for income were also obtained without the additional differences-in-differences technique. Results were qualitatively the same, and are available upon request.

5.5 Conclusion and discussion

We find that a health shock — measured by an acute hospitalization — significantly and permanently reduces the probability of continued employment. For working-age individuals in the Netherlands the probability of remaining in work falls by more than 7 percentage points 3 years after admission, or 10 percent in relative terms, and the effect remains stable up to 6 years after the health shock. The vast majority of the individuals leaving employment are observed to be claiming disability benefits. In addition, and despite the relative generosity of the Dutch Disability Benefits scheme (OECD, 2003), the concomitant effects on personal and household income are non-negligible. The drop in personal disposable income is around 1,600 Euro on a yearly basis three years after the initial shock. The effects on household disposable income are more than twice as large, reaching an amount of 3,600 Euro three years after the health shock, or around 8 percent in relative terms. The 6 year results indicate that these effects are non-vanishing in the long-run, such that the drop in disposable income can be considered permanent, implying a reduction in household consumption and welfare (Deaton, 2002b; Attanasio and Davis, 1996; Blundell *et al.*, 2008).

We find that, on average, there are important spillover effects to other household members, possibly reflecting a fall in the earnings of the spouse as a result of meeting the care needs of the disabled person. This explanation is supported by the fact that the drop in household income is larger for those experiencing a more severe shock to health. Also, the hypothesis is corroborated by the larger drop in household income for those that move onto DI compared to those that are able to remain at work, for which there are no spillover effects at all. DI is designed to insure personal earnings and it does not offer protection against the impact of disability on the earnings of other household members.

While Charles (1999) finds that wives are likely to increase their labour supply in case their husbands fall ill, and husbands are more likely to reduce their labour supply if their wives receive a health shock, we find exactly the opposite. Spillover effects are found only in the case of illness of a male, which seems to indicate that wives meet the care needs of their husbands and reduce their labour supply to do so. In contrast, when a woman experiences a health shock her personal income falls, but, on average, household income

[14]Results available upon request.

remains constant. It appears that males are able to compensate for the earnings loss of their wives. These findings challenge Charles' (1999) claim that illness induces the spouse to substitute for the previous main activity of the disabled partner.

A second feature of the findings is that the effects are not homogeneous. We find that those who remain at work do not experience a large drop in income — 4 percent in relative terms — while those that transit into disability face a 30 percent relative income loss. Remaining at work thus provides the best protection against income loss after a health shock. This means that recent policy changes focusing on job protection, provided they are successful, are likely to protect incomes as well. However, not all individuals who incur a health shock are capable of continuing to work. Our analysis reveals that individuals experiencing the most severe health shocks are least likely to remain in employment and most likely to move onto DI.

An extremely important finding is that the drop in disposable income following illness varies substantially. We confirm evidence by Charles (2003) that earnings losses are larger for older individuals, possibly since more human capital is destroyed and because they have less incentive to invest in disability specific human capital. But the incentives provided by the Dutch DI system, which favour the elderly, are likely to play a role, as is the fact that older individuals are more likely to experience a more severe health shock. Next, and most importantly, the lowest socio-economic classes — those in the bottom quartile of personal income, and the lower educated — are both the most vulnerable to the health-induced risk of job loss, and their job loss translates into higher relative drops in personal and household disposable income. The lower socioeconomic groups suffer from double jeopardy: they are most likely to suffer a shock to their health (e.g. Smith, 2005c) and they are least protected against the employment and income losses following such a shock. From the fact that, on average, individuals in the lowest income quartile lose around 20% of personal income following a health shock, while those in the top quartile lose only 4%, it follows that ill-health not only raises income inequality, but also places a non-negligible proportion of low-income individuals at risk of poverty.

Differential effects are also found by education. The higher educated are more likely to remain at work following an acute hospital admission than their lower educated peers. However, there are no differences by education in the absolute loss of income following a health shock, although the relative loss is larger among lower educated. This suggests that while the higher educated are better able to remain in employment, they may have to change to a lower paying job, to cut back on hours of work hours, or to miss out on promotion. The lower educated are more likely to exit from working altogether and obtain compensation from Disability Insurance.

Although the small numbers do not allow us to obtain conclusive results regarding the effects on the self-employed, the analysis points out that this group likely faces less protection from formal mechanisms, as their probability of loosing their job due to a health shock is as high as 13 percent, and the income effects are larger, although non-significant. However, one would expect larger negative effects in the later years as the special DI scheme for the self-employed was abolished in 2004, and only voluntary insurance is currently available. This is a particular worrisome result as this group was also found not to save enough for their retirement in the Netherlands (van der Lecq and Oerlemans, 2009), which enhances their vulnerability to a health shock.

Last, but not least, we find public sector workers are most protected. Since the employer is less exposed to competition, and the salary is more regulated and less tied to productivity, individuals working in the public sector are far less likely to lose their jobs after a health shock, nor do they face a decrease of disposable income.

The fact that the lower socioeconomic groups suffer double jeopardy is clearly a major concern. Those most susceptible to receiving a health shock, are least protected against a fall in income in case one arrives. To tackle this problem, it would be more efficient, but not easy, to reduce health shocks in this vulnerable group rather than trying to mitigate the effects of a health shock by increasing replacement rates, given the moral hazard this would induce. Whether the average actual replacement rates are below or above the desired level, and should provoke any concern for policy-makers, is arguable. At 70% of personal income for those that move onto DI, it seems unlikely that many would consider the Dutch system ungenerous. But the spillover effects on the incomes of other household members warrant further investigation. If these reflect a voluntary increase in joint leisure (see e.g. Hurd, 1990), then there is less concern for household welfare, although the moral hazard effect of DI on the labour supply of partners should be considered. On the other hand, if spouses are forced to decrease their labour participation in order to meet care giving responsibilities, then the impact on household welfare is greater and a reassessment of current policy is warranted.

The most important limitation of our study is limited power due to small sample sizes. Given our datasets and empirical strategy, there is trade-off between controlling for a rich set of observable characteristics to satisfy the conditional independence assumption by matching on the greatest possible number of variables, while at the same time ensuring that there are sufficient observations left in our treatment group. In this study we err on the side of caution by including as many characteristics as possible, at the expense of a smaller treatment group. An important advantage is that given this rich set of control

variables the results are extremely robust to all specification checks. Still, replication on a larger dataset would be informative.

Tables

Table 5.1: Descriptive statistics of the evolution of percentages in employment outcomes and average disposable income by treatment status

	Treated	Control		Treated	Control
Working			**Retired**		
$t-1$	1.000	1.000	$t-1$	0.000	0.000
t	0.877	0.893	t	0.008	0.005
$t+1$	0.816	0.830	$t+1$	0.021	0.012
$t+2$	0.738	0.782	$t+2$	0.037	0.021
$t+3$	0.660	0.742	$t+3$	0.056	0.029
Unemployed			**No Income**		
$t-1$	0.000	0.000	$t-1$	0.000	0.000
t	0.003	0.004	t	0.038	0.034
$t+1$	0.010	0.006	$t+1$	0.031	0.048
$t+2$	0.016	0.009	$t+2$	0.047	0.048
$t+3$	0.024	0.011	$t+3$	0.042	0.047
Disabled			**Personal Income**		
$t-1$	0.000	0.000	$t-1$	19,297	18,348
t	0.013	0.005	t	19,990	19,514
$t+1$	0.044	0.009	$t+1$	20,710	20,564
$t+2$	0.052	0.014	$t+2$	21,384	21,532
$t+3$	0.069	0.016	$t+3$	20,594	21,605
Social Security			**Household Income**		
$t-1$	0.000	0.000	$t-1$	40,845	40,415
t	0.005	0.002	t	41,176	41,883
$t+1$	0.000	0.003	$t+1$	42,310	43,193
$t+2$	0.005	0.003	$t+2$	42,394	44,643
$t+3$	0.013	0.004	$t+3$	40,387	43,622

Notes: Average percentages in the listed states for treatment (N=398) and control group (N=32,085). Year $t-1$ is the base year, at which all individuals were working. In year t the shock found place.

Table 5.2: Descriptive statistics and coefficients of the variables included in the propensity score

Variable	Treated	Control	Coefficient	Dataset
Demographics				
Age 18-24 – Males	0.09	0.10	−0.12	RIO
Age 25-34 – Males	0.12	0.13	−0.11	RIO
Age 35-44 - Males	0.18	0.15		RIO
Age 45-54 - Males	0.18	0.15	0.00	RIO
Age 55-64 - Males	0.10	0.05	0.15	RIO
Age 18-24 - Females	0.05	0.08	−0.07	RIO
Age 25-34 - Females	0.07	0.10	−0.04	RIO
Age 35-44 - Females	0.09	0.12		RIO
Age 45-54 - Females	0.09	0.10	0.04	RIO
Age 55-64 – Females	0.03	0.02	0.19	RIO
Men	0.67	0.58	0.37**	GBA
Married	0.66	0.65		GBA
Never married	0.20	0.25	−0.04	GBA
Widow	0.01	0.01	−0.10	GBA
Divorced	0.10	0.06	0.17**	GBA
Household members	3.32	3.38	0.00	RIO
House owner	0.66	0.72	−0.10	POLS
Ethnicity				
Native	0.87	0.85		GBA
Non-native Non-Western	0.04	0.04	−0.14	GBA
Non-native Western	0.07	0.07	−0.05	GBA
Wave				
Wave 1998	0.42	0.44	−0.08	POLS
Wave 1999	0.25	0.23		POLS
Wave 2000	0.21	0.21	−0.04	POLS
Wave 2001	0.13	0.12	−0.05	POLS
City Size				
< 10,000	0.05	0.07		GBA
10-50,000	0.54	0.54	0.05	GBA
50-100,000	0.15	0.15	0.10	GBA
> 100,000	0.25	0.25	0.09	GBA

Table 5.2: (continued)

Variable	Treated	Control	Coefficient	Dataset
Province				
Groningen	0.05	0.04	0.03	GBA
Friesland	0.07	0.08		GBA
Drenthe	0.05	0.04	0.01	GBA
Overijssel	0.06	0.08	-0.21^{**}	GBA
Flevoland	0.03	0.02	0.02	GBA
Gelderland	0.13	0.12	-0.06	GBA
Utrecht	0.05	0.06	-0.21^{*}	GBA
Noord-Holland	0.10	0.13	-0.17^{*}	GBA
Zuid-Holland	0.14	0.17	-0.18^{**}	GBA
Zeeland	0.05	0.03	0.15	GBA
Noord-Brabant	0.17	0.15	-0.03	GBA
Limburg	0.10	0.08	-0.02	GBA
Occupation				
Regular employee	0.77	0.8	-0.10	RIO
Civil Servant	0.08	0.09	0.10	RIO
CEO/Executive	0.01	0.01	-0.20	RIO
Self-Employed	0.14	0.10		RIO
Hours work/week	33.30	31.79	0.00	POLS
Public sector	0.19	0.20	-0.02	POLS
Private sector	0.40	0.39	-0.11^{*}	POLS
Unknown sector	0.41	0.40		POLS
Blue collar	0.24	0.19	0.09	POLS
White collar	0.35	0.38	0.00	POLS
Unknown collar	0.40	0.43		POLS
Education				
Primary school	0.12	0.08	0.03	POLS
Lower Secondary school	0.26	0.21		POLS
Higher Secondary school	0.33	0.33	0.02	POLS
Higher education	0.11	0.17	-0.19^{**}	POLS
Income				
Income ratio (personal / household income)	0.54	0.50	-0.01	RIO
Equivalent household income	23,641.30	23,331.56	0.00	RIO

Table 5.2: (continued)

Variable	Treated	Control	Coefficient	Dataset
Self-Reported Health				
Good Self-Assessed Health	0.79	0.89	−0.07	POLS
Fair Self-Assessed Health	0.13	0.08		POLS
Poor Self-Assessed Health	0.08	0.03	0.19	POLS
Hampered in daily actvs	0.17	0.09	0.17***	POLS
GP visit	0.22	0.22	−0.01	POLS
Life-style variables				
Smoking	0.41	0.35	0.04	POLS
Sports Frequently	0.14	0.12	0.09	POLS
Sports Modestly	0.34	0.41	−0.01	POLS
Sports Infrequently	0.04	0.04		POLS
Sports Never	0.49	0.43	0.03	POLS

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Column 1 and 2 present the means of the different variables separately for treatment (N=398) and control group (N=32,085), respectively. Column 3 presents the coefficients of the probit estimation of the propensity score. In addition to the variables listed, we included interactions between gender and civil-servants (coefficient -0.39**), being a house owner (coefficient -0.03), and self-assessed health (coefficient -0.15 for good self-assessed health, and -0.11 for poor self-assessed health). Column 4 indicates from which data set the variables were retrieved.

Table 5.3: Average Treatment Effects on the Treated of an Acute Hospitalization on Employment status and Disposable Income

Health Shock	**Acute**	**Acute + # Nights > 3**	**Diagnosis**
Probability of working			
t	-0.014	-0.019	-0.024
$t+1$	-0.009	-0.026	-0.048^{***}
$t+2$	-0.037^{**}	-0.055^{*}	-0.059^{***}
$t+3$	-0.076^{***}	-0.096^{***}	-0.075^{***}
Probability of being unemployed			
t	-0.002	-0.004	-0.002
$t+1$	0.004	0.007	0.011
$t+2$	0.007	0.009	0.006
$t+3$	0.012	0.020^{*}	0.014^{*}
Probability of being on disability benefits			
t	0.006	0.012	0.017^{**}
$t+1$	0.035^{***}	0.058^{***}	0.028^{***}
$t+2$	0.038^{***}	0.063^{***}	0.041^{***}
$t+3$	0.052^{***}	0.085^{***}	0.047^{***}
Probability of being on Social Security benefits			
t	0.003	0.007	0.001
$t+1$	-0.003	-0.003^{***}	0.002
$t+2$	0.002	0.006	-0.001
$t+3$	0.010^{*}	0.010	0.004
Probability of being retired			
t	0.002	-0.001	0.002
$t+1$	0.009	0.015	0.008
$t+2$	0.015	0.033^{**}	0.021^{**}
$t+3$	0.026^{*}	0.044^{***}	0.034^{***}
Probability of having no income			
t	0.004	0.002	0.011
$t+1$	-0.018^{*}	-0.028^{***}	-0.004
$t+2$	-0.002	-0.014	-0.006
$t+3$	-0.006	-0.022^{***}	-0.013
Personal income (in 1000 Euro)			
t	-0.458	-0.987	0.373
$t+1$	-0.796^{*}	0.006	-0.283
$t+2$	-0.903^{**}	-0.468	-0.558
$t+3$	-1.631^{***}	-1.247^{**}	-0.074
Household income (in 1000 Euro)			
t	-1.229	0.124	-0.326
$t+1$	-1.882^{***}	-0.067	-0.982
$t+2$	-2.863^{***}	-1.623	-2.291
$t+3$	-3.634^{***}	-3.327^{**}	-1.308

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Estimates based on POLS 1998-2001 sample with follow-up of RIO 1998-2005, N=31,882 of which 395 are in the treatment group. Column 1 gives the base case using acute hospitalizations (NT = 395), Column 2 uses the combination of acute and number of nights more than three (NT = 231) and Column 3 uses diagnoses based on cancer, respiratory, or circulatory diseases as indicator (NT = 415).

Table 5.4: Average Treatment Effects on the Treated by Sex and Age.

	Males	**Females**	**Age $<$ 50**	**Age $\geq$ 50**
Probability of working				
t	0.000	-0.050	-0.007	-0.019
$t+1$	-0.001	-0.036	0.001	-0.026
$t+2$	-0.051^{*}	-0.020	-0.016	-0.082
$t+3$	-0.079^{***}	-0.082^{*}	-0.042	-0.148^{***}
Probability of being unemployed				
t	0.000	-0.005	0.000	-0.009^{***}
$t+1$	0.006	0.000	0.006	-0.004
$t+2$	0.011	-0.002	0.007	0.002
$t+3$	0.008	0.020	0.015^{*}	-0.001
Probability of being on disability benefits				
t	0.015	-0.007^{***}	0.006	0.011
$t+1$	0.032^{**}	0.044^{*}	0.027^{***}	0.054^{**}
$t+2$	0.044^{***}	0.030^{*}	0.017^{*}	0.098^{***}
$t+3$	0.062^{***}	0.036^{*}	0.029^{***}	0.117^{***}
Probability of being on Social Security benefits				
t	0.007	-0.003^{***}	0.005	-0.001^{**}
$t+1$	-0.002^{***}	-0.003^{***}	-0.003^{***}	-0.002^{***}
$t+2$	0.001	0.004	0.000	0.008
$t+3$	0.005	0.02	0.007	0.019
Probability of being retired				
t	0.004	-0.003^{***}	-0.000^{***}	0.002
$t+1$	0.015	-0.007^{***}	-0.001^{***}	0.023
$t+2$	0.026^{*}	-0.014^{***}	0.003	0.028
$t+3$	0.033^{**}	0.006	0.002	0.063
Probability of having no income				
t	-0.006	0.032	0.003	0.005
$t+1$	-0.020^{***}	0.001	-0.015	-0.012
$t+2$	0.004	-0.007	0.003	-0.012
$t+3$	-0.001	-0.008	-0.002	-0.017
Personal income (in 1000 Euro)				
t	-0.133	-1.238^{***}	-0.549	0.230
$t+1$	-0.714	-1.112^{**}	-0.778^{*}	-0.307
$t+2$	-0.898^{*}	-1.170^{**}	-0.876^{*}	-0.367
$t+3$	-1.650^{***}	-1.807^{***}	-1.435^{***}	-1.192
Household income (in 1000 Euro)				
t	-0.842	-2.120^{***}	-1.256^{*}	-0.734
$t+1$	-1.877^{**}	-1.961	-1.895	-0.909
$t+2$	-3.123^{**}	-2.437	-2.604^{**}	-2.408
$t+3$	-4.787^{***}	-1.442	-2.722^{**}	-4.677^{**}

* p-value $<$ 0.1, ** p-value $<$ 0.05, *** p-value $<$ 0.01

Notes: Number of treated individuals (NT) for males is 267, for females NT=131, for those under 50 NT=292, while for those over 50 NT=106. Data as in Table 5.3.

Table 5.5: Average Treatment Effects on Disposable Income (in 1000 Euro) for those that stay at work, and for those that enter Disability Insurance schemes.

	Remain at Work	**Disability Recipients**
Disposable Personal Income		
t	-0.113	-1.487^{*}
$t+1$	0.151	-3.385^{***}
$t+2$	-0.112	-4.420^{***}
$t+3$	-1.031^{*}	-6.565^{***}
Disposable Household Income		
t	-0.601	-2.825^{*}
$t+1$	0.499	-4.476^{*}
$t+2$	-0.534	-6.414^{*}
$t+3$	-1.819	-8.756^{***}

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Column 1 presents the results for those that remain at work (NT=235), while column 2 presents the results for those that rely on disability insurance in any of the years (NT=108). Data as in Table 5.3.

Table 5.6: Long-Run Average Treatment Effects on the Treated of an Acute Hospitalization on Employment status and Disposable Income

Probability of working		**Probability of being retired**	
1999	−0.005	1999	−0.005***
2000	−0.023	2000	0.008
2001	−0.047	2001	0.015
2002	−0.104***	2002	0.009
2003	−0.111***	2003	0.027
2004	−0.095***	2004	0.009
2005	−0.114***	2005	0.006
Probability of being unemployed		**Probability of having no income**	
1999	−0.004	1999	0.010
2000	0.001	2000	−0.006
2001	0.001	2001	0.009
2002	0.010	2002	0.007
2003	0.014	2003	−0.007
2004	0.009	2004	0.008
2005	0.027*	2005	0.001
Prob being on disability benefits		**Personal income (in 1000 Euro)**	
1999	−0.005	1999	−0.116
2000	0.035**	2000	−0.309
2001	0.033**	2001	−0.282
2002	0.069***	2002	−1.749**
2003	0.044**	2003	−0.820
2004	0.051***	2004	−1.478*
2005	0.058***	2005	−1.706
Prob being on Social Security benefits		**Household income (in 1000 Euro)**	
1999	0.004	1999	−0.575
2000	−0.002	2000	−0.721
2001	0.010	2001	−2.599*
2002	0.015	2002	−4.840***
2003	0.028**	2003	−4.322***
2004	0.015	2004	−5.012**
2005	0.008	2005	−5.328***

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Estimates based on POLS 1998 sample with follow-up of RIO 1998-2005, N=14,380 of which 235 are in the treatment group.

Table 5.7: Average Treatment Effects on the Treated by Personal Income and Education

	Bottom Quartile	Top Quartile	Lower Educated	Higher Educated
Probability of working				
t	-0.028	-0.012	-0.038	0.020
$t+1$	0.017	0.002	-0.021	0.031
$t+2$	-0.059	-0.036	-0.064	0.012
$t+3$	-0.079	-0.055^{*}	-0.113^{**}	-0.018
Probability of being unemployed				
t	-0.004^{***}	0.004	-0.008	0.003
$t+1$	0.008	0.002	-0.003	0.007
$t+2$	0.020	-0.010^{***}	0.015	-0.002
$t+3$	0.004	-0.012^{***}	0.010	0.014
Probability of being on disability benefits				
t	-0.005	0.006	0.024	-0.005
$t+1$	0.041^{*}	0.012	0.065^{***}	0.017
$t+2$	0.038	0.046^{**}	0.068^{***}	0.026^{*}
$t+3$	0.035	0.053^{***}	0.082^{***}	0.036^{**}
Probability of being on Social Security benefits				
t	0.021	0.000	0.004	0.005
$t+1$	-0.005^{***}	0.000	-0.004^{***}	-0.001^{***}
$t+2$	0.019	0.000	0.002	-0.002^{***}
$t+3$	0.032	0.000	0.016	-0.002^{***}
Probability of being retired				
t	-0.002^{***}	0.004	-0.001	0.007
$t+1$	-0.004^{***}	0.027	0.004	0.013
$t+2$	-0.011^{***}	0.030	0.026	0.010
$t+3$	0.011	0.033	0.060^{***}	0.010
Probability of having no income				
t	-0.015	0.013	0.007	-0.005
$t+1$	-0.045	-0.008^{***}	-0.029^{*}	-0.012
$t+2$	0.015	-0.005^{***}	0.018	0.006
$t+3$	0.011	-0.005^{***}	-0.033^{**}	0.013
Personal income (in 1000 Euro)				
t	-0.318	-0.304	-0.308	-0.589
$t+1$	-1.113^{**}	-0.288	-0.859	-1.083
$t+2$	-1.195	-0.834	-0.725	-1.490^{**}
$t+3$	-2.436^{***}	-1.244	-1.899^{***}	-1.579^{**}
Household income (in 1000 Euro)				
t	1.030	-0.884	-0.786	-1.620
$t+1$	-3.587	0.019	-1.710	-2.965^{**}
$t+2$	-3.737	-1.348	-1.464	-4.601^{***}
$t+3$	-4.534	-1.893	-3.390	-3.460^{**}

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Bottom and Top Quartile are based upon Personal Disposable Income. Number of treated (NT) individuals (NT) for the bottom quartile of personal income is 83, while for the top quartile it is 116. For the lower educated (primary school and lower vocational education) NT=150, for the higher educated (higher secondary + higher vocational + university education) NT=174. Data as in Table 5.3.

Table 5.8: Average Treatment Effects on the Treated by Occupational Sector

	Public Sector	Private Sector	Self-Employed
Probability of working			
t	-0.010	0.007	-0.042
$t+1$	0.004	0.003	-0.007
$t+2$	0.001	-0.022	-0.034
$t+3$	-0.025	-0.052	-0.105^{*}
Probability of being unemployed			
t	-0.003^{***}	0.003	-0.001
$t+1$	-0.004^{***}	0.006	0.019
$t+2$	-0.005^{***}	0.008	-0.001
$t+3$	-0.009^{***}	0.030^{*}	-0.002^{*}
Probability of being on disability benefits			
t	-0.002^{***}	0.007	-0.004^{***}
$t+1$	-0.008^{***}	0.043^{***}	0.004
$t+2$	0.001	0.039^{**}	0.037
$t+3$	0.029	0.056^{***}	0.051
Probability of being on Social Security benefits			
t	0.000	0.006	0.017
$t+1$	-0.001^{**}	-0.002^{***}	-0.006^{***}
$t+2$	-0.001^{**}	-0.002^{***}	0.017
$t+3$	-0.002^{**}	0.003	0.038
Probability of being retired			
t	-0.003^{***}	-0.004^{***}	-0.005^{**}
$t+1$	-0.008^{***}	0.009	0.030
$t+2$	-0.003	0.034^{**}	0.012
$t+3$	0.018	0.026^{*}	0.002
Probability of having no income			
t	-0.007^{***}	0.007	0.060
$t+1$	-0.010^{***}	-0.010	0.014
$t+2$	-0.012^{***}	-0.008	0.010
$t+3$	-0.014^{***}	-0.015^{***}	-0.016
Personal income (in 1000 Euro)			
t	0.056	0.275	-2.378
$t+1$	0.388	-1.003	-1.692
$t+2$	-0.293	-1.054	-1.227
$t+3$	-0.076	-1.595^{*}	-2.034
Household income (in 1000 Euro)			
t	-0.429	-0.366	-6.466^{*}
$t+1$	2.304	-2.077^{*}	-5.129^{*}
$t+2$	0.613	-2.423	-5.343
$t+3$	0.410	-3.402^{*}	-4.950

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Number of treated individuals (NT) for public sector workers (governmental institutions, education, health services, environmental services, culture, recreation, defense) is 76, for private sector workers (all other sectors) is 158, and for the self-employed NT = 55. Data as in Table 5.3.

Table 5.9: Relative Average Treatment Effects on the Treated in year $t+3$

	Working	**Disability**	**Personal Income**	**Household Income**
All	−10.26%	309.48%	−7.25%	−8.36%
Males	−10.39%	474.19%	−6.24%	−10.89%
Females	−11.37%	168.50%	−12.41%	(−3.37%)
Age <50	(−5.47%)	229.85%	−6.47%	−6.19%
Age ≥50	−23.83%	315.30%	(−5.44%)	−11.60%
Stay at work			−4.32%	−4.06%
Disabled			−30.58%	−18.55%
Bottom Quartile	−13.49%	199.79%	−21.44%	−10.32%
Top Quartile	−6.85%	441.44%	(−3.85%)	(−3.93%)
Lower Educated	−16.61%	329.12%	−9.48%	(−8.37%)
Higher Educated	(−2.38%)	280.47%	−6.63%	−7.70%
Public sector	(−3.01%)	(190.64%)	(−0.04%)	(0.01%)
Private sector	(−6.66%)	334.42%	−6.60%	−8.20%
Self-employed	−13.21%	144.22%	(−9.77%)	(−12.18%)

Notes: The Relative Treatment Effects are calculated on basis of equation (5.2). Relative Treatment effects between parenthesis represent those of which the $ATET$ is non-significant. Strictly speaking, since the numerator of the $RATET$ is statistically indistinguishable (at 10%) from zero, the ratio is zero too.

Table 5.10: Robustness Checks

	Default Bandwidth	Small Bandwidth	Large Bandwidth	More bootstrap reps
Probability of working				
t	−0.014	−0.008	−0.014	−0.014
$t+1$	−0.009	0.001	−0.010	−0.009
$t+2$	−0.037**	−0.024	−0.037*	−0.037**
$t+3$	−0.076***	−0.060***	−0.077***	−0.076***
Probability of being unemployed				
t	−0.002	−0.003	−0.002	−0.002
$t+1$	0.004	0.003	0.004	0.004
$t+2$	0.007	0.005	0.007	0.007
$t+3$	0.012	0.010	0.012	0.012
Probability of being on disability benefits				
t	0.006	0.004	0.007	0.006
$t+1$	0.035***	0.028***	0.035***	0.035***
$t+2$	0.038***	0.032***	0.039***	0.038***
$t+3$	0.052***	0.046***	0.053***	0.052***
Probability of being on Social Security benefits				
t	0.003	0.003	0.003	0.003
$t+1$	−0.003	−0.004	−0.003	−0.003
$t+2$	0.002	0.002	0.002	0.002
$t+3$	0.010*	0.009*	0.010*	0.010*
Probability of being retired				
t	0.002	0.000	0.002	0.002
$t+1$	0.009	0.004	0.009	0.009
$t+2$	0.015	0.005	0.015*	0.015
$t+3$	0.026*	0.014	0.027***	0.026*
Probability of having no income				
t	0.004	0.007	0.005	0.004
$t+1$	−0.018*	−0.012	−0.016*	−0.018*
$t+2$	−0.002	0.002	−0.002	−0.002
$t+3$	−0.006	−0.001	−0.007	−0.006
Personal income (in 1000 Euro)				
t	−0.458	−0.133	0.438	−0.458
$t+1$	−0.796*	−0.235	0.110	−0.796*
$t+2$	−0.903**	−0.421	−0.158	−0.903**
$t+3$	−1.631***	−1.135*	−1.034*	−1.631***
Household income (in 1000 Euro)				
t	−1.229	−0.271	−0.864	−1.229
$t+1$	−1.882***	−0.188	−1.012	−1.882***
$t+2$	−2.863***	−1.195	−2.163**	−2.863***
$t+3$	−3.634***	−1.785	−3.183***	−3.634***

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Column 1 gives the base case using acute hospitalizations, and an Epanechnikov kernel with bandwidth 0.06. Column 2 uses a smaller bandwidth of only 0.01. Column 3 uses a larger bandwidth of 0.10, while column 4 is similar to the base case but uses more bootstrap replications (200 instead of 50). Data as in Table 5.3.

Part III

Underlying Mechanisms

Chapter 6

Healthy, Wealthy and Wise: a Life Cycle Model of Socioeconomic Disparities in Health

Understanding of the substantial disparity in health between low and high socioeconomic status (SES) groups is hampered by the lack of a comprehensive theoretical framework to interpret empirical facts and to predict yet untested relations. We present a life-cycle model that incorporates multiple mechanisms that explain (jointly) a large part of the observed disparities in health by SES. In our model, lifestyle factors (preventive care, healthy and unhealthy consumption), working conditions (physical and psychosocial health stresses), living conditions (housing, neighborhood social environment) and curative care are mechanisms through which SES (education, income, wealth) and health are related. Our model predicts that greater initial wealth, permanently higher income (over the life cycle) and a higher level of education induce individuals to invest more in curative and in preventive care, shift consumption toward healthy consumption and enable individuals to afford healthier working and living environments. Jointly these gradually lead to cumulative health advantage with age.

This chapter is based upon:

Galama, T.J. and H. van Kippersluis (2010), "Healthy, Wealthy and Wise: a Life Cycle Model of Socioeconomic Disparities in Health", Mimeo RAND Corporation, Santa Monica.

6.1 Introduction

Disparities in health across socioeconomic status (SES) groups — often called the SES health gradient — are substantial. For example, Case and Deaton (2005) show how in the United States, a 60 year old high-income (top quartile of family income) male, on average, reports to be in similar health to a 20 year old low-income male (bottom quartile). Not only do low SES individuals start adulthood with a lower level of health but their health also deteriorates faster with age than their high SES peers. Similar patterns hold for other measures of SES, such as education and wealth and other indicators of health, such as onset of chronic diseases, disability and mortality (e.g., Adler *et al.*, 1994; Smith, 1999; Marmot, 1999; Van Doorslaer *et al.*, 2008). Low SES individuals are more likely to suffer from disability and have a significantly shorter life expectancy. In cross sectional data the disparity in health between low and high SES groups appears to increase over the life cycle until ages 50-60, after which it narrows. This pattern is strikingly consistent across countries with relatively low levels of protection from loss of work and health risks, such as the US, and those with stronger welfare systems, such as the Netherlands (House *et al.*, 1994; Kunst and Mackenbach, 1994; Preston and Elo, 1995; Smith, 1999; 2005a; 2007; Case and Deaton, 2005; Van Kippersluis *et al.*, 2010).

The significant social and economic patterning of disease suggests that social interventions have great potential for improving the health of, in particular, disadvantaged groups (CSDH, 2008). Addressing health inequalities requires a detailed understanding of the complex interaction between dimensions of SES and those of health. Recent significant contributions to the understanding of socioeconomic disparities in health have concentrated on the identification of causal effects, but have stopped short of uncovering the underlying mechanisms that produce the causal relationships. For example, education is found to have a causal protective effect on health (Lleras-Muney, 2005; Oreopoulos, 2006; Silles, 2009) but it is not known exactly how the more educated achieve their health advantage.

Understanding of the relative importance of underlying mechanisms responsible for the observed relationships is hampered by the lack of a sufficiently comprehensive theory. Case and Deaton (2005) argue that it is extremely difficult to understand the relationships between health, education, income, wealth and labor-force status without some guiding theoretical framework. A good theoretical framework ought to guide empirical research not only in testing causal effects but also in revealing the underlying mechanisms. Without such knowledge it is difficult to design policies that are effective in reducing disparities (Deaton, 2002a). Integrating the roles of proposed mechanisms and their long-term effect

into a comprehensive framework allows researchers to assess the relative importance of each proposed mechanism, the interaction between mechanisms, and to disentangle the differential patterns of causality. It is no surprise then that several authors (e.g., Case and Deaton, 2005; Cutler *et al.*, 2008) have pointed to the absence of a unifying theory of SES and health over the life cycle and have emphasized the importance of developing one.

A suitable framework in which multiple mechanisms and their cumulative long-term effects can be studied is a structural model of SES and health over the life cycle. Health disparities, as well as SES differences (e.g., wealth) accumulate over the life course, and are considerably larger at later ages. Case and Deaton (2005) have attempted to develop a model for the role of work and consumption behavior in explaining the SES and health gradient. Their starting point is the canonical life cycle model of the demand for health and medical care, due to Grossman (1972a, 1972b). Case and Deaton (2005) present a simplified Grossman model and extend the model to include the detrimental effect of hard/risky labor and of unhealthy consumption behavior on health. However, the authors find that the Grossman model is not able to explain a number of the most salient features of the SES health gradient. For example, Case and Deaton (2005) claim that while the Grossman model can explain differences in the *level* of health between low and high SES groups it cannot explain differences in the *rate* of health deterioration. In other words, it cannot account for the widening of the SES health gradient with age through late middle age or early late life. Other problems with the predictions and properties of the Grossman model have been pointed out in the literature (see Grossman, 2000, for a review and rebuttal of these).

The aim of this paper is to develop a conceptual framework for health and socioeconomic status over the life cycle. Our starting point is the Grossman model (Grossman 1972a, 1972b) and the extended version presented by Case and Deaton (2005). Our contribution is as follows. First, we address three significant problems identified in the literature with the Grossman model by introducing decreasing-returns-to-scale (DRTS) in the health production function (as in Ehrlich and Chuma, 1990). This addresses (i) the indeterminacy problem ("bang-bang" solution) for investment in medical care, (ii) the relative inability of the model to reproduce the observed negative relation between health and the demand for medical care (e.g., Zweifel and Breyer, 1997)[1], and (iii) the model's

[1]It is not entirely correct to assert that the Grossman model always produces the incorrect sign for the relationship between health and investment in curative care. As Grossman (2000; p. 369) shows, for the pure investment model and assuming that the "natural" deterioration rate increases with age (a necessary assumption for the health stock to decline with age in Grossman's formulation [not so for a

lack of capacity to explain differences in the health deterioration rate (not just the level of health) between different socioeconomic groups (Case and Deaton, 2005). Addressing these problems has been essential: a DRTS health production function can account for a greater number of observed empirical patterns and suggests that the Grossman model provides a suitable foundation for the development of a life-cycle model of the SES-health gradient.

Yet, utilization of medical services and access to care explain only part of the association between SES and health (e.g., Adler *et al.*, 1993). Our second contribution is therefore to incorporate many potential mechanisms in the model that could explain disparities in health by SES and to include a multitude of potential bi-directional pathways between health and dimensions of SES. One important concept in our work is "job-related health stress", which can be interpreted broadly and can range from physical working conditions (e.g., hard labor) to the psychosocial aspects of work (e.g., low status, limited control, repetitive work, etc). The notion here is that job-related health stress can include any aspect of work that is detrimental to health and as such is associated with a wage premium (a compensating wage differential). Other important features of the model are lifestyle factors (preventive care, healthy and unhealthy consumption), and curative (medical) care. The model integrates a life cycle approach, and the concepts of financial, human and health capital (Muurinen and Le Grand, 1985). The focus is on understanding the SES-health gradient as the outcome of rational (constrained) individual behaviour, and the framework applies to individuals who have completed their education and participate (or have participated) in the labor-force.

We find that greater initial wealth, permanently higher income (over the life cycle) and a higher level of education induce individuals to invest more in curative and in preventive care, shift consumption toward healthy consumption, and enable individuals to afford healthier working environments (associated with lower levels of physical and psychosocial health stresses) and living environments. The mechanism through which initial wealth, permanent income and education operate is by increasing the marginal cost of, and demand for, curative care. The greater marginal cost of curative care in turn increases the health benefit of (and hence demand for) preventive care and healthy consumption, and the health cost of (and hence reduced demand for) unhealthy working

DRTS formulation]), investment in curative care increases with age while the health stock falls if the elasticity of the marginal production benefit of health with respect to health is less than one (Grossman refers to this as the MEC schedule). This produces a negative correlation between health and medical care. In a DRTS formulation however the relation between health and investment in curative care is more intuitive and follows directly from the first-order condition for health.

and living environments, and unhealthy consumption. Jointly these gradually lead to cumulative health advantage with age. Our model thus holds considerable promise in explaining empirical health patterns. Such a model has not been available before and economists have highlighted the significance of its development (e.g., Case and Deaton, 2005; Cutler *et al.*, 2008).

The paper is organized as follows. Section 6.2 briefly reviews the literature on health disparities by SES to determine the essential components required in a comprehensive framework. The relation between SES and health is complex and developing a theory requires simplification and a focus on the essential mechanisms relating SES and health. We highlight potential explanations for the SES health gradient that a) explain a large part of the gradient and b) are relatively straightforward to include in our theoretical framework. Based on these findings we develop our theoretical formulation. Section 6.3 presents and discusses first-order conditions and the characteristics of the model solutions. The section also highlights potential mechanisms through which SES and health influence each other, how the socioeconomic gradient in health emerges, and discusses the role of institutions. In section 6.4 we discuss the importance of mechanisms that were not included, potential ways of including additional mechanisms and provide direction for future research and theoretical development. In section 6.5 we conclude.

6.2 Components of a model capturing the SES-health gradient

In this section we review the literature on health disparities by SES to determine the essential components required in a comprehensive framework (section 6.2.2). Based on these findings we extend and refine prior theoretical work and present our theoretical formulation (section 6.2.3).

6.2.1 Background

A significant body of research across multiple disciplines (including epidemiology, sociology, demography, psychology, evolutionary biology and economics) has been devoted to documenting and explaining the substantial disparity in health between low and high socioeconomic status (SES) groups. Progress has been made in recent years in characterizing the relationships between the various dimensions of SES and health over the life cycle and in understanding the relative importance and directions of causal pathways. Epidemiological research has used longitudinal studies to examine the role of behavioral,

material, psychosocial and healthcare related pathways in explaining SES-health associations (Avendano *et al.*, 2006a; 2006b; House *et al.*, 1990; 1994; Huisman *et al.*, 2008; Lantz *et al.*, 1998; Lynch *et al.*, 1997; Marmot *et al.*, 1997a; Skalicka *et al.*, 2009; Van Lenthe *et al.*, 2002; van Oort *et al.*, 2005; Yen and Kaplan, 1999). Economists have recently re-emphasized the importance of the reverse impact of health on SES through ability to work (Smith, 1999; 2005a; 2007; Case and Deaton, 2005). These studies suggest that education is the key dimension of SES for which there appears to be robust evidence of a substantial causal protective effect on health and that an important part of the health differences by financial indicators of SES can be explained by the fact that bad health impinges on the ability to work, thereby reducing income. Further, these studies highlight the importance of health behaviors (such as smoking, drinking and exercise), curative and preventive care, psychosocial and environmental risk factors, neighborhood social environment, acute and chronic psychosocial stress, social relationships and supports, sense of control, fetal and early childhood conditions, and physical, chemical, biological and psychosocial hazards and stressors at work.

6.2.2 Key components of a conceptual framework

The life cycle:
Health is modelled as a stock that deteriorates over the life cycle and its deterioration can be counteracted by health investments. A natural starting point for a theory of the relation between health and SES is a model of life cycle utility maximization. Since we are interested in explaining health and SES over the life cycle, the model should include health, education, income, assets (wealth), and health investment as a function of age.

Financial capital, human capital and health capital:
Muurinen and Le Grand (1985) emphasize that there are three types of capital: financial capital (assets), human capital (education, knowledge, experience) and health capital (our bodies). The Grossman model contains financial capital and provides a framework for the concept of health capital. We include human capital by assuming a Mincer-type wage equation in which the more educated and more experienced earn higher wages (Mincer, 1974). We also allow the more educated to be more efficient consumers and producers of curative (medical) and preventive care (based on the interpretation, as in Grossman, of education as a productivity factor in own time inputs and in identifying and seeking effective care).

Working environment and lifestyle factors:
Low SES individuals more often perform risky, manual labor than high SES individuals, and their health deteriorates faster as a consequence (Marmot *et al.*, 1997b; Schrijvers *et al.*, 1998; Borg and Kristensen, 2000). Case and Deaton (2005) find that those who are employed in manual occupations have worse health than those who work in professional occupations and that the health effect of occupation operates at least in part independently of the personal characteristics of the workers. Cutler *et al.* (2008) present similar results using mortality as an indicator of health. Schrijvers *et al.* (1998) use Dutch cross-sectional data to study the impact of working conditions on the association between occupational class and self-reported health. Hazardous physical working conditions are more prevalent in lower occupational classes, and this explains a substantial part (for males up to 83 percent) of the association between health and occupational (social) class.

Extensive research further suggests an important role of lifestyle factors, particularly smoking, in explaining SES disparities in health (Khang *et al.*, 2009; Mackenbach *et al.*, 2004). Using three different data sets from the UK and the US, Marmot *et al.* (1997a) find that features of the psycho-social working environment, social circumstances outside work, and health behavior account for much of the social gradient in health (see also House *et al.*, 1994). Fuchs (1986) argues that in developed countries, it is personal lifestyles that cause the greatest variation in health. Some epidemiological studies estimate that around two thirds of the social gradient in health deterioration could be explained by working environment and life style factors alone (Borg and Kristensen, 2000).

The evidence presented above highlights the potential role of both the physical and psychosocial aspects of the working environment in explaining the SES gradient in health. We introduce the choice of working environment and other health-related behavior in our theoretical framework by allowing an individual's health decline to be partly endogenous, involving individual decisions concerning life style factors and working environment.

First, individuals choose their level of undesirable job characteristics which potentially have health consequences, denoted as "job-related health stress". The concept of job-related health stress can be interpreted broadly and can range from physical working conditions (e.g., hard or risky labor) to psychosocial aspects of work (e.g., low social status, lack of control, repetitive work, etc). The decision to engage in unhealthy labor is governed by the relative benefit of a possible wage premium — a compensating wage differential (Smith, 1776; Thaler and Rosen, 1975; Viscusi, 1978; 1979) — versus the cost in terms of a higher health deterioration rate. Evidence is strong that there is a wage premium for jobs with higher mortality risk (Smith, 1978), and also for less serious, non-fatal, health risks (e.g. Viscusi, 1978; Olson, 1981; Duncan and Holmlund, 1983).

Second, individuals engage in preventive care (such as check up doctor visits), healthy consumption (such as the consumption of healthy foods, sports and exercise, but also living in a "healthy" neighborhood with low levels of pollution and crime and with a high quality service infrastructure), and unhealthy consumption (such as smoking, excessive alcohol consumption). Preventive care and healthy consumption are associated with health benefits in that they lower the health deterioration rate, either instantaneously or after a substantial period of exposure. Healthy consumption also provides direct utility whereas preventive care is assumed to solely provide health benefits (similar to curative care, individuals demand preventive care solely for the health benefits it provides). The distinction between healthy consumption and preventive care could in practice be difficult for some activities and could differ across individuals (e.g., some individuals exercise because they derive utility from it, whereas others solely exercise because it is healthy). Unhealthy consumption provides consumption benefits (utility) but increases the health deterioration rate.

A multitude of potential pathways between health and SES and vice versa: As Cutler *et al.* (2008) note, the mechanisms linking the various dimensions of SES to health are diverse. Some cause health, some are caused by health and some are jointly determined with health.

- *Education on health:* Education is found to have a causal effect on health and mortality (Lleras-Muney, 2005; Oreopoulos, 2006; Smith, 2007; Silles, 2009). However, Cutler *et al.* (2008) note that the mechanisms by which education affects health are not well understood. While consumption behavior and curative and preventive care can partly explain the effect of education on health, it remains largely unclear why more educated individuals behave in a healthier manner (Cutler *et al.*, 2008). Education increases earnings (e.g., Mincer, 1974), and in turn wealth, and thereby enables purchases of health investment (though this may also increase the opportunity cost of time). Education potentially increases the efficiency of curative and preventive care usage and time inputs into the production of health investment (Grossman 1972a; 1972b). It appears that the higher educated are better able at managing their diseases (Goldman and Smith, 2002), and high SES individuals appear to benefit more from new knowledge and new technology (Lleras-Muney and Lichtenberg, 2002; Glied and Lleras-Muney, 2003).

- *Health on education:* In our formulation we take as our starting point the age by which the majority of individuals have completed their education and joined the

labor force (e.g., around age 25 or so), and treat the level of education obtained by young adults as well as their initial level of health as initial conditions. In other words, we do not include the potential influence of childhood health on education and treat education as being predetermined by the time individuals join the labor-force[2].

- *Income or wealth on health:* Smith (2007) finds no effect of financial measures of SES (income, wealth and change in wealth) on changes in health in the US. Financial indicators of SES do not seem to cause the onset of health problems at any age (Smith, 2007). Cutler *et al.* (2008) provide an overview of empirical findings and conclude that the evidence points to no or a very limited impact of income or wealth on health. Yet, this view is not unequivocally accepted. Replication is still needed and controversy remains on the extent to which these findings apply uniformly to different population segments. For example, Lynch *et al.* (1997) suggest that accumulated exposure to economic hardship causes bad health, and Herd *et al.* (2008) argue that there might be causal effects of financial resources on health at the bottom of the income or wealth distribution.

  Regardless of this debate we include a number of plausible pathways by which income and wealth could causally impact health. For example, income and wealth enable purchases of curative and preventive care and thereby potentially allow for better health maintenance. The impact of financial resources on health is likely to depend on the manner of health care provision in a country. In the case of market provision, income, wealth and employment may determine access to health care, whereas in the case of universal health care provision these factors may be less important. On the other hand, higher wages are associated with greater opportunity costs, which would reduce the amount of time devoted to health maintenance. Further, more affluent workers may choose safer working (associated with lower level of job-related health stress) and living environments since safety is a normal good (Viscusi 1978, 1993).

- *Health on income and wealth:* Healthy individuals are more productive, earn higher wages and are able to accrue greater wealth (Currie and Madrian, 1999; Contoyannis and Rice, 2001).

Labor-force participation:

Studies have shown that perhaps the most dominant causal relation between health and

[2]Our model is therefore limited to explaining the formation of disparities in health from early adulthood till old age but not during childhood or the fetal period.

dimensions of SES is the causal impact that health has on one's ability to work and hence produce income and wealth (e.g., Smith, 2005a; 2007; Case and Deaton, 2005; Van Doorslaer *et al.*, 2008). Our formulation allows for endogenous retirement but in this paper we treat the decision as exogenous to facilitate derivations.

Mortality:
An essential component of the disparity in health by SES is the observed difference in mortality between SES groups. Further, length of life might be an important determinant of the age of retirement and the level of consumption and health investment over the life-course. Ehrlich and Chuma (1990) argue that the demand for health must be derived in conjunction with that for longevity and consumption. Our formulation allows for endogenous length of life but in this paper we treat the decision as exogenous to facilitate derivations.

6.2.3 Theoretical formulation

In this section we formalize the above discussion on the features of a theoretical framework for the SES health gradient over the life cycle. The aim is to understand the SES-health gradient as the outcome of rational constrained individual behavior.

We present the Grossman model for the demand for health (Grossman 1972a, 1972b, 2000) in continuous time (see also Wagstaff, 1986; Wolfe, 1985; Ehrlich and Chuma, 1990; Zweifel and Breyer, 1997) with seven essential additional features. First, we assume decreasing-returns-to-scale (DRTS) to investments in curative care. This solves the indeterminacy problem ("bang-bang" solution) for investment in curative care (Ehrlich and Chuma, 1990), ensures that investment in curative (medical) care is non-negative (for the usual assumptions of functional forms), reproduces the observed negative relation between health and the demand for medical care, and explains differences in the level of health as well as the rate of health deterioration between low and high SES groups (for a detailed discussion of the properties of a Grossman model with a DRTS health production function, see Galama, 2010). With these essential issues addressed, the model provides a foundation for a framework of the SES health gradient.

Second, we introduce the notion that individuals may accept risky and/or unhealthy work environments, in exchange for higher pay (Muurinen, 1982; Case and Deaton, 2005). In other words, we will explore solutions in which the decision to rapidly "wear one's body down" (i.e., to perform "hard" labor or engage in work with psychosocial health risks) is endogenous. Third, we include the possibility that consumption affects the "natural"

health deterioration rate, and fourth that individuals engage in preventive care to slow the natural rate of deterioration. This will allow us to include consumption behavior, which we interpret broadly to include decisions regarding housing and neighborhood, and our new concept of endogenous preventive care[3]. Fifth, we include the decision to withdraw from the labor force (Galama *et al.*, 2009). Sixth, we include endogenous length of life (Ehrlich and Chuma, 1990)[4]. Lastly, the causal effect of education on income is included in a straightforward manner by assuming a Mincer-type wage relation, in which earnings are increasing in the level of education and the level of experience of workers (e.g., Mincer, 1974).

With the exception of the above seven additional features the discussion below follows the usual formulation of the Grossman model (e.g., Grossman, 1972a; 1972b; 2000; Wagstaff, 1986; Zweifel and Breyer, 1997). Health is treated as a form of human capital (health capital) and individuals derive both consumption (health provides utility) and production benefits (health increases earnings) from it. The demand for health investment (broadly interpreted as curative and/or preventive care) is a derived demand: individuals demand "good health", not the consumption of curative or preventive care. Consumption and health investment constitute both own-time inputs and goods or services purchased in the market.

Individuals maximize the life-time utility function

$$\int_0^T U(t)e^{-\beta t}dt, \tag{6.1}$$

where T denotes total (endogenous) life time, β is a subjective discount factor and individuals derive utility $U(t) \equiv U[C_h(t), C_u(t), H(t)]$ from healthy consumption $C_h(t)$, unhealthy consumption $C_u(t)$ and from health $H(t)$. Time t is measured from the time the majority of individuals have completed their education and joined the labor force (e.g., around age 25 or so). Utility increases with healthy consumption $\partial U(t)/\partial C_h(t) \geq 0$, unhealthy consumption $\partial U(t)/\partial C_u(t) \geq 0$ and with health $\partial U(t)/\partial H(t) \geq 0$. Further, we assume diminishing marginal benefits: $\partial^2 U(t)/\partial^2 C_h(t) \leq 0$, $\partial^2 U(t)/\partial^2 C_u(t) \leq 0$ and $\partial^2 U(t)/\partial^2 H(t) \leq 0$.

The objective function (6.1) is maximized subject to the following dynamic equations,

$$\dot{H}(t) = I_m(t)^{\alpha} - d(t)H(t), \tag{6.2}$$

$$\dot{A}(t) = \delta A(t) + Y(t) - p_{X_h}(t)X_h(t) - p_{X_u}(t)X_u(t) - p_m(t)m_m(t) - p_p(t)m_p(t), \tag{6.3}$$

[3]It is useful to interpret the endogenous variables as bundles of goods and services (e.g., various consumption goods/services) or composite environmental factors (e.g., various physical and psychosocial health stresses).

[4]However, to facilitate derivations we treat mortality and retirement as exogenous in this work.

the total time budget Ω,

$$\Omega = \tau_w(t) + \tau_{I_m}(t) + \tau_{I_p}(t) + \tau_{C_h}(t) + \tau_{C_u}(t) + s[H(t)], \tag{6.4}$$

and we have initial and end conditions: $H(0)$, $H(T)$, $A(0)$ and $A(T)$ are given[5].

$\dot{H}(t)$ and $\dot{A}(t)$ in equations (6.2) and (6.3) denote time derivatives of health $H(t)$ and assets $A(t)$. Health (equation 6.2) can be improved through investment in curative (medical) care $I_m(t)$ and deteriorates at the "natural" health deterioration rate $d(t) \equiv d[t, C_h(t), C_u(t), z(t), I_p(t); \xi(t)]$. The health production function $I_m(t)^\alpha$ is assumed to exhibit decreasing-returns-to-scale ($0 < \alpha < 1$; see Ehrlich and Chuma, 1990). The "natural" deterioration rate depends endogenously on healthy consumption $C_h(t)$, unhealthy consumption $C_u(t)$, job-related health stress $z(t)$, and investment in preventive care $I_p(t)$ and on a vector of exogenous variables $\xi(t)$. Consumption can be healthy ($\partial d(t)/\partial C_h(t) \leq 0$; e.g., healthy foods, healthy neighborhood) or unhealthy ($\partial d(t)/\partial C_u(t) > 0$; e.g., smoking). Preventive care is modelled analogous to curative care as an activity that provides no utility ($\partial U(t)/\partial I_p(t) = 0$) but is demanded for its health benefits ($\partial d(t)/\partial I_p(t) < 0$). Greater job-related health stress $z(t)$ accelerates the "aging" process ($\partial d(t)/\partial z(t) > 0$).

Assets $A(t)$ (equation 6.3) provide a return δ (the interest rate), increase with income $Y(t)$ and decrease with purchases in the market of healthy consumption goods $X_h(t)$, unhealthy consumption goods $X_u(t)$, curative care $m_m(t)$ and preventive care $m_p(t)$ at prices $p_{X_h}(t)$, $p_{X_u}(t)$, $p_m(t)$ and $p_p(t)$, respectively. Income $Y(t) \equiv Y[H(t), z(t); E, x(t)]$ is assumed to be an increasing function of health $H(t)$ ($\partial Y(t)/\partial H(t) > 0$) and an increasing function in job-related health stress $z(t)$ ($\partial Y(t)/\partial z(t) > 0$; Case and Deaton, 2005). Further, income depends exogenously on the consumer's stock of knowledge (an individual's human capital exclusive of health capital), usually assumed to be a function of years of schooling E and years of working experience $x(t)$ (e.g., Mincer, 1974).

Goods and services $m_m(t)$ and $m_p(t)$ as well as own time inputs $\tau_{I_m}(t)$ and $\tau_{I_p}(t)$ are used in the production of curative care $I_m(t)$ and preventive care $I_p(t)$, respectively. Similarly, goods $X_h(t)$ and $X_u(t)$ purchased in the market and own time inputs $\tau_{C_h}(t)$ and $\tau_{C_u}(t)$ are used in the production of healthy and unhealthy consumption, $C_h(t)$ and $C_u(t)$, respectively. The efficiencies of production are assumed to be a function of the consumer's stock of knowledge E as it is generally believed that the more educated are

[5] In Grossman's original formulation (Grossman 1972a, 1972b) length of life T is determined by a minimum health level $H_{\min}$. If health falls below this level $H(t) \leq H_{min}$ an individual dies, hence $H(T) \equiv H_{\min}$.

more efficient at investing in health (see, e.g., Grossman, 2000),

$$I_m(t) \equiv I_m[m_m(t), \tau_{I_m}(t); E], \tag{6.5}$$
$$I_p(t) \equiv I_p[m_p(t), \tau_{I_p}(t); E], \tag{6.6}$$
$$C_h(t) \equiv C_h[X_h(t), \tau_{C_h}(t); E], \tag{6.7}$$
$$C_u(t) \equiv C_u[X_u(t), \tau_{C_u}(t); E]. \tag{6.8}$$

Further, we implicitly assume that curative care $I_m(t)$, preventive care $I_p(t)$ and job-related health stress $z(t)$ are non-negative. We do so by assuming DRTS of the health production function in investment in curative care (see equation 6.2) and diminishing marginal benefits for job-related health stress and for investment in preventive care. The notion here is that one cannot "sell" one's health through negative curative care (see Galama and Kapteyn, 2009) or negative preventive care nor can one "buy" health through negative job-related health stress[6].

The total time available in any period Ω is the sum of all possible uses $\tau_w(t)$ (work), $\tau_{I_m}(t)$ (curative care), $\tau_{I_p}(t)$ (preventive care), $\tau_{C_h}(t)$ (healthy consumption), $\tau_{C_u}(t)$ (unhealthy consumption) and $s[H(t)]$ (sick time). The resulting time budget constraint is shown in equation (6.4).

We follow Grossman (1972a, 1972b, 2000) and assume that income $Y(t)$ is a function of the wage rate $w(t)$ times the amount of time spent working $\tau_w(t)$,

$$Y(t) = w(t)\left\{\Omega - \tau_{I_m}(t) - \tau_{I_p}(t) - \tau_{C_h}(t) - \tau_{C_u}(t) - s[H(t)]\right\}. \tag{6.9}$$

The wage rate $w(t) \equiv w[t, z(t); E, x(t)]$ is a function of job-related health stress $z(t)$

$$w(t) = w_*(t)[1 + z(t)]^{\gamma_w}, \tag{6.10}$$

where $\gamma_w \geq 0$ and $w_*(t) \equiv w_*[E, x(t)]$ represents the "effortless" wage rate, i.e., the wage rate associated with the least job-related health stress $z(t) = 0$[7]. The effortless wage rate

[6] In earlier versions of the model we explicitly assumed non negativity by introducing multipliers $q_{I_m}(t)$, $q_{I_p}(t)$ and $q_z(t)$ associated with the condition that, respectively, curative care is non negative, $I_m(t) \geq 0$, preventive care is non negative, $I_p(t) \geq 0$, and job-related health stress is non negative, $z(t) \geq 0$. This is not necessary for DRTS technologies, for diminishing marginal benefits and choice of suitable functional forms which ensure that optimal solutions for curative care $I_m(t)$, preventive care $I_p(t)$ and job-related health stress $z(t)$ are non negative.

[7] Our model concerns individuals who participate in the labor force. Given that our frame of reference is the labor force we associate $z(t) = 0$ with the least amount of job-related health stress possible in employment, and since there is no obvious scale to job-related health stress we employ the simple relationship shown in equation (6.10).

$w_*(t)$ is a function of the consumer's education E and experience $x(t)$ (e.g., Mincer, 1974),

$$w_*(t) = w_E e^{\rho_E E + \beta_x x(t) - \beta_{x^2} x(t)^2}, \tag{6.11}$$

where education E is expressed in years of schooling, $x(t)$ is years of working experience, and ρ_E, β_x and β_{x^2} are constants, assumed to be positive.

Thus, we have the following optimal control problem: the objective function (6.1) is maximized with respect to the control functions $X_h(t)$, $\tau_{C_h}(t)$, $X_u(t)$, $\tau_{C_u}(t)$, $m_m(t)$, $\tau_{I_m}(t)$, $m_p(t)$, $\tau_{I_p}(t)$ and $z(t)$ and subject to the constraints (6.2, 6.3 and 6.4). The Hamiltonian (see, e.g., Seierstad and Sydsaeter, 1977; 1987) of this problem is:

$$\Im = U(t)e^{-\beta t} + q_H(t)\dot{H}(t) + q_A(t)\dot{A}(t), \tag{6.12}$$

where $q_H(t)$ is the adjoint variable associated with the differential equation (6.2) for health $H(t)$ and $q_A(t)$ is the adjoint variable associated with the differential equation (6.3) for assets $A(t)$.

The conditions for the optimal retirement age R and the optimal length of life T are

$$\Im(R) = 0, \tag{6.13}$$

$$\Im(T) = 0. \tag{6.14}$$

We have thus arrived at a life cycle model that incorporates labor force participation, healthy and unhealthy consumption (including housing, neighborhood, and social environment), health, curative (medical) and preventive care, job-related physical and psychosocial health stresses, wealth and mortality (all as endogenous variables).

6.3 Solutions

In this section we solve the optimal control problem conditional on retirement age R and length of life T (i.e., for fixed exogenous R and T). We first present the first-order conditions for optimization (section 6.3.1). Next we discuss the characteristics of the solutions (section 6.3.2) and the predictions for the relations between health and SES (section 6.3.3).

6.3.1 First-order conditions

The first-order condition for maximization of (6.1) with respect to health is

$$\frac{\partial U(t)}{\partial H(t)} = q_A(0)\left[\sigma_H(t) - \varphi_H(t)\right] e^{(\beta-\delta)t}, \tag{6.15}$$

where the Lagrange multiplier $q_A(0)$ is the shadow price of life-time wealth (see, e.g, Case and Deaton, 2005), $\sigma_H(t) \equiv \sigma_H[t, I_m(t), C_h(t), C_u(t), z(t), I_p(t); E, x(t), \xi(t)]$ is the user cost of health capital at the margin

$$\sigma_H(t) \equiv \pi_{I_m}(t)\left[d(t) + \delta - \widetilde{\pi_{I_m}}(t)\right], \tag{6.16}$$

$\pi_{I_m}(t) \equiv \pi_{I_m}[t, I_m(t), z(t); E, x(t)]$ is the marginal cost of curative care $I_m(t)$

$$\pi_{I_m}(t) \equiv \frac{p_m(t) I_m(t)^{1-\alpha}}{\alpha[\partial I_m(t)/\partial m_m(t)]} = \frac{w(t) I_m(t)^{1-\alpha}}{\alpha[\partial I_m(t)/\partial \tau_{I_m}(t)]}, \tag{6.17}$$

and $\varphi_H(t) \equiv \varphi_H[t, H(t), z(t); E, x(t)]$ is the marginal production benefit of health

$$\varphi_H(t) \equiv \frac{\partial Y(t)}{\partial H(t)}. \tag{6.18}$$

The symbol $\sim$ is used to denote the relative time derivative of a function: $\widetilde{f(t)} \equiv \frac{\partial f(t)}{\partial t} f(t)^{-1}$. Note that the marginal cost of investment in curative care $\pi_{I_m}(t)$ (equation 6.17) increases with the level of investment in curative care $I_m(t)$ (contrast this with equation 10 in Grossman, 2000) due to decreasing-returns-to-scale of the health production function $I_m(t)^\alpha$ ($0 < \alpha < 1$; see equation 6.2).

The first-order condition for maximization of (6.1) with respect to healthy consumption is

$$\frac{\partial U(t)}{\partial C_h(t)} = q_A(0)\left[\pi_{C_h}(t) - \varphi_{dC_h}(t)\right] e^{(\beta-\delta)t}, \tag{6.19}$$

where $\pi_{C_h}(t) \equiv \pi_{C_h}[t, C_h(t), z(t); E, x(t)]$ is the marginal cost of healthy consumption $C_h(t)$

$$\pi_{C_h}(t) \equiv \frac{p_{X_h}(t)}{\partial C_h(t)/\partial X_h(t)} = \frac{w(t)}{\partial C_h(t)/\partial \tau_{C_h}(t)}, \tag{6.20}$$

and $\varphi_{dC_h}(t) \equiv \varphi_{dC_h}[t, H(t), I_m(t), C_h(t), C_u(t), z(t), I_p(t); E, x(t), \xi(t)]$ is the additional health benefit associated with healthy consumption

$$\varphi_{dC_h}(t) \equiv -\pi_{I_m}(t)\frac{\partial d(t)}{\partial C_h(t)} H(t). \tag{6.21}$$

Similarly, the first-order condition for maximization of (6.1) with respect to unhealthy consumption is

$$\frac{\partial U(t)}{\partial C_u(t)} = q_A(0)\left[\pi_{C_u}(t) + \pi_{dC_u}(t)\right] e^{(\beta-\delta)t}, \tag{6.22}$$

where $\pi_{C_u}(t) \equiv \pi_{C_u}[t, C_u(t), z(t); E, x(t)]$ is the marginal cost of unhealthy consumption $C_u(t)$

$$\pi_{C_u}(t) \equiv \frac{p_{X_u}(t)}{\partial C_u(t)/\partial X_u(t)} = \frac{w(t)}{\partial C_u(t)/\partial \tau_{C_u}(t)}, \tag{6.23}$$

and $\pi_{dC_u}(t) \equiv \pi_{dC_u}[t, H(t), I_m(t), C_h(t), C_u(t), z(t), I_p(t); E, x(t), \xi(t)]$ is the additional health cost associated with unhealthy consumption

$$\pi_{dC_u}(t) \equiv \pi_{I_m}(t)\frac{\partial d(t)}{\partial C_u(t)}H(t). \tag{6.24}$$

The first-order condition for maximization of (6.1) with respect to job-related health stress is

$$\pi_{dz}(t) \quad = \quad \varphi_z(t), \tag{6.25}$$

where $\pi_{dz}(t) \equiv \pi_{dz}[t, H(t), I_m(t), C_h(t), C_u(t), z(t), I_p(t); E, x(t), \xi(t)]$ is the marginal cost of job-related health stress

$$\pi_{dz}(t) \equiv \pi_{I_m}(t)\frac{\partial d(t)}{\partial z(t)}H(t), \tag{6.26}$$

and $\varphi_z(t) \equiv \varphi_z[t, H(t), z(t); E, x(t)]$ is the marginal production benefit of job-related health stress

$$\varphi_z(t) \equiv \frac{\partial Y(t)}{\partial z(t)}. \tag{6.27}$$

Lastly, the first-order condition for maximization of (6.1) with respect to preventive care is

$$\pi_{I_p}(t) \quad = \quad \varphi_{dp}(t), \tag{6.28}$$

where $\pi_{I_p}(t) \equiv \pi_{I_p}[t, z(t), I_p(t); E, x(t)]$ is the marginal cost of preventive care $I_p(t)$

$$\pi_{I_p}(t) \equiv \frac{p_p(t)}{\partial I_p(t)/\partial m_p(t)} = \frac{w(t)}{\partial I_p(t)/\partial \tau_{I_p}(t)}, \tag{6.29}$$

and $\varphi_{dp}(t) \equiv \varphi_{dp}[t, H(t), I_m(t), C_h(t), C_u(t), z(t), I_p(t); E, x(t), \xi(t)]$ is the marginal benefit of preventive care

$$\varphi_{dp}(t) \equiv -\pi_{I_m}(t)\frac{\partial d(t)}{\partial I_p(t)}H(t). \tag{6.30}$$

The five first-order equations (6.15, 6.19, 6.22, 6.25 and 6.28) define the dynamics of the problem we are interested in. The "traditional" Grossman model is a special case of our model and is defined by the first-order equations (6.15) and (6.19) for an exogenous "natural" deterioration rate $d(t)$ (consumption does not improve or worsen the health deterioration rate). The first-order conditions (6.19), (6.22) and (6.25) are similar (but not identical) to those presented by Case and Deaton (2005).

6.3.2 Characteristics of the solutions

In this section we provide a qualitative and intuitive description of the nature of the solutions for health $H(t)$, investment in curative care $I_m(t)$, investment in preventive care $I_p(t)$, healthy consumption $C_h(t)$ and unhealthy consumption $C_u(t)$ and job-related health stress $z(t)$.

In the remainder, we assume[8]:

- diminishing marginal utilities of healthy $C_h(t)$ and unhealthy consumption $C_u(t)$ and of health $H(t)$,
- diminishing marginal benefits of health $\varphi_H(t)$, of job-related health stress $\varphi_z(t)$ and of investment in preventive care $\varphi_{dp}(t)$,
- diminishing returns to scale (DRTS) in the health production function $I_m(t)^\alpha$, and
- constant (CRTS) or diminishing (DRTS) returns to scale in the inputs (goods/services purchased in the market and own-time) for investment in curative care $I_m(t)$, preventive care $I_p(t)$, healthy consumption $C_h(t)$ and unhealthy consumption $C_u(t)$.

The assumption of DRTS of the health production function (equation 6.2) in investment in curative care $I_m(t)$ is critical and a distinguishing feature of this work with respect to prior formulations based on the Grossman model (with the exception of the work by Ehrlich and Chuma, 1990; see also Dustmann and Windmeijer, 2000, and Liljas, 2000). As Ehrlich and Chuma (1990) have shown, this assumption firstly addresses the indeterminacy problem ("bang-bang" solution) for investment in curative care. Second, as discussed in Galama (2010), it reproduces the observed negative relation between health and investment in curative care (see the critique by Zweifel and Breyer, 1997), and finally it allows for differences in the health deterioration rate (not just the level) between SES groups (see the critique by Case and Deaton, 2005). This latter point is crucial: unlike alternative life-cycle models of health, medical care, and SES, our formulation can explain the formation of disparities in health by SES with age.

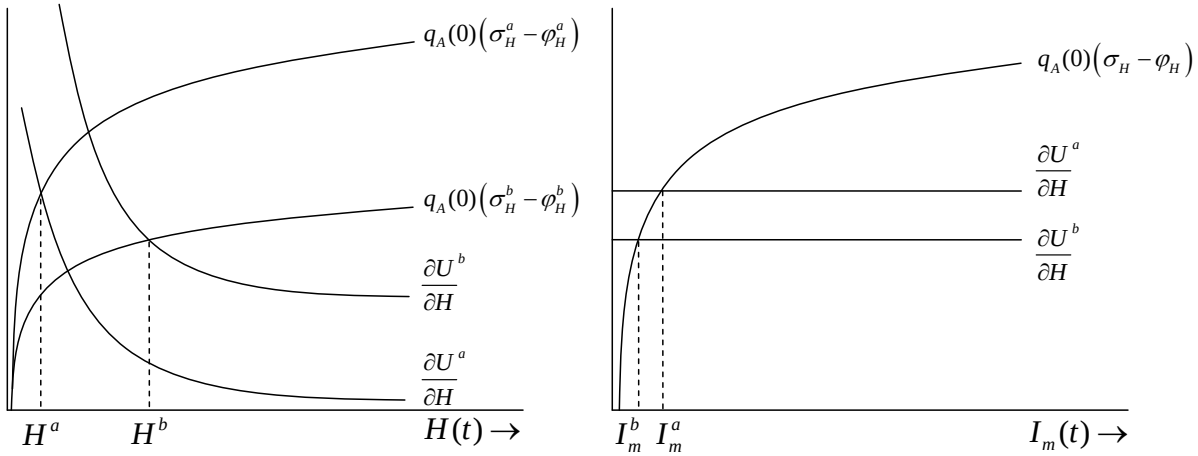
[8]Optimal solutions for the state functions $A(t)$, $H(t)$ and the control functions $X_h(t)$, $\tau_{C_h}(t)$, $X_u(t)$, $\tau_{C_u}(t)$, $m_m(t)$, $\tau_{I_m}(t)$, $m_p(t)$, $\tau_{I_p}(t)$ and $z(t)$ exist if the Hamiltonian $\Im$ (see equations 6.2, 6.3 and 6.12) is concave in each of the state and control functions and differentiable w.r.t. the state and control functions (see, e.g., Seierstad and Sydsaeter, 1977; 1987). These conditions are met for the usual assumptions of differentiability of functions and for the usual assumptions of constant or diminishing returns to scale.

To aid interpretation of the first-order conditions we provide specific functional forms in the Appendix for the utility function $U(t)$, and the relation between income $Y(t)$ on the one hand and health, years of schooling and years of working experience on the other. Further, we make the dependence of the "natural" deterioration rate $d(t)$ on job-related health stress, healthy and unhealthy consumption, and preventive care explicit. Finally, we specify the relations between the outputs consumption $C_h(t)$, $C_u(t)$, and health investments $I_m(t)$ and $I_p(t)$ and the inputs of own-time and goods/services purchased in the market.

The solutions for health and curative care

The solutions for health and for investment in curative care are both determined by the first-order condition (6.15) which equates the consumption benefit of health $\partial U(t)/\partial H(t)$ with the cost of maintaining the health stock $q_A(0)[\sigma_H(t) - \varphi_H(t)]e^{(\beta-\delta)t}$. Because of the diminishing marginal utility of health, the optimal level of the health stock is high for wealthy individuals[9], low user cost of health capital $\sigma_H(t)$ and a high production benefit of health $\varphi_H(t)$.

Figure 6.1: Marginal benefit versus marginal cost of health



Notes: In labelling the curves we have omitted the term $e^{(\beta-\delta)t}$.

Consider the relation between health $H(t)$ and investment in curative care $I_m(t)$ for an individual. Figure 6.1 shows a simple stylized graph of the marginal benefit and marginal cost of health as a function of health $H(t)$ (left-hand side) and as a function of investment in curative care $I_m(t)$ (right-hand side). Consider the left-hand figure first. The marginal consumption benefit of health (labelled $\partial U^a/\partial H$) is downward sloping in

[9] $q_A(0)$ is decreasing in life-time wealth. See, e.g. Wagstaff (1986).

health to represent the diminishing marginal utility of greater health. The user cost of health capital $\sigma_H(t)$ (see equation 6.16) is assumed to be independent of the level of health $H(t)$ and the production benefit of health $\varphi_H(t)$ (see equation 6.18) is assumed to exhibit diminishing returns in health. The resulting curve (labelled $q_A(0)(\sigma_H^a - \varphi_H^a)$) is upward sloping in health. The level of health for which the two curves intersect determines the health stock H^a.

Now consider the right-hand side of Figure 6.1. The first-order condition for health (equation 6.15) also determines the level of investment in curative care I_m^a. The optimal level of investment in curative care $I_m(t)$ is higher for greater life-time wealth (lower $q_A(0)$), higher marginal consumption benefit of health $\partial U(t)/\partial H(t)$, and higher marginal production benefit of health $\varphi_H(t)$. The marginal cost of curative care $\pi_{I_m}(t)$ and hence the user cost of health capital $\sigma_H(t)$ is increasing in the level of curative care (see equations 6.16, 6.17 and 6.48)[10]. The marginal production benefit of health $\varphi_H(t)$ (see equations 6.9, 6.18 and 6.45) is independent of the level of investment in curative care $I_m(t)$. The resulting curve is upward sloping (labelled $q_A(0)(\sigma_H - \varphi_H)$). Further, the marginal utility of health $\partial U(t)/\partial H(t)$ is independent of the level of investment in curative care $I_m(t)$: this is shown as the horizontal line (labelled $\partial U^a/\partial H$). The intersection of the two curves determines the optimal level of investment in curative care I_m^a.

Now consider an optimal solution with a lower level of investment in curative care ($I_m^b < I_m^a$) while holding all exogenous variables and functions constant. Since the first-order condition for investment in curative care (equation 6.15) also determines the level of health, we expect an associated change in the level of the health stock. As discussed before, the marginal cost of curative care $\pi_{I_m}(t)$ is increasing in the level of curative investment $I_m(t)$. Thus a lower level of investment in curative care $I_m^b < I_m^a$ is, ceteris paribus, associated with a lower user cost of health capital at the margin $\sigma_H(t) = \pi_{I_m}(t)[d(t) + \delta - \widetilde{\pi_{I_m}}(t)]$. This is shown as a shift downward in the curve of the net marginal cost of health capital (labelled $q_A(0)[\sigma_H^b - \varphi_H^b]$)[11] and the stock of health is higher[12]. However, a

[10] As noted before, we assume DRTS in curative care of the health production function (equation 6.2; $0 < \alpha < 1$) and CRTS or DRTS in curative care for the production of investment in curative care (equation 6.47; $0 < \alpha_{I_m} + \beta_{I_m} \leq 1$).

[11] Note that the marginal benefit of health $\varphi_H(t)$ is not a function of the level of investment in curative care $I_m(t)$ (see equation 6.45). It is a function of the health stock $H(t)$ and that relation is unchanged and shown in Figure 6.1.

[12] Implicitly this amounts to assuming that the indirect effect of a change in the level of investment in preventive care $I_m(t)$ on health $H(t)$, through changes in the resulting optimal levels of investment in preventive care $I_p(t)$, healthy consumption $C_h(t)$, unhealthy consumption $C_u(t)$ and job-related health stress $z(t)$, is small and dominated by the direct effect of change in investment in preventive care $I_m(t)$ on health $H(t)$, through the first-order condition for health and investment in curative care (equation

higher stock of health potentially shifts the marginal utility of health $\partial U(t)/\partial H(t)$ as well (curve labelled $\partial U^b/\partial H$). It seems plausible that such a shift would be neutral (no shift) or in the upward direction as there are no a priori reasons to expect that improved health would diminish the marginal utility of consumption. The level of health H^b associated with lower investment in curative care ($I_m^b < I_m^a$) is higher ($H^b > H^a$). In other words, we find that healthy individuals invest less in curative care. This finding is supported by casual observation (the healthy do not go to the doctor) and by numerous empirical studies that find a strong negative correlation between measures of health and measures of curative (medical) care usage (see Galama and Kapteyn, 2009, for an overview of the empirical literature).

In contrast, for a CRTS health production function as employed in the Grossman literature (except Ehrlich and Chuma, 1990; Dustmann and Windmeijer, 2000) the marginal cost of investment in curative care $\pi_{I_m}(t)$ is independent of the level of investment (see equation 6.48 for $\alpha = 1$ and $\alpha_{I_m} + \beta_{I_m} = 1$). This leads to the problematic (i.e., unconfirmed) prediction that healthy individuals invest more (not less) in curative care (see, e.g., Zweifel and Breyer, 1997; Wagstaff 1986). It also produces an indeterminacy problem ("bang-bang" solution; e.g., Ehrlich and Chuma, 1990) as in this case both the marginal benefit (labelled $\partial U^a/\partial H$) and the marginal net cost (labelled $q_A(0)(\sigma_H - \varphi_H)$) are independent of the level of investment in curative care $I_m(t)$. Thus, both would be horizontal lines in the right-hand side of Figure 6.1 and without intersecting there would not be an optimal solution for investment in curative care. Only in the peculiar case where both lines exactly overlap a solution exists, allowing any (undetermined) level of investment in curative care. Further, in this case the marginal cost of curative care $\pi_{I_m}(t)$ is exogenously determined by the price of medical care $p_m(t)$, the wage rate $w(t)$ and the efficiency of curative care $\mu_{I_m}(t)$ (see equation 6.48). Hence there is often a mismatch between the actual and "desired" health stock. As a result one has to assume that individuals can instantaneously adjust their health to a "desired" level (Grossman, 2000). However, in a DRTS formulation the marginal cost of curative care $\pi_{I_m}(t)$ is endogenous: for any level

6.15). This would be true for small transitory endogenous variation in the "natural" deterioration rate $d(t)$ through choices made in working environment and in life style (operating through $C_h(t)$, $C_u(t)$, $z(t)$ and $I_p(t)$). In other words, if $\partial d(t)/\partial C_h(t)$, $\partial d(t)/\partial C_u(t)$, $\partial d(t)/\partial z(t)$ and $\partial d(t)/\partial I_p(t)$ are small. It also requires that changes in the optimal level of job-related health stress $z(t)$ affect the marginal cost of curative care $\pi_{I_m}(t)$ (e.g., equation 6.48) and the marginal benefit of health $\varphi_H(t)$ (e.g., equation 6.45) modestly (such that the downward shift is not undone). While transitory changes are assumed to be small, gradually, as time passes, lower levels of healthy consumption, curative and preventive care and higher levels of unhealthy consumption and job-related health stress lead to cumulative disadvantage over the life cycle.

of health $H(t)$ you can find a corresponding level of investment in curative care $I_m(t)$ that satisfies the first order condition for health (equation 6.15).

A more rigorous discussion of our finding that investment in curative care is negatively correlated with health can be found in Galama (2010) for the standard Grossman model with a DRTS health production function. The model presented here differs from that in Galama (2010) due to the introduction of life-style and working environment factors, which complicates a qualitative discussion.

Stylized representations of the first-order conditions

Figure 6.2 provides a stylized representation of the first-order conditions for healthy consumption $C_h(t)$ (equation 6.19), unhealthy consumption $C_u(t)$ (equation 6.22), job-related health stress $z(t)$ (equation 6.25) and investment in preventive care $I_p(t)$ (equation 6.28).

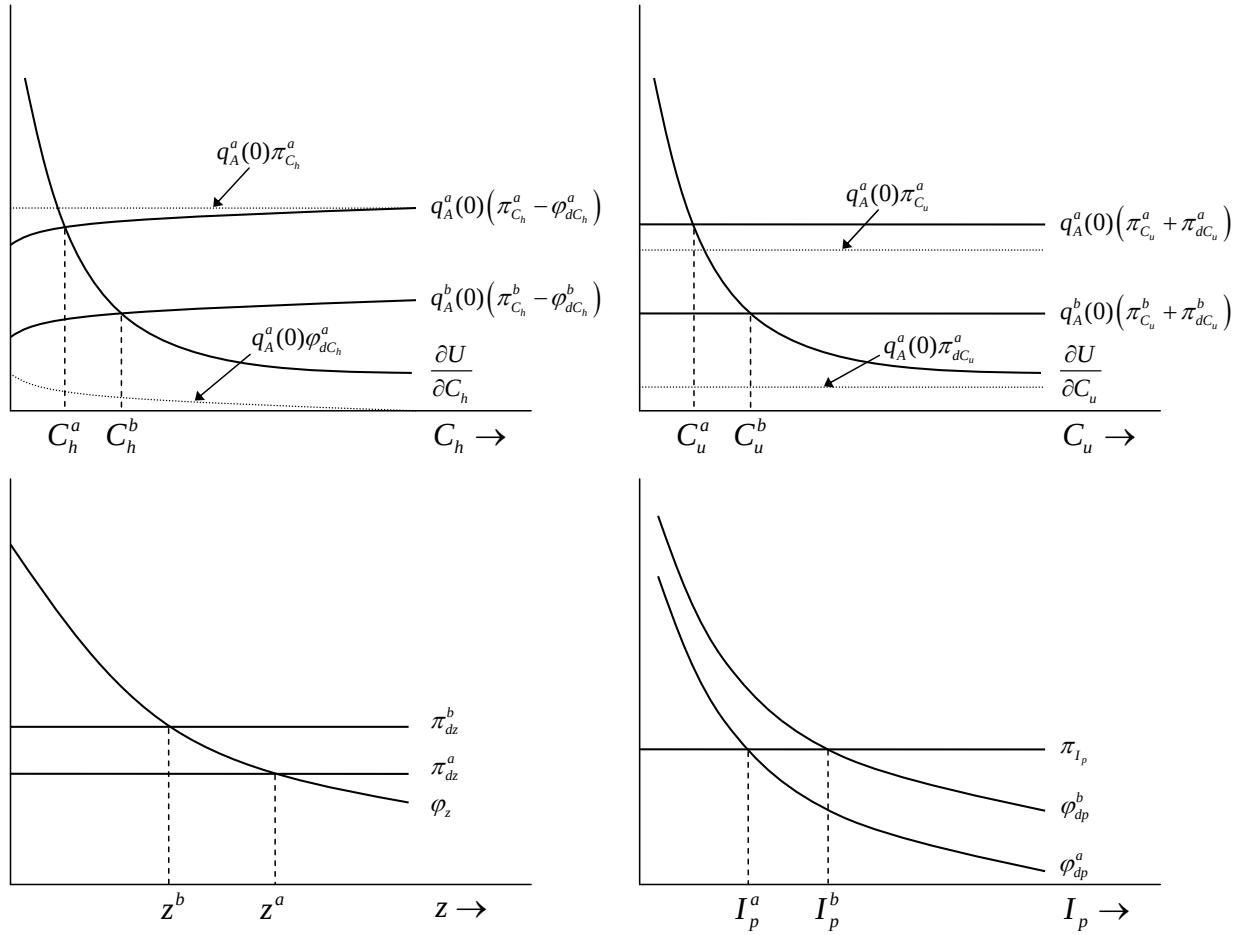
Healthy and unhealthy consumption

The top-left corner of Figure 6.2 shows the first-order condition for healthy consumption $C_h(t)$ (equation 6.19) which equates the marginal utility of healthy consumption (solid line labelled $\partial U/\partial C_h$) to the net marginal cost of healthy consumption (solid line labelled $q_A^a(0)(\pi_{C_h}^a - \varphi_{dC_h}^a)$)[13]. The marginal utility of healthy consumption is downward sloping to represent the diminishing marginal utility of greater consumption. The net marginal cost of healthy consumption increases with the marginal cost of healthy consumption $\pi_{C_h}(t)$ and decreases with the additional health benefit $\varphi_{dC_h}(t)$. A priori it is not clear whether the relationship between the inputs (good/services and own time) and healthy consumption $C_h(t)$ exhibits decreasing- or increasing-returns-to-scale (see equation 6.55 for an example functional form). For simplicity assume constant returns to scale so that the marginal cost of healthy consumption $\pi_{C_h}(t)$ is independent of the level of consumption (dotted horizontal line labelled $q_A^a(0)\pi_{C_h}^a$). Diminishing marginal additional health benefit $\varphi_{dC_h}(t)$ is represented by a downward sloping curve (dotted line labelled $q_A^a(0)\varphi_{dC_h}^a$). The net marginal cost of healthy consumption (solid line labelled $q_A^a(0)(\pi_{C_h}^a - \varphi_{dC_h}^a)$) is the difference between the two dotted curves. The point where the marginal utility of healthy consumption $\partial U/\partial C_h$ crosses the net marginal cost of healthy consumption $q_A^a(0)(\pi_{C_h}^a - \varphi_{dC_h}^a)$ defines the optimal solution for healthy consumption C_h^a (vertical dashed line).

The top-right corner of Figure 6.2 shows the first-order condition for unhealthy consumption $C_u(t)$ (equation 6.22). The first-order condition is similar to the condition for

[13] In labelling the curves we have omitted the term $e^{(\beta-\delta)t}$.

Figure 6.2: Marginal benefits versus marginal costs of life-style variables



Notes: healthy consumption $C_h(t)$ (top-left), unhealthy consumption $C_u(t)$ (top-right), job-related health stress $z(t)$ (bottom-left) and investment in preventive care $I_p(t)$ (bottom-right). In labelling the curves we have omitted the term $e^{(\beta-\delta)t}$.

healthy consumption described in the preceding paragraph. The difference lies in the additional cost of unhealthy consumption (dotted line labelled $q_A^a(0)\pi_{dC_u}^a$) which has to be added rather than subtracted from the marginal cost of unhealthy consumption (dotted line labelled $q_A^a(0)\pi_{C_u}^a$) to obtain the net marginal cost of unhealthy consumption (solid line labelled $q_A^a(0)(\pi_{C_u}^a + \pi_{dC_u}^a)$). The figure shows CRTS in the additional cost of unhealthy consumption $\pi_{dC_u}(t)$, since it is unclear a priori whether the effect of unhealthy consumption on the deterioration rate exhibits in- or decreasing returns to scale. The point where the marginal utility of unhealthy consumption $\partial U/\partial C_u$ crosses the net marginal cost of unhealthy consumption $q_A^a(0)(\pi_{C_u}^a + \pi_{dC_u}^a)$ defines the optimal solution for unhealthy consumption C_u^a (vertical dashed line).

Assuming a simple functional relationship between the inputs (good/services and own time) and healthy consumption $C_h(t)$ one finds that the marginal cost of healthy consumption $\pi_{C_h}(t)$ increases with the price of consumption goods $p_{X_h}(t)$ and the wage rate $w(t)$ (see equation 6.55 in the Appendix). Equation (6.56) presents a similar relationship for the marginal cost of unhealthy consumption $\pi_{C_u}(t)$. The marginal costs of consumption $\pi_{C_h}(t)$ and $\pi_{C_u}(t)$ thus represent the *direct* monetary costs of consumption in terms of the price of goods and the opportunity cost of time (the wage rate). The additional health benefit of healthy consumption $\varphi_{dC_h}(t)$ and the additional cost of unhealthy consumption $\pi_{dC_u}(t)$, on the other hand, represent the *indirect* savings (healthy consumption) and losses (unhealthy consumption) in investment in curative care: they are the product of the marginal cost of investment in curative care $\pi_{I_m}(t)$ and the "amount" of health saved $[\partial d(t)/\partial C_h(t)]H(t)$ (healthy consumption; see equation 6.21) or health lost $[\partial d(t)/\partial C_u(t)]H(t)$ (unhealthy consumption; see equation 6.24).

Further, as one can see from Figure 6.2, the optimal level of healthy consumption C_h^a is higher due to the additional benefit of healthy consumption in comparison to a situation in which this benefit were absent[14]. The opposite is true for the optimal level of unhealthy consumption C_u^a which is lower due to the additional cost of unhealthy consumption.

Job-related health stress

The bottom-left corner of Figure 6.2 shows the first-order condition for job-related health stress $z(t)$ (equation 6.25) which equates the marginal benefit to the marginal cost of job-related health stress. The marginal benefit of job-related health stress (equation 6.27) is downward sloping (solid line labelled φ_z) to represent diminishing marginal benefit. The marginal cost of job-related health stress (equation 6.26; solid line labelled π_{dz}^a) is shown

[14] The solid line labelled $q_A^a(0)(\pi_{C_h}^a - \varphi_{dC_h}^a)$ is shifted downwards compared to the dotted line labelled $q_A^a(0)\pi_{C_h}^a$.

with constant returns to scale, since it is unclear a priori whether it would exhibit in- or decreasing returns to scale[15]. The optimal solution for job-related health stress z^a is indicated by the vertical dashed line.

Investment in preventive care

The bottom-right corner of Figure 6.2 represents the first-order condition for investment in preventive care $I_p(t)$ (equation 6.28). Analogous to the first-order condition for job-related health stress the marginal benefit of investment in preventive care (equation 6.30) is downward sloping (solid line labelled $\varphi^a_{I_p}$) to represent diminishing marginal benefit and the marginal cost (equation 6.29; solid line labelled π_{I_p}) is shown with constant returns to scale, since it is unclear a priori whether it would exhibit in- or decreasing returns to scale[16]. The optimal solution for investment in preventive care I^a_p is indicated by the vertical dashed line.

6.3.3 Health and socioeconomic status

Individuals have partial control over their health through choices made in investment in curative care $I_m(t)$, *life style* (operating through healthy consumption $C_h(t)$, unhealthy consumption $C_u(t)$ and investment in preventive care $I_p(t)$) and in *working environment* (operating through job-related health stress $z(t)$). In this section we explore the (cumulative) effect on health over the life cycle through choices made in curative care, in life style and in working environment. Our emphasis will be on exploring differences in constraints (e.g., differences in wealth, skills, experience, education, prices and initial health) rather than differences in preferences (e.g., differences in time preferences and in the utility of health and consumption). Naturally differences in the level of patience and in preferences for health and consumption exist across individuals. Research on the possible role of differences in preferences in explaining the SES health gradient is important[17]. In the

[15]Note that the marginal cost of job-related health stress is a function of job-related health stress $z(t)$ through the marginal cost of investment in curative care $\pi_{I_m}(t)$ (see equation 6.47) which is a function of the wage rate $w(t)$ (see equation 6.10) and through the dependence of the deterioration rate on job-related health stress (e.g., equation 6.46).

[16]Note that whereas the marginal cost of job-related health stress (equation 6.26) is a function of the marginal cost of investment in curative care $\pi_{I_m}(t)$ and the "natural" deterioration rate $d(t)$, in the case of investment in preventive care it is the marginal benefit (not the cost; see equation 6.30) that shows a comparable dependence on $\pi_{I_m}(t)$ and $d(t)$.

[17]For example, Fuchs (1982, 1986) has argued that the strong correlation between education and health may be due to differences in the time preferences of individuals, which affects investments in both education and health and helps to explain variations in cigarette smoking, diet, and exercise. Other third

following discussion, however, we focus on explaining behavior as the result of rational decisions based on differences in the constraints faced by individuals.

In the following discussion we define a high SES individual as (i) wealthy (life-time wealth) and (ii) wise (highly educated, experienced, skilled). Common measures for SES employed in empirical research are wealth, income and education. In the following subsections we discuss the relations between wealth and health, income and health and education and health. We consider two individuals a and b who differ in one particular SES indicator, but are otherwise identical. Both individuals have the same initial level of health $H(t)$, are of the same age t, face the same environments (e.g., same interest rate δ), and have the same preferences (i.e., same utility function $U[C_h(t), C_u(t), H(t)]$ and same time preference β). We are interested in the predictions of our model for the subsequent evolution of health for these two individuals, given a ceteris paribus change in one SES indicator. We also briefly discuss the formation of the SES health gradient and the role of institutions.

Wealth and health: pure "asset" effect

Consider two individuals a and b who differ in life-time wealth $q_A(0)$. Individual b has greater life-time wealth, i.e., $q_A^a(0) > q_A^b(0)$ but is otherwise identical. Because of the similarities between the two individuals the difference in life-time wealth is to be interpreted as due to differences in endowed physical capital (e.g., assets $A(0)$).

Because both individuals have the same health stock $H(t)$ it follows that $q_A^a(0)[\sigma_H^a(t) - \varphi_H^a(t)] = q_A^b(0)[\sigma_H^b(t) - \varphi_H^b(t)]$ (equation 6.15) and hence $\sigma_H^a(t) - \varphi_H^a(t) < \sigma_H^b(t) - \varphi_H^b(t)$. Assume that both individuals have the same income function $Y[H(t)]$, i.e., approximately $\varphi_H^a(t) \sim \varphi_H^b(t)$ and hence $\sigma_H^a(t) < \sigma_H^b(t)$[18]. Assuming further that the secondary effect of changes in the control functions $C_h(t)$, $C_u(t)$, $z(t)$ and $I_p(t)$ on the "natural" deterioration rate is small, we find $\pi_{I_m}^a(t) < \pi_{I_m}^b(t)$ (see equation 6.16).

---

factors of interest that may produce a spurious correlation between education and health are general ability and intelligence (see e.g. Deary, 2008). Our model provides a framework for analyzing the role of such differences in individual preferences.

[18] This amounts to assuming that, as a first approximation, the optimal level of job-related health stress $z(t)$ is not much different as a result of the difference in life-time wealth (see equation 6.45 and note that both individuals have the same level of initial health $H(t)$ and are assumed to be identical in their level of education E and experience $x(t)$).

Investment in curative care:
As discussed earlier, $\pi^a_{I_m}(t) < \pi^b_{I_m}(t)$ implies $I^a_m(t) < I^b_m(t)$ (see equation 6.48)[19]. Thus our model predicts that wealthy individuals invest more in curative care $I_m(t)$. As a result, even without the inclusion of additional potential mechanisms responsible for the SES health gradient beside utilization of curative care, the model predicts differences in the rate of "effective" health deterioration between high and low SES individuals due to different levels of investment in curative care $I_m(t)$. In other words, a DRTS health production function as advocated here can address the criticism levelled by Case and Deaton (2005).

Healthy and unhealthy consumption:
The top-left figure of Figure 6.2 shows the shift in the level of healthy consumption (C^b_h; vertical dashed line). As a result of greater life-time wealth ($q^b_A(0) < q^a_A(0)$) the product of the shadow price of life-time wealth $q^b_A(0)$ and the marginal cost of healthy consumption $\pi^b_{C_h}$ shifts downward. The additional benefit of healthy consumption $\varphi_{dC_h}(t)$ (equation 6.21) is proportional to the product of the marginal cost of investment in curative care and the health stock $\pi_{I_m}(t)H(t)$. Since individual a and b are assumed to be equally healthy we find that wealthy individuals ($\pi^b_{I_m}(t) > \pi^a_{I_m}(t)$; see prior discussion) have greater additional benefit from healthy consumption $\varphi_{dC_h}(t)$. The net result of greater additional benefit from healthy consumption $q^b_A(0)\varphi^b_{dC_h}$ remains however relatively unchanged because of the competing effect of greater life-time wealth (lower $q^b_A(0) < q^a_A(0)$)[20]. On the one hand the marginal cost of investment in curative care $\pi_{I_m}(t)$ is greater for wealthier individuals. This increases the benefit of healthy consumption $\varphi_{dC_h}(t)$. On the other hand greater wealth enables purchases of curative care to compensate for health losses, approximately balancing the effect. The resulting net marginal cost of healthy consumption (solid line labelled $q^b_A(0)(\pi^b_{C_h} - \varphi^b_{dC_h})$) is shown as being lower in level (the "wealth shift" through $q_A(0)$) but with the same slope ($q^b_A(0)\varphi^b_{dC_h} \sim q^a_A(0)\varphi^a_{dC_h}$). The optimal solution for healthy consumption of a wealthier individual C^b_h (vertical dashed line) is higher than that of a poorer individual ($C^b_h > C^a_h$). For simplicity, the marginal utility of

[19] Again, we assume that, as a first approximation, the optimal level of job-related health stress $z(t)$ is not much different as a result of the difference in life-time wealth so that the wages of individual a and b are comparable (see, e.g., equation 6.10).

[20] Since $q^a_A(0)[\sigma^a_H(t) - \varphi^a_H(t)] = q^b_A(0)[\sigma^b_H(t) - \varphi^b_H(t)]$, and assuming as before that $\varphi^b_H(t) \sim \varphi^a_H(t)$ we have $q^b_A(0)\sigma^b_H(t) \sim q^a_A(0)\sigma^a_H(t)$ and hence $q^a_A(0)\pi^a_{I_m} \sim q^b_A(0)\pi^b_{I_m}$. As a result $q^b_A(0)\varphi^b_{dC_h}(t) = q^b_A(0)\pi^b_{I_m}(t)\partial d(t)/\partial C^b_h(t)H(t) \sim q^a_A(0)\pi^a_{I_m}(t)\partial d(t)/\partial C^b_h(t)H(t) \sim q^a_A(0)\varphi^a_{dC_h}(t)$.

healthy consumption $\partial U/\partial C_h$ is shown as unchanged[21]. Thus it is plausible that wealthy individuals consume more healthy consumption than the less wealthy ($C_h^b > C_h^a$).

The top-right figure of Figure 6.2 shows the shift in the level of unhealthy consumption (C_u^b; vertical dashed line). As with healthy consumption, greater life-time wealth ($q_A^b(0) < q_A^a(0)$) shifts the product of the shadow price of life-time wealth $q_A^b(0)$ and the marginal cost of unhealthy consumption $\pi_{C_u}^b$ downward. Analogous to healthy consumption, the net result of the additional cost of unhealthy consumption is small. On the one hand, higher demand for curative care increases the additional cost of unhealthy consumption ($\pi_{I_m}^b(t) > \pi_{I_m}^a(t)$ and hence $\pi_{dC_u}^b(t) > \pi_{dC_u}^a(t)$). On the other hand greater wealth enables purchases of curative care to compensate for health losses. The marginal utility of unhealthy consumption $\partial U/\partial C_u$ is shown as unchanged because a priori it is unclear whether the marginal utility of unhealthy consumption increases or decreases as a result of changes in the level of healthy consumption. The resulting optimal level of unhealthy consumption is higher $C_u^b > C_u^a$. These patterns suggest that wealthy individuals consume more unhealthy goods/services than less wealthy individuals ($C_u^b > C_u^a$) but because wealthy individuals have greater additional benefit from healthy consumption $\varphi_{dC_h}(t)$ and greater additional cost of unhealthy consumption $\pi_{dC_u}(t)$, wealthy individuals consume less unhealthy consumption as a share of their total consumption than less wealthy individuals ($C_u^b/(C_h^b + C_u^b) < C_u^a/(C_h^a + C_u^a)$).

However, it is plausible that healthy and unhealthy consumption are substitutes. Greater wealth ($q_A^b(0) < q_A^a(0)$) could then lead to a solution in which the level of healthy consumption increases ($C_h^b(t) > C_h^a(t)$) while the level of unhealthy consumption decreases ($C_u^b(t) < C_u^a(t)$). Smoking might be an example of unhealthy consumption that fits this pattern, if e.g. smoking reduces the marginal utility of exercise. High SES individuals smoke substantially less than low SES individuals.

Job-related health stress and investment in preventive care:
The first-order conditions for job-related health stress $z(t)$ and for investment in preventive care show no direct dependence on life-time wealth (no dependence on $q_A(0)$). This is because the wealth effect is the same for the marginal cost as it is for the marginal benefit for both job-related health stress and for investment in preventive care (in the first-order

[21] This assumption would be correct if the utility function were additive, consisting of separate functions for healthy and for unhealthy consumption, e.g. $U(t) = f_h[C_h(t), H(t)] + f_u[C_u(t), H(t)]$. However, for other forms of the utility function changes in the level of healthy consumption $C_h(t)$ and unhealthy consumption $C_u(t)$ potentially shift the marginal utility of healthy consumption upward or downward, depending on whether healthy and unhealthy consumption are substitutes or complements and depending on whether the optimal solution for unhealthy consumption is higher or lower for the wealthier individual.

conditions for job-related health stress and for investment in preventive care [see the Appendix] both the marginal cost and the marginal benefit are multiplied by $q_A(0)$). Intuitively, higher wealth decreases the marginal benefit of job-related health stress since there is no strong need to increase wages. Yet, at the same time greater wealth decreases the marginal cost of job-related health stress, since there are more financial resources available to invest in curative care to compensate for any detrimental impact on health. These effects cancel out. For the cost and benefit of preventive care a similar story can be told.

However, there is an indirect effect of greater life-time wealth. Both the marginal cost of job-related health stress $\pi_{dz}(t)$ (equation 6.26) and the marginal benefit of preventive care $\varphi_{dp}(t)$ (equation 6.30) are proportional to the product of the marginal cost of investment in curative care and health $\pi_{I_m}(t)H(t)$. Higher wealth (individual b) implies $\pi^a_{I_m}(t) < \pi^b_{I_m}(t)$ (see previous discussion). Thus wealthier individuals have greater marginal cost of job-related health stress $\pi_{dz}(t)$ and greater marginal benefit of preventive care $\varphi_{dp}(t)$. Consequently the optimal level of job-related health stress is lower ($z^b < z^a$; bottom-left corner of Figure 6.2) and the optimal level of investment in preventive care is higher ($I^b_p > I^a_p$; bottom-right corner of Figure 6.2) for wealthy individuals compared to less-affluent peers.

The effect of wealth on health:

In sum, wealthy individuals invest more in curative care $I_m(t)$. Associated with greater wealth is a higher marginal cost of curative care $\pi_{I_m}(t)$. As a result, wealthy individuals derive greater benefit from healthy consumption $\varphi_{dC_h}(t)$ and from preventive care $\varphi_{dp}(t)$ because of the savings these represent in terms of (costly) curative care. Their optimal levels of healthy consumption and investment in preventive care are higher. Similarly, wealthy individuals have greater cost of unhealthy consumption $\pi_{dC_u}(t)$ and of job-related health stress $\pi_{dz}(t)$ because these behavioral choices result in additional costly curative care. Wealthy individuals invest more in health through higher investment in curative and preventive care, engage in work that is more conducive to health (jobs associated with lower levels of job-related health stress) and shift from unhealthy toward healthy consumption. As a result their health deteriorates slower. Wealth protects health by encouraging healthy life styles and enabling individuals to work in healthy environments. It has to be noted that the effect of wealth on health disappears in the pure investment formulation of the model, i.e. if health does not enter the utility function.

Income and health: pure "wage" effect

Income $Y(t)$ is a function of the wage rate $w(t)$ times the amount of time spent working $\tau_w(t)$ (see equation 6.9). Again, consider two individuals a and b but this time the difference is in their level of income. Individual b earns a higher income than individual a ($Y^b(t) > Y^a(t)$) and hence earns a higher wage rate $w^b(t) > w^a(t)$[22]. It is important to distinguish between transitory and permanent wage responses.

Transitory wage change: In our model of perfect certainty a transitory change in wage does not affect the parameter $q_A(0)$ (e.g., life-time wealth) as the change is fully anticipated by the individual. Such a response is referred to as an evolutionary wage change (along an individual's wage profile). The effect of a transitory wage change is best understood as occurring within a person. Assume that for situation b the wage change has occurred while it has not (yet) occurred in situation a. Alternatively, one can assume that individual a's wage has not (yet) changed while individual b's wage rate has increased ($w^b(t) > w^a(t)$; but otherwise individual a and b follow similar wage trajectories).

A transitory increase in the wage rate $w(t)$ increases firstly the opportunity cost of time, but secondly also increases the production benefits of health (see equations 6.9, 6.18 and 6.45). As before, the marginal cost of investment in curative care $\pi_{I_m}(t)$ is determined by the first-order condition for health (equation 6.15). Because health $H(t)$ is as yet unchanged (it is a stock), the net marginal cost of maintaining the health stock is unchanged and $\sigma_H^a(t) - \varphi_H^a(t) = \sigma_H^b(t) - \varphi_H^b(t)$. Both the user cost of health capital $\sigma_H(t)$ (through the marginal cost of investment in curative care $\pi_{I_m}(t)$) and the marginal production benefit of health $\varphi_H(t)$ increase with the wage rate $w(t)$. It follows that $\varphi_H^b(t) > \varphi_H^a(t)$ and $\sigma_H^b(t) > \sigma_H^a(t)$. The net effect of a wage increase on the level of investment in curative care $I_m(t)$ is consequently unknown and investment in curative care could either increase or decrease.

Both the direct cost $\pi_{C_h}(t)$ (equation 6.20) and the additional marginal benefit $\varphi_{dC_h}(t)$ (equation 6.21) of healthy consumption increase with the wage rate $w(t)$ and the net result of a wage increase on the level of healthy consumption $C_h(t)$ is unknown. Similarly the marginal benefit $\varphi_z(t)$ (equation 6.26) and the marginal cost $\pi_{dz}(t)$ (equation 6.27) of job-related health stress $z(t)$, and the marginal benefit $\varphi_{dp}(t)$ (equation 6.30) and marginal cost $\pi_{I_p}(t)$ (equation 6.29) of investment in preventive care $I_p(t)$ increase with the wage

[22]Higher income $Y(t)$ implies that individual b receives a higher wage rate $w(t)$ than individual a because the direct effect of higher wages is to increase income (the wage rate multiplied by the time spent working) while the secondary effect is, ceteris paribus, an increase in time spent working (own-time inputs in health investment and consumption decrease as a result of the increased opportunity cost of time; see equations 6.50, 6.60, 6.61 and 6.62). Both effects increase income $Y(t)$.

rate $w(t)$[23]. The net result of a wage increase on the level of job-related health stress $z(t)$ and the level of investment in preventive care $I_p(t)$ is thus unknown. The only exception to this pattern is unhealthy consumption. Both the direct cost $\pi_{C_u}(t)$ (equation 6.23) and the additional marginal cost $\pi_{dC_u}(t)$ (equation 6.24) of unhealthy consumption increase with the wage rate $w(t)$ and the result of a wage increase is a lower level of unhealthy consumption ($C_u^b < C_u^a$). In sum, because an increase in transitory wages increases both the various marginal costs and the marginal benefits, it could be either good or bad for health. The only unambiguous prediction is that a transitory increase in the wage rate leads to a reduction in unhealthy consumption.

If, however, the marginal production benefit of health $\varphi_H(t)$ is small compared to the user cost of health capital $\sigma_H(t)$[24], then $\sigma_H^b(t) \sim \sigma_H^a(t)$ and $\pi_{I_m}^b(t) \sim \pi_{I_m}^a(t)$. A higher wage rate then reduces the level of investment in curative care $I_m^b < I_m^a$, see equation (6.48). In contrast, the additional benefit of healthy consumption ($\varphi_{dC_h}(t)$; equation 6.21), the additional cost of unhealthy consumption ($\pi_{dC_u}(t)$; equation 6.24), the marginal cost of job-related health stress ($\pi_{dz}(t)$; equation 6.26) and the marginal benefit of preventive care ($\varphi_{dp}(t)$; equation 6.30) are unchanged in this scenario. The direct cost of healthy consumption ($\pi_{C_h}(t)$; equation 6.20) and unhealthy consumption ($\pi_{C_u}(t)$; equation 6.23) however increase with the wage rate $w(t)$ due to higher opportunity cost (see equations 6.55 and 6.56). Hence the level of healthy $C_h(t)$ and unhealthy $C_u(t)$ consumption is reduced. In addition, the marginal benefit of job-related health stress $\varphi_z(t)$ increases with the wage rate (equation 6.27) as does the marginal cost of investment in preventive care $\pi_p(t)$ (equations 6.27 and 6.44). As a result the level of job-related health stress $z(t)$ is higher and the level of investment in preventive care $I_p(t)$ lower. Thus, on balance, if the production benefit of health is small, a transitory wage change is bad for health. As a result of the greater opportunity cost of time the level of investment in health (curative and preventive) and the level of healthy consumption is lower and the level of job-related health stress is higher. The only exception is that the level of unhealthy consumption is lower.

Permanent wage change: Now consider again two individuals a and b. Individual b has a permanently higher level of wages, i.e., person b has greater life-time wealth (and

[23]In our formulation the marginal benefit of job-related health stress is increasing in the wage rate. Case and Deaton (2005) in their narrower definition of $z(t)$ as manual, risky labor (i.e., not including the psychosocial aspects of work), assume that the marginal benefit of additional manual labor is lower among those with higher wages.

[24]Note that it is always true that $\sigma_H(t) \geq \varphi_H(t)$, otherwise the investment in curative care would finance itself through negative net marginal costs of maintaining the health stock and individuals would achieve infinite health.

hence $q_A^b(0) < q_A^a(0)$). The effect of a "pure" asset increase (mostly due to a difference in endowed physical capital, e.g., assets $A(t)$) was described in section 6.3.3. In the case of a permanent wage increase, however, there is apart from the life-time wealth effect also the competing effect of the greater opportunity cost of time. There are reasons to believe that the wealth effect (greater life-time wealth) dominates the effect of the opportunity cost of time (higher current wages). First, this is consistent with the result by Dustmann and Windmeijer (2000) and Contoyannis *et al.* (2004) that a transitory wage increase affects health negatively while a permanent wage change affects health positively. Second, it is consistent with the rich literature on SES and health that consistently finds that high income individuals are healthier than low income individuals.

For simplicity, assume that the permanent effect dominates the transitory effect. The net result of a permanent wage change is somewhere in-between the "pure" asset effect described in section 6.3.3 and the situation described above for the transitory wage effect on health, with the former effect assumed to dominate the latter effect. Permanently higher wages are good for health as they increase investment in curative and preventive care, lower the level of job-related health stress and shift consumption toward healthy consumption. Our model suggests that the health benefit of a "pure" asset endowment would be larger than the effect of a "comparable" change in permanent life-time wages (similar change in the shadow-price of life-time wealth $q_A(0)$) due to the competing effect of the increased opportunity cost of time.

As a last point, one may be tempted to conclude that individuals invest less in curative care during middle and old age because of the high opportunity cost of time associated with high earnings at these ages (see equation 6.11). However, as health deteriorates with age the demand for curative care increases (see section 6.3.2). If the latter effect dominates the model is capable of reproducing the observation that young individuals invest little, the middle-aged invest more and the elderly invest substantially in curative care.

The effect of income on health:

In sum, a transitory increase in wages could affect health positively or negatively, depending on whether the higher marginal production benefit outweighs the increased opportunity cost of time. A permanent increase in income operates similar to an increase in wealth and is beneficial to health, though the wealth effect is moderated by the increased opportunity cost of time.

Education and health: the additional "efficiency" effect

Consider two individuals a and b who differ in their level of education E. Individual b has obtained more education but is otherwise identical. As a result, individual b has a higher wage rate $w_*(t)$ than individual a, which follows directly from the Mincer-type wage equation defined in (6.11). Thus the effect of education is similar to the effect of a permanent wage increase and the discussion presented in section 6.3.3 applies here as well.

But education also improves the efficiency of investment in curative care $\mu_{I_m}(t)$, investment in preventive care $\mu_{I_p}(t)$, and to a lesser extent healthy consumption $\mu_{C_h}(t)$ and unhealthy consumption $\mu_{C_u}(t)$[25]. Assume the effects of a permanent wage increase as described in section 6.3.3 and consider the additional effect of increased efficiency of investment in curative and preventive care. The marginal cost of investment in curative care $\pi_{I_m}(t)$ is determined by the first-order condition for health (equation 6.15). For a permanent wage change we found $\pi^b_{I_m} > \pi^a_{I_m}$ (or $\pi^b_{I_m} \sim \pi^a_{I_m}$ if the marginal production benefit of health $\varphi_H(t)$ is small compared to the user cost of health capital at the margin $\sigma_H(t)$). Since the marginal cost of investment in curative care $\pi_{I_m}(t)$ increases in the level $(I_m(t))$ and decreases in the efficiency $(\mu_{I_m}(t))$ of investment in curative care, a higher efficiency due to education implies a higher level of investment in curative care compared to the pure "wage" effect.

A higher efficiency of investment in preventive care μ_{I_p} lowers the marginal cost of preventive care $\pi_{I_p}(t)$ (equation 6.54) while the marginal benefit $\varphi_{dp}(t)$ (equation 6.30) is unchanged compared to the pure "wage" effect described in section 6.3.3[26]. Thus the optimal level of investment in preventive care is higher compared with the pure "wage" effect. If the efficiencies of healthy and unhealthy consumption do not (or only moderately) respond to education then the levels of healthy and unhealthy consumption are unchanged compared to the pure "wage" effect (section 6.3.3).

The effect of education on health:

An increase in education operates much in the same way as an increase in permanent income (the pure "wage" effect). But education has an independent effect on health,

[25] Grossman (1972a, 1972b) assumes that the higher educated are more efficient producers and consumers of curative care. We extend his definition to preventive care. However, it is less clear whether the higher educated are more efficient producers and consumers of consumption goods and services.

[26] The marginal cost of investment in curative care $\pi_{I_m}(t)$ is determined by the first-order condition for health (equation 6.15) and hence increased efficiency μ_{I_m} does not change the marginal cost $\pi_{I_m}(t)$, but rather leads to an increase in the level of investment in curative care.

over and above generating greater life-time income and wealth, through enhancing the efficiency of curative and preventive care. This leads to a higher demand for both curative and preventive care, while the effect on healthy and unhealthy consumption is ambiguous.

The formation of the SES health gradient

As the discussion in section 6.3.3 suggests greater initial wealth, permanently higher income (over the life cycle) and a higher level of education induce individuals to invest more in curative and in preventive care, shift consumption toward healthy consumption and enable individuals to afford healthier working and living environments. As a result, even for individuals who are initially equally healthy, the health trajectories of high and low SES individuals will begin to diverge. As the health of high SES individuals deteriorates at a slower pace, higher levels of health further reinforce the divergence of health trajectories as the marginal cost of unhealthy lifestyles, (e.g., unhealthy consumption and job-related health stress; see equations 6.24 and 6.26) and the marginal benefit of healthy lifestyles (e.g., healthy consumption and preventive care; see equations 6.21 and 6.30) are greater for higher levels of the health stock. Curative care is an exception. Greater health is associated with a lower level of investment in curative care. However, there is a competing effect in that high SES individuals invest more in curative care. In addition, the higher the health stock, the greater the earnings (e.g., see equation 6.9) such that reverse causality (from health to SES) could further reinforce the SES health gradient. Jointly these mechanisms gradually lead to cumulative health advantage with age.

The role of institutions

The extent to which wealth, life-time earnings and education relate to health outcomes is reasonably expected to depend on the institutional organization of the labor market and capital markets. In this paper we assume perfect capital and insurance markets[27]. Another important institutional feature is the organization of the health care market. In countries with universal health care coverage and low deductibles (such as, e.g., in Northern Europe) the price of curative care is small and health care is affordable to everyone. Still, the observed SES health gradient over the life cycle is strikingly consistent across countries with relatively low levels of protection from loss of work and health risks, such as the US, and those with stronger welfare systems, such as the Netherlands (e.g.,

[27]Note that an imperfect capital market could contribute to the association between income and health if individuals with little income cannot borrow and as such underinsure and invest too little in their health.

Smith, 1999; 2005a; 2007; Case and Deaton, 2005; Van Kippersluis *et al.*, 2010). A legitimate question to answer is to what extent our model explains this phenomenon.

In the following we focus on discussing the effect of wealth, i.e. differences in $q_A(0)$. The discussion of the effect of life-time earnings and education is similar (and follows the reasoning in sections 6.3.3). The marginal cost of curative care $\pi_{I_m}(t)$ is governed by the first-order condition for health (equation 6.15) and is largely determined by the level of the health stock $H(t)$ and by wealth $q_A(0)$ (assuming similar preferences). Thus, even if the price of curative care $p_m(t)$ is very small, the marginal cost of curative care $\pi_{I_m}(t)$ in general is not. The price of curative care is only one component that determines the marginal cost of curative care; the others are the opportunity cost of time (wages $w(t)$), the efficiency of curative care $\mu_{I_m}(t)$ and the level of curative care investment $I_m(t)$ (e.g., see equation 6.48 for an example functional form)[28]. A low price of curative care is likely to result in greater demand for curative care $I_m(t)$[29], but not in a lower marginal cost (assuming similar opportunity cost of time and efficiency of care).

Further, the marginal cost of curative care $\pi_{I_m}(t)$ cannot be constant across individuals with differences in health $H(t)$ and wealth $q_A(0)$. Wealthy individuals demand more curative care $I_m(t)$ and have a higher marginal cost of curative care $\pi_{I_m}(t)$ compared to their less affluent peers (assuming similar health). As the marginal cost of unhealthy lifestyles, (e.g., unhealthy consumption and job-related health stress; see equations 6.24 and 6.26) and the marginal benefit of healthy lifestyles (e.g., healthy consumption and preventive care; see equations 6.21 and 6.30) are proportional to the marginal cost of curative care $\pi_{I_m}(t)$, our model predicts not only differences in curative care between low SES and high SES individuals but also in the choice of working and living environment, the demand for preventive care and the level of healthy and unhealthy consumption. Consequently, higher wealth not only translates in greater investment in curative care, but also encourages healthy lifestyles and the choice of healthy working and living environments, irrespective of the way in which health care is provided. Hence also in countries with

[28]Inserting $p_m(t) = 0$ in equation (6.48) results in $\pi_{I_m}(t) = 0$ and is in conflict with the first-order condition for health (equation 6.15) which demands that the marginal cost of curative care $\pi_{I_m}(t)$ is positive. If the marginal cost of curative care could be negligible, individuals would demand infinite health. For low prices of curative care $p_m(t)$ it is reasonable to assume that the marginal cost of curative care is dominated by the opportunity cost of time. This suggests that the example functional form (equation 6.48) is not valid for low prices $p_m(t)$. Note that these specifications are not strictly part of the formulation and are provided to aid interpretation of the model. Other suitable functional forms can be constructed.

[29]However, in countries with universal health care coverage and low deductibles there are other means than price (out-of-pocket expenses) to reduce the consumption of care. For example the primary physician may act as gate keeper.

universal health care coverage and low deductibles there will be a significant SES health gradient.

6.4 Discussion

Two important mechanisms that are part of our formulation but that we did not explicitly take into account to facilitate derivations are differential mortality and health-related labor-force withdrawal. With regard to the former, the Grossman model provides a natural way to include length of life. In Grossman's original formulation (Grossman 1972a; 1972b) length of life is determined by a minimum health level $H_{\min}$. If health falls below this level $H(t) \leq H_{min}$ an individual dies. Ehrlich and Chuma (1990) specify a demand function for longevity in the Grossman model and argue that the demand for health must be derived in conjunction with that for longevity and consumption. With regard to the latter, as emphasized by Smith (1999; 2005a) and Case and Deaton (2005), reverse causality from health to income through labor force participation could be an important mechanism explaining the SES-health gradient. In our model, this could be incorporated by an endogenous retirement age (as in Galama *et al.*, 2009).

To include endogenous length of life and endogenous retirement, one has to assume suitable functional forms for the utility function $U(t)$, the relation between income $Y(t)$ and health, years of schooling and years of working experience, the dependence of the "natural" deterioration rate $d(t)$ on job-related health stress, healthy and unhealthy consumption, and preventive care and the relations between the outputs consumption $C_h(t)$ and $C_u(t)$ and health investments $I_m(t)$ and $I_p(t)$ and the inputs of own-time and goods/services purchased in the market. Using these assumed relations one can then in principle solve the first-order conditions and obtain the optimal solutions for healthy consumption $C_h(t)$, unhealthy consumption $C_u(t)$ and health $H(t)$ (as well as others). These solutions can then be inserted into the "indirect utility function", $V(R,T)$

$$V(R,T) \equiv \int_0^R U(t)e^{-\beta t}dt + \int_R^T U(t)e^{-\beta t}dt. \tag{6.31}$$

Given the increasing complexity of the problem one may have to resort to numerically solving for the optimal length of life T and the optimal retirement age R through identifying the maximal indirect utility $V(R,T)$ by varying R and T. Such analyses have been performed for retirement R (Galama *et al.*, 2009) and for length of life T (Ehrlich and Chuma, 1990).

Another potentially important mechanism, which we do not explicitly take into account, is the influence of the wider social context and social relationships of the family or neighborhood on health (House *et al.*, 1988; Robert, 1998; Kawachi and Berkman, 2003). Less affluent areas are more polluted, have lower quantity and quality of municipal services (such as policing, fire protection, and sanitation), have higher crime rates, and are associated with unhealthy lifestyles (Robert, 1998). Also, the social isolation induced by poor quality and quantity of social contacts is an important risk factor for health (House *et al.*, 1988).

To some extent these factors are exogenous to an individual; the neighborhood in which you live, including poor access to schooling and sanitation are not always controllable. In our model, this is partly captured, by the exogenous part of the deterioration rate (exogenous environmental factors). However, it is likely that social factors, such as the neighborhood in which you live, the household size, and the number of social contacts are partly endogenous to socioeconomic status (Robert, 1998). Some of this can be captured in our model by extending the definition of healthy consumption to include housing. Living in an affluent neighborhood is an expensive, yet health-promoting and utility-generating choice of individuals. However, the choice of neighborhood (housing) is a constrained choice: low SES individuals cannot afford to live in more affluent areas.

6.5 Conclusions

The aim of this paper is to provide a theory of the relation between health and socioeconomic status (SES) over the life cycle. Our life-cycle model incorporates multiple mechanisms that could explain (jointly) a large part of the observed disparities in health by SES. In our model, lifestyle factors (preventive care, healthy and unhealthy consumption), working conditions (physical and psychosocial health stresses), living conditions (housing, neighborhood social environment), curative care and the constraining effect of health on work are mechanisms through which SES (education, income, wealth) and health are related.

Our model is one of cumulative advantage in health (House *et al.*, 1994; Lynch *et al.*, 1997), and is consistent with fundamental cause theory (House *et al.*, 1990; Link and Phelan, 1995; Phelan and Link, 2005) which suggests that the pathways that cause disease may change over time and across populations but that the association between SES and health is maintained. Our model broadly describes the role of fundamental indicators of SES such as education, wealth and income (fundamental causes) and their relation to

health. This is not a theory of detailed disease pathways but rather a theory of social and economic resources and the way these relate to health.

Compared to Grossman (1972a; 1972b) and Case and Deaton (2005) the model presented in this paper contains several improvements and extensions: (i) The introduction of decreasing-returns-to-scale (DRTS) to investment in curative care addresses the issue of the indeterminacy of curative care, predicts differences in the effective health deterioration rate (as well as the level of health), and reproduces the observed negative relation between health and curative (medical) care. (ii) We have included the concept of healthy consumption (as well as unhealthy consumption as in Case and Deaton, 2005) and allow the demand for consumption to be governed both by the direct monetary price of consumption as well as the indirect health benefit (healthy consumption) or indirect health cost (unhealthy consumption). Case and Deaton (2005) on the other hand consider an unhealthy consumption good whose price is only paid in terms of health. (iii) We have broadened the concept of "job-related health stress" to include not only hard/risky labor (as in Case and Deaton, 2005) but also psychosocial aspects of work that are detrimental to health. (iv) We have argued that the effect of housing and neighborhood social environment can be included by extending the definition of healthy consumption as well as exogenous environmental factors to include relevant aspects of housing and neighborhood characteristics. (v) We have introduced the concept of preventive care. (vi) We have introduced length of life (mortality; as in Ehrlich and Chuma, 1990) and labor-force withdrawal (retirement; as in Galama *et al.*, 2009) as potentially important in explaining the SES health gradient, though in order to facilitate derivations we have treated them as exogenous in this work.

Considering the effect of differences in wealth between otherwise identical individuals, we find that individuals with greater initial wealth (pure "asset" effect) invest more in curative and in preventive care, have higher levels of both healthy and unhealthy consumption, yet shift their consumption toward healthy consumption, and are able to afford healthier working environments (associated with lower levels of job-related health stress). If healthy and unhealthy consumption are substitutes it is even possible that greater initial wealth leads to a reduction in unhealthy consumption. Smoking might be an example of this: high SES individuals smoke less than low SES individuals.

The mechanism through which initial wealth operates is by increasing the marginal cost of, and demand for, curative care. The greater marginal cost of curative care in turn increases the health benefit of (and hence demand for) preventive care and healthy consumption, and the health cost of (and hence reduced demand for) unhealthy working and living environments, and unhealthy consumption. Greater financial resources induce

healthy lifestyles, greater investment in curative care and protect individuals from the health risks of physical working conditions (e.g., hard labor) and/or psychosocial aspects of work (e.g., low status, limited control, repetitive work, etc) that are detrimental to health. Gradually, as time passes, higher levels of curative and preventive care, a shift toward healthy consumption and lower job-related physical and psychosocial stress lead to cumulative health advantage over the life cycle.

Considering the effect of differences in the wage rate between individuals (the pure "wage" effect), we found that a transitory wage increase (along an individual's wage profile, i.e., differences within an individual) is potentially bad for health as a result of the increased opportunity cost of time. However, the effect is hard to predict given the competing effect of the greater production benefit of health. A permanent wage increase on the other hand (i.e., an increase in life-time wages; differences across individuals) operates similar to an increase in initial wealth (pure "asset" effect) but its effect is moderated by the higher opportunity cost of time. Empirical analyses (e.g., Dustmann and Windmeijer, 2000; Contoyannis *et al.*, 2004) suggest a positive effect on health from a permanent wage increase and a negative effect from a transitory wage increase. This suggests that the permanent effect dominates the transitory (opportunity cost) effect. High income (over the life cycle) individuals, like wealthy individuals, invest more in curative and preventive care, engage in work that is more conducive to health (lower level of job-related health stress) and shift consumption toward healthy consumption.

The effect of a higher level of education is similar to an increase in permanent income, but with the additional effect of increasing the efficiency of the production and consumption of curative and preventive care. Compared to an increase in permanent income of the same magnitude (the pure "wage" effect), investment in curative and preventive care is higher due to the more efficient use of goods/services and own time. Compared to a ceteris paribus change in wealth (the pure "asset" effect), the effect of a higher education level could be either more beneficial or less beneficial to health, depending on whether the effect of greater efficiency outweighs the greater opportunity cost of higher wages. Given the strong effect of education on health outcomes observed in empirical studies (e.g., Grossman, 2000; Lleras-Muney, 2005; Oreopoulos, 2006; Cutler and Lleras-Muney, 2008) even when controlling for various other socioeconomic indicators, it is plausible that the greater efficiency more than compensates for the higher opportunity costs.

Our model makes a number of predictions. First, it predicts a gradual increase in health disparities with age as lower levels of healthy consumption, curative and preventive care and higher levels of unhealthy consumption and job-related health stress for low SES individuals lead to cumulative disadvantage in health over the life cycle.

Results from earlier studies (Ehrlich and Chuma, 1990; Galama *et al.*, 2009) suggest that the more rapidly worsening health of low SES individuals could lead to early withdrawal from the labor force and shorter life spans. Early withdrawal from the labor force may attribute to further cumulative disadvantage (widening of the SES health gradient) as the associated loss of income disproportionally affects low SES individuals. Shorter life spans of low SES individuals might explain the observed narrowing of SES health disparities beginning in late middle age as high SES individuals survive into old age. Properly accounting for labor-force withdrawal and mortality may require numerical approaches to solving the model (see section 6.4).

Second, the model predicts strong positive effects on health of high initial wealth (pure "asset" effect), permanently high wage income (pure "wage" effect) and education. Education potentially has the greater positive impact of the SES indicators because it not only gives rise to permanently higher income over the life cycle (which is qualitatively similar to wealth) but also increases the efficiency of investment in curative and preventive care.

Third, we expect empirical studies to find a strong causal impact of education, permanent income and initial wealth on health after substantial exposure, but not of current income. Our model predicts a strong effect of education on health and empirical studies are likely to confirm this prediction because education is generally obtained in childhood and early adulthood and hence ample time has passed by the time individuals are observed. The same cannot be said for income because current income is not a good measure of cumulative exposure to low income. Further, most empirical studies will control for education, and as such essentially control for life-time wealth. Thus the comparison is made between individuals of very similar life-time wealth and is more likely to pick up a transitory rather than a permanent wage effect on health. As we discussed, the sign of a transitory wage effect is ambiguous and could be small because of competing effects.

Fourth, our model suggests that the SES health gradient could be strong in countries with universal health care coverage and low deductions as well as in countries with large uninsured populations and high out-of-pocket expenditures as the marginal cost of curative care is determined by the level of the health stock on the one hand and wealth, life-time earnings and education on the other. In our model the organizational structure of health care in a country affects the level of investment in curative care but does not alter the marginal cost of care. And, it is the marginal cost of curative care that determines lifestyles and the choice of living and working environments. Hence also in countries with universal health care coverage and low deductibles there will be a significant SES health gradient.

Empirical estimation of the model is needed to assess the relative importance of mechanisms, study interactions between mechanisms, and disentangle the different patterns of causality. This will require developing structural- and reduced-form relations. Model estimates may allow for a comprehensive, as opposed to partial, explanation for the relations between SES and health, and to simulate the long-term effects of policy interventions.

Our model includes major mechanisms identified in a review of the literature as explaining (jointly) a large part of the observed disparities in health by SES. Given the complexity (e.g., Cutler *et al.*, 2008) of the various relations between SES and health, we have focussed on potential explanations that a) explain a large part of the gradient and b) are relatively straightforward to include in our theoretical framework. Two important mechanisms that are part of our formulation but were not explicitly taken into account to facilitate derivations are differential mortality (see Ehrlich and Chuma 1990) and health-related labor-force withdrawal (see Galama *et al.*, 2009). Given the increasing complexity of the model one may have to resort to numerically solving for the optimal length of life T and the optimal retirement age R (see section 6.4). An important extension of our model would be to incorporate insights from the literature on socioeconomic differences in the evolution of child health (e.g., Case *et al.*, 2002; Currie and Stabile, 2003; Currie *et al.*, 2007; Murasko, 2008), and from the literature on the impact of fetal and early-childhood conditions on health in adulthood (e.g., Barker *et al.*, 1993; Case *et al.*, 2005; van den Berg *et al.*, 2006). This might be feasible by including the production of health by the family (including the health of the child) similar to, e.g., Jacobson (2000) and Bolin *et al.* (2001; 2002a; 2002b). The role of the wider social context, social relationships, and other psycho-social risk factors (House *et al.*, 1988; 1994; Robert, 1998; Kawachi and Berkman, 2003) can partially be captured in our model by extending the definition of healthy consumption to include choice of housing / neighborhood social environment (see section 6.4). This might be further extended by including social capital similar to, e.g., Bolin *et al.* (2003). Insights from the behavioral-economic and psychological literature regarding myopia and lack of self-control (e.g., Blanchflower *et al.*, 2009) might be incorporated following Laibson (1998). Uncertainty (e.g., health shocks) could be included similar to, e.g., Cropper (1977), Dardanoni and Wagstaff (1990) and Liljas (1998).

Appendix

First-order conditions

Associated with the Lagrangian (equation 6.12) we have the following conditions:

$$\begin{aligned}
\dot{q}_A(t) &= -\frac{\partial \Im(t)}{\partial A(t)} \Rightarrow \\
\dot{q}_A(t) &= -\delta q_A(t) \Leftrightarrow \\
q_A(t) &= q_A(0)e^{-\delta t}, \qquad (6.32)
\end{aligned}$$

$$\begin{aligned}
\dot{q}_H(t) &= -\frac{\partial \Im(t)}{\partial H(t)} \Rightarrow \\
\dot{q}_H(t) &= q_H(t)d(t) - \frac{\partial U(t)}{\partial H(t)}e^{-\beta t} - q_A(0)\frac{\partial Y(t)}{\partial H(t)}e^{-\delta t} \\
&= q_H(t)d(t) - \frac{\partial U(t)}{\partial H(t)}e^{-\beta t} - q_A(0)\varphi_H(t)e^{-\delta t}, \qquad (6.33)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Im(t)}{\partial X_h(t)} &= 0 \Rightarrow \\
\frac{\partial U(t)}{\partial C_h(t)} &= q_A(0)\frac{p_{X_h}(t)}{\partial C_h(t)/\partial X_h(t)}e^{(\beta-\delta)t} + q_H(t)\frac{\partial d(t)}{\partial C_h(t)}H(t)e^{\beta t} \\
&\equiv q_A(0)\,\pi_{C_h}(t)e^{(\beta-\delta)t} - q_H(t)\frac{\varphi_{dC_h}(t)}{\pi_{I_m}(t)}e^{\beta t}, \qquad (6.34)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Im(t)}{\partial \tau_{C_h}(t)} &= 0 \Rightarrow \\
\frac{\partial U(t)}{\partial C_h(t)} &= q_A(0)\frac{w(t)}{\partial C_h(t)/\partial \tau_{C_h}(t)}e^{(\beta-\delta)t} + q_H(t)\frac{\partial d(t)}{\partial C_h(t)}H(t)e^{\beta t} \\
&\equiv q_A(0)\,\pi_{C_h}(t)e^{(\beta-\delta)t} - q_H(t)\frac{\varphi_{dC_h}(t)}{\pi_{I_m}(t)}e^{\beta t}, \qquad (6.35)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Im(t)}{\partial X_u(t)} &= 0 \Rightarrow \\
\frac{\partial U(t)}{\partial C_u(t)} &= q_A(0)\frac{p_{X_u}(t)}{\partial C_u(t)/\partial X_u(t)}e^{(\beta-\delta)t} + q_H(t)\frac{\partial d(t)}{\partial C_u(t)}H(t)e^{\beta t} \\
&\equiv q_A(0)\,\pi_{C_u}(t)e^{(\beta-\delta)t} + q_H(t)\frac{\pi_{dC_u}(t)}{\pi_{I_m}(t)}e^{\beta t}, \qquad (6.36)
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \Im(t)}{\partial \tau_{C_u}(t)} &= 0 \Rightarrow \\
\frac{\partial U(t)}{\partial C_u(t)} &= q_A(0)\frac{w(t)}{\partial C_u(t)/\partial \tau_{C_u}(t)}e^{(\beta-\delta)t} + q_H(t)\frac{\partial d(t)}{\partial C_u(t)}H(t)e^{\beta t} \\
&\equiv q_A(0)\,\pi_{C_u}(t)e^{(\beta-\delta)t} + q_H(t)\frac{\pi_{dC_u}(t)}{\pi_{I_m}(t)}e^{\beta t}, \qquad (6.37)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Im(t)}{\partial m_m(t)} &= 0 \Rightarrow \\
q_H(t) &= q_A(0)\left\{\frac{p_m(t)I_m(t)^{1-\alpha}}{\alpha[\partial I_m(t)/\partial m_m(t)]}\right\}e^{-\delta t} \\
&\equiv q_A(0)\,\pi_{I_m}(t)e^{-\delta t}, \qquad (6.38)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Im(t)}{\partial \tau_{I_m}(t)} &= 0 \Rightarrow \\
q_H(t) &= q_A(0)\left\{\frac{w(t)I_m(t)^{1-\alpha}}{\alpha[\partial I_m(t)/\partial \tau_{I_m}(t)]}\right\}e^{-\delta t} \\
&\equiv q_A(0)\,\pi_{I_m}(t)e^{-\delta t}, \qquad (6.39)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Im(t)}{\partial z(t)} &= 0 \Rightarrow \\
0 &= q_H(t)\frac{\partial d(t)}{\partial z(t)}H(t) - q_A(0)\frac{\partial Y(t)}{\partial z(t)}e^{-\delta t} \\
&\equiv q_H(t)\frac{\pi_{dz}(t)}{\pi_{I_m}(t)} - q_A(0)\,\varphi_z(t)e^{-\delta t}, \qquad (6.40)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Im(t)}{\partial m_p(t)} &= 0 \Rightarrow \\
0 &= q_H(t)\frac{\partial d(t)}{\partial I_p(t)}H(t) + q_A(0)\frac{p_p(t)}{\partial I_p(t)/\partial m_p(t)}e^{-\delta t} \\
&\equiv -q_H(t)\frac{\pi_{dI_p}(t)}{\pi_{I_m}(t)} + q_A(0)\,\pi_{I_p}(t)e^{-\delta t}, \qquad (6.41)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Im(t)}{\partial \tau_{I_p}(t)} &= 0 \Rightarrow \\
0 &= q_H(t)\frac{\partial d(t)}{\partial I_p(t)}H(t) + q_A(0)\frac{w(t)}{\partial I_p(t)/\partial \tau_{I_p}(t)}e^{-\delta t}, \\
&\equiv -q_H(t)\frac{\pi_{dI_p}(t)}{\pi_{I_m}(t)} + q_A(0)\,\pi_{I_p}(t)e^{-\delta t}, \qquad (6.42)
\end{aligned}
$$

Equation (6.34) or (6.35) combined with (6.38) or (6.39) provide the first-order condition for maximization of (6.1) with respect to consumption (equation 6.19). Similarly, equation (6.36) or (6.37) combined with (6.38) or (6.39) provide the first-order condition for maximization of (6.1) with respect to consumption (equation 6.22). Using (6.38) or (6.39) to obtain an expression for $\dot{q}_H(t)$ and substituting the results for $q_H(t)$ and $\dot{q}_H(t)$ in (6.33) we find the first-order condition for maximization of (6.1) with respect to health (equation 6.15). Combining equations (6.38) or (6.39) and (6.40) to eliminate $q_H(t)$ we find the first-order condition for maximization of (6.1) with respect to job-related health stress (equation 6.25). Lastly, combining equations (6.38) or (6.39) and (6.41) or (6.42) to eliminate $q_H(t)$ we find the first-order condition for maximization of (6.1) with respect to preventive care (equation 6.28).

Simple functional forms

To aid interpretation of the first-order conditions it is helpful to specify suitable functional forms for the utility function $U(t)$, the relation between income $Y(t)$ and health, years of schooling and years of working experience, the dependence of the "natural" deterioration rate $d(t)$ on job-related health stress, healthy and unhealthy consumption, and preventive care and the relations between the ouputs consumption $C_h(t)$ and $C_u(t)$ and health investments $I_m(t)$ and $I_p(t)$ and the inputs of own-time and goods/services purchased in the market.

Assume that sick time is a power law in health

$$s[H(t)] = \Omega \left\{ 1 - [H(t)/H_{max}]^{\beta_2} \right\}, \tag{6.43}$$

where β_2 is a positive constant.[30]

We assume a Mincer-type wage relation for the "effortless" wage rate $w_*(t)$ (e.g., Mincer, 1974; see equation 6.11). Thus wages $w(t)$ (equations 6.10 and 6.11) increase with years of schooling E, experience $x(t)$ and the level of job-related health stress $z(t)$.

We then find (see equation 6.9),

$$Y[H(t), z(t); E, x(t)] = w_*(t)[1 + z(t)]^{\gamma_w} \Omega \left[\frac{H(t)}{H_{max}} \right]^{\beta_2} [1 - \epsilon(t)] \tag{6.44}$$

[30] Wagstaff (1986) and Grossman (1972a, 1972b, 2000) assume $s(t) = \beta_1 H(t)^{-\beta_2}$. In keeping with the literature, we follow their formulation but note that negative values of β_1 and β_1 can be allowed as long as $\beta_1 \beta_2 > 0$. We specify sick time instead as $s(t) = \beta_0 - \beta_1 H(t)^{\beta_2}$. We assume that the health stock has natural bounds between 0 and a maximal value H_{max}. The relevant properties of sick time are then as follows: $\lim_{H \to H_{max}} s[H(t)] = 0$ and $\lim_{H \downarrow 0} s[H(t)] = \Omega$. These properties lead to expression (6.43).

where the term $\Omega[H(t)/H_{max}]^{\beta_2}$ represents healthy time $h(t) \equiv \Omega - s[H(t)]$ and $\epsilon(t) \equiv [\tau_{I_m}(t)+\tau_{I_p}(t)+\tau_{C_h}(t)+\tau_{C_u}(t)]/h(t)$ is the ratio of the total time input into the production of curative care, preventive care and consumption to the total amount of healthy time $h(t)$.

Further, using equations (6.9), (6.10) and (6.44), we have

$$\varphi_H(t) = \beta_2 \Omega H_{max}^{-\beta_2} w_*(t)[1+z(t)]^{\gamma_w} H(t)^{-(\beta_2+1)}. \tag{6.45}$$

We assume the health deterioration rate $d(t)$ to be a function of various endogenous as well as exogenous functions

$$d(t) = d_* e^{\beta_t t + \beta_\xi \xi(t)}[1+z(t)]^{\gamma_d}[1+C_h(t)]^{-\gamma_{C_h}}[1+C_u(t)]^{\gamma_{C_u}}[1+I_p(t)]^{-\gamma_p}, \tag{6.46}$$

where $\xi(t)$ is a vector of exogenous environmental variables (e.g., exogenous living conditions) that affect the deterioration rate. The deterioration rate $d(t)$ increases with job-related health stress $z(t)$, decreases with healthy consumption $C_h(t)$, increases with unhealthy consumption $C_u(t)$ and decreases with investment in preventive care $I_p(t)$; $\gamma_d \geq 0$, $\gamma_{C_h} \geq 0$, $\gamma_{C_u} \geq 0$ and $\gamma_p \geq 0$. For minimal job-related health stress ($z(t) = 0$), "health neutral" consumption ($\gamma_{C_h} = \gamma_{C_u} = 0$) and no preventive care ($I_p(t) = 0$) individuals "age" at an exogenous rate $d(t) = d(0)e^{\beta_t t + \beta_\xi[\xi(t)-\xi(0)]}$ (as in Cropper, 1981; Wagstaff, 1986).

Assume a production function of the following form

$$I_m(t) = \mu_{I_m}(t) m_m(t)^{\alpha_{I_m}} \tau_{I_m}(t)^{\beta_{I_m}}, \tag{6.47}$$

where $\mu_{I_m}(t)$ is an efficiency factor and α_{I_m} and β_{I_m} are the elasticities of investment in curative care $I_m(t)$ with respect to goods and services $m_m(t)$ purchased in the market (e.g., curative care) and with respect to own-time $\tau_{I_m}(t)$, respectively. The production function exhibits DRTS for $0 < \alpha_{I_m} + \beta_{I_m} < 1$, CRTS for $\alpha_{I_m} + \beta_{I_m} = 1$ and increasing-returns-to-scale (IRTS) for $\alpha_{I_m} + \beta_{I_m} > 1$. Using equations (6.17) and (6.47) we have

$$\begin{aligned} \pi_{I_m}(t) &= \frac{p_m(t)^{\alpha_{I_m}/(\alpha_{I_m}+\beta_{I_m})}\, w(t)^{\beta_{I_m}/(\alpha_{I_m}+\beta_{I_m})}}{\alpha \mu_{I_m}(t)^{1/(\alpha_{I_m}+\beta_{I_m})} \alpha_{I_m}(t)^{\alpha_{I_m}/(\alpha_{I_m}+\beta_{I_m})} \beta_{I_m}(t)^{\beta_{I_m}/(\alpha_{I_m}+\beta_{I_m})}} \\ &\times I_m(t)^{(1-\alpha)+[1-(\alpha_{I_m}+\beta_{I_m})]/(\alpha_{I_m}+\beta_{I_m})}. \end{aligned} \tag{6.48}$$

Further

$$m_m(t) = \frac{p_m(t)^{-\beta_{I_m}/(\alpha_{I_m}+\beta_{I_m})}\, w(t)^{\beta_{I_m}/(\alpha_{I_m}+\beta_{I_m})}}{\mu_{I_m}(t)^{1/(\alpha_{I_m}+\beta_{I_m})} \alpha_{I_m}(t)^{-\beta_{I_m}/(\alpha_{I_m}+\beta_{I_m})} \beta_{I_m}(t)^{\beta_{I_m}/(\alpha_{I_m}+\beta_{I_m})}} I_m(t)^{1/(\alpha_{I_m}+\beta_{I_m})} \tag{6.49}$$

$$\tau_{I_m}(t) = \frac{p_m(t)^{\alpha_{I_m}/(\alpha_{I_m}+\beta_{I_m})}\, w(t)^{-\alpha_{I_m}/(\alpha_{I_m}+\beta_{I_m})}}{\mu_{I_m}(t)^{1/(\alpha_{I_m}+\beta_{I_m})} \alpha_{I_m}(t)^{\alpha_{I_m}/(\alpha_{I_m}+\beta_{I_m})} \beta_{I_m}(t)^{-\alpha_{I_m}/(\alpha_{I_m}+\beta_{I_m})}} I_m(t)^{1/(\alpha_{I_m}+\beta_{I_m})} \tag{6.50}$$

We can derive similar relationships for investment in preventive care $I_p(t)$, healthy consumption $C_h(t)$ and unhealthy consumption $C_u(t)$, assuming the following analogous functional forms:

$$I_p(t) = \mu_{I_p}(t) m_p(t)^{\alpha_{I_p}} \tau_{I_p}(t)^{\beta_{I_p}}, \tag{6.51}$$

$$C_h(t) = \mu_{C_h}(t) X_{C_h}(t)^{\alpha_{C_h}} \tau_{C_h}(t)^{\beta_{C_h}}, \tag{6.52}$$

$$C_u(t) = \mu_{C_u}(t) X_{C_u}(t)^{\alpha_{C_u}} \tau_{C_u}(t)^{\beta_{C_u}}, \tag{6.53}$$

which, using equations (6.20), (6.23) and (6.30), lead to

$$\pi_{I_p}(t) = \frac{p_p(t)^{\alpha_{I_p}/(\alpha_{I_p}+\beta_{I_p})}\, w(t)^{\beta_{I_p}/(\alpha_{I_p}+\beta_{I_p})}}{\mu_{I_p}(t)^{1/(\alpha_{I_p}+\beta_{I_p})} \alpha_{I_p}^{\alpha_{I_p}/(\alpha_{I_p}+\beta_{I_p})} \beta_{I_p}^{\beta_{I_p}/(\alpha_{I_p}+\beta_{I_p})}} I_p(t)^{[1-(\alpha_{I_p}+\beta_{I_p})]/(\alpha_{I_p}+\beta_{I_p})}, \tag{6.54}$$

$$\pi_{C_h}(t) = \frac{p_{X_h}(t)^{\alpha_{C_h}/(\alpha_{C_h}+\beta_{C_h})}\, w(t)^{\beta_{C_h}/(\alpha_{C_h}+\beta_{C_h})}}{\mu_{C_h}(t)^{1/(\alpha_{C_h}+\beta_{C_h})} \alpha_{C_h}^{\alpha_{C_h}/(\alpha_{C_h}+\beta_{C_h})} \beta_{C_h}^{\beta_{C_h}/(\alpha_{C_h}+\beta_{C_h})}} C_h(t)^{[1-(\alpha_{C_h}+\beta_{C_h})]/(\alpha_{C_h}+\beta_{C_h})}, \tag{6.55}$$

$$\pi_{C_u}(t) = \frac{p_{X_u}(t)^{\alpha_{C_u}/(\alpha_{C_u}+\beta_{C_u})}\, w(t)^{\beta_{C_u}/(\alpha_{C_u}+\beta_{C_u})}}{\mu_{C_u}(t)^{1/(\alpha_{C_u}+\beta_{C_u})} \alpha_{C_u}^{\alpha_{C_u}/(\alpha_{C_u}+\beta_{C_u})} \beta_{C_u}^{\beta_{C_u}/(\alpha_{C_u}+\beta_{C_u})}} C_u(t)^{[1-(\alpha_{C_u}+\beta_{C_u})]/(\alpha_{C_u}+\beta_{C_u})}, \tag{6.56}$$

where the efficiency factors $\mu(t)$ and the elasticities α and β are defined analogous to those for investment in preventive care $I_m(t)$ (see above). Further

$$m_p(t) = \frac{p_p(t)^{-\beta_{I_p}/(\alpha_{I_p}+\beta_{I_p})}\, w(t)^{\beta_{I_p}/(\alpha_{I_p}+\beta_{I_p})}}{\mu_{I_p}(t)^{1/(\alpha_{I_p}+\beta_{I_p})} \alpha_{I_p}(t)^{-\beta_{I_p}/(\alpha_{I_p}+\beta_{I_p})} \beta_{I_p}(t)^{\beta_{I_p}/(\alpha_{I_p}+\beta_{I_p})}} I_p(t)^{1/(\alpha_{I_p}+\beta_{I_p})}, \tag{6.57}$$

$$X_{C_h}(t) = \frac{p_{X_h}(t)^{-\beta_{C_h}/(\alpha_{C_h}+\beta_{C_h})}\, w(t)^{\beta_{C_h}/(\alpha_{C_h}+\beta_{C_h})}}{\mu_{C_h}(t)^{1/(\alpha_{C_h}+\beta_{C_h})} \alpha_{C_h}(t)^{-\beta_{C_h}/(\alpha_{C_h}+\beta_{C_h})} \beta_{C_h}(t)^{\beta_{C_h}/(\alpha_{C_h}+\beta_{C_h})}} C_h(t)^{1/(\alpha_{C_h}+\beta_{C_h})}, \tag{6.58}$$

$$X_{C_u}(t) = \frac{p_{X_u}(t)^{-\beta_{C_u}/(\alpha_{C_u}+\beta_{C_u})}\, w(t)^{\beta_{C_u}/(\alpha_{C_u}+\beta_{C_u})}}{\mu_{C_u}(t)^{1/(\alpha_{C_u}+\beta_{C_u})} \alpha_{C_u}(t)^{-\beta_{C_u}/(\alpha_{C_u}+\beta_{C_u})} \beta_{C_u}(t)^{\beta_{C_u}/(\alpha_{C_u}+\beta_{C_u})}} C_u(t)^{1/(\alpha_{C_u}+\beta_{C_u})}, \tag{6.59}$$

and

$$\tau_{I_p}(t) = \frac{p_p(t)^{\alpha_{I_p}/(\alpha_{I_p}+\beta_{I_p})}\, w(t)^{-\alpha_{I_p}/(\alpha_{I_p}+\beta_{I_p})}}{\mu_{I_p}(t)^{1/(\alpha_{I_p}+\beta_{I_p})} \alpha_{I_p}(t)^{\alpha_{I_p}/(\alpha_{I_p}+\beta_{I_p})} \beta_{I_p}(t)^{-\alpha_{I_p}/(\alpha_{I_p}+\beta_{I_p})}} I_p(t)^{1/(\alpha_{I_p}+\beta_{I_p})}, \tag{6.60}$$

$$\tau_{C_h}(t) = \frac{p_{X_h}(t)^{\alpha_{C_h}/(\alpha_{C_h}+\beta_{C_h})}\, w(t)^{-\alpha_{C_h}/(\alpha_{C_h}+\beta_{C_h})}}{\mu_{C_h}(t)^{1/(\alpha_{C_h}+\beta_{C_h})} \alpha_{C_h}(t)^{\alpha_{C_h}/(\alpha_{C_h}+\beta_{C_h})} \beta_{C_h}(t)^{-\alpha_{C_h}/(\alpha_{C_h}+\beta_{C_h})}} C_h(t)^{1/(\alpha_{C_h}+\beta_{C_h})}, \tag{6.61}$$

$$\tau_{C_u}(t) = \frac{p_{X_u}(t)^{\alpha_{C_u}/(\alpha_{C_u}+\beta_{C_u})}\, w(t)^{-\alpha_{C_u}/(\alpha_{C_u}+\beta_{C_u})}}{\mu_{C_u}(t)^{1/(\alpha_{C_u}+\beta_{C_u})} \alpha_{C_u}(t)^{\alpha_{C_u}/(\alpha_{C_u}+\beta_{C_u})} \beta_{C_u}(t)^{-\alpha_{C_u}/(\alpha_{C_u}+\beta_{C_u})}} C_u(t)^{1/(\alpha_{C_u}+\beta_{C_u})}, \tag{6.62}$$

Chapter 7

Discussion

Health differs by socioeconomic status (SES). Low SES individuals not only start adulthood with a lower level of health, but their health also deteriorates faster with age than their high SES peers. From the analyses employed in this thesis and the related literature two causal pathways prove particularly important in generating inequalities in health by socioeconomic status. First, education has a large and non-vanishing effect on health throughout life, even persisting beyond the age of 80. This thesis confirmed that education is not only *'the most important correlate of good health'* (Grossman, 2003), but part of this association stems from a genuine, causal effect of education on health outcomes, including mortality. Education is shown to significantly reduce mortality even in old age. Our analysis reveals that, for Dutch males surviving to the age of 81, an additional year of schooling reduces the probability that they will die before reaching 89 by 2-3 percentage points, or 4-6% relative to the baseline probability.

Second, this thesis confirmed that there is a large effect of health on income that operates through constrained employment. It is shown that ill-health significantly and permanently reduces both personal and household disposable income. In the Netherlands, the drop in personal disposable income is around 1,600 Euro on a yearly basis three years after the initial shock. The effects on household disposable income are more than twice as large, reaching an amount of 3,600 Euro three years after the health shock, or around 8 percent in relative terms, indicating that there are spill-over effects from the health shock to other household members.

In contrast, although large health differences across income groups are visible, consensus among economists is growing that these do not derive from a causal impact of income on health (Adams *et al.*, 2003; Contoyannis *et al.*, 2004; Smith, 2005a; Case and Deaton, 2005; Frijters *et al.*, 2005; Cutler and Lleras-Muney, 2008). This view is however not unequivocally accepted. Replication is still needed and controversy remains on the extent

to which these findings apply uniformly to different population segments (Marmot, 2002; Herd *et al.*, 2008).

The theoretical model presented in this thesis predicts that the main mechanism through which higher socioeconomic status is translated into higher health is by increasing the marginal cost of, and demand for, curative care. The greater marginal cost of curative care increases the health benefit of (and hence demand for) preventive care and healthy consumption, and the health cost of (and hence reduced demand for) unhealthy working and living environments, and unhealthy consumption. Greater financial resources induce healthy lifestyles, greater investment in curative care and protect individuals from the health risks of physical working conditions (e.g., hard labor) and/or psychosocial aspects of work (e.g., low status, limited control, repetitive work, etc) that are detrimental to health. Gradually, as time passes, higher levels of curative and preventive care, a shift toward healthy consumption and lower job-related physical and psychosocial stress lead to cumulative health advantage over the life cycle.

In contrast, low SES individuals have little human and financial capital, such that they have few options but to resort to their health capital in production and consumption. Given low wealth and low wages they have to increase their nominal wages by accepting unhealthier working conditions such as manual labor, low status, limited control, and/or repetitive work. Also, given lower wages and less wealth, the benefits of health protection are lower. Relative to higher SES individuals they will invest less in preventive care and healthy consumption, and more in unhealthy consumption as a result. What should be emphasized here is that the choices of lower socioeconomic groups of an unhealthier lifestyle and working conditions represent *optimal, though heavily constrained, choices* (Case and Deaton, 2005).

This thesis propagates a theory of cumulative disadvantage over the life cycle. This would start with early-life conditions and parental background affecting health in childhood and education. We do not cover the literature on the lifetime economic and health consequences of early childhood conditions, but the evidence base to support causal effects is growing (Barker, 1995; Case *et al.*, 2002; 2005; Currie and Stabile, 2003; Currie *et al.*, 2007; van den Berg *et al.*, 2006; van den Berg and Lindeboom, 2007). From adulthood, SES and health start interacting and influencing each other. The theoretical model in this thesis predicts a gradual increase in health inequalities with age, as lower levels of healthy consumption, curative and preventive care and higher levels of unhealthy consumption and job-related health stress for low SES individuals lead to cumulative disadvantage in health over the life cycle. More rapidly worsening health of low SES individuals could lead to early withdrawal from the labor force and shorter life-spans. So, the health-related

earnings losses that are partly responsible for the income-health gradient are themselves, to some extent, the result of socioeconomic differences in lifetime opportunities and their impact on health. Beyond that, it is shown in this thesis that even after accounting for the higher risk of a sudden deterioration of health — a health shock — the lower socioeconomic groups are least protected against income loss in case such a shock arrives. This represents double jeopardy for lower SES groups and is likely to exacerbate socioeconomic inequalities in health.

The objective of this thesis is to uncover the causal pathways and underlying mechanisms responsible for socioeconomic differences in health by taking an economic approach. As shown in this thesis, an economic approach has the potential to significantly advance the field. The statistical rigor of econometric analyses triggers it possible to arrive at causal effects by isolating exogenous variation in socioeconomic status and health. Moreover, economic theory provides an extremely useful behavioral framework to an improved understanding of constrained individual decision-making in the health sphere. The combination of econometrics and economics is a distinguishing feature of this thesis and proved extremely valuable towards understanding socioeconomic differences in health. For example, chapter 4 confirms that education causally affects mortality exploiting a change in the Dutch compulsory schooling law in a Regression Discontinuity Design (RDD). After having confirmed the existence of a causal effect, the economic model in chapter 6 proposes some mechanisms through which an individual's education might affect health behaviors, and in turn mortality.

Policy implications are numerous. First, regarding the effect of education on health; education not only raises earnings over the life cycle (e.g. Card, 1999), it also extends the horizon of the lifetime. This is an important finding in the context of rising education levels and the ageing of populations worldwide. As more and better educated individuals reach old age, we can anticipate that mortality rates among the elderly will fall further and populations will become even more 'grey'. Also, pensions will be stretched further to meet the consumption needs of an extended old age. The other side of the coin is that poorly educated individuals die earlier, enjoying a less extended period of retirement. There is a double injustice here. Not only does a lack of education lead to a deprivation of life itself, but it implies a lower return on investments in pensions made over the working life.

Second consideration is that while disability insurance (DI) is developed to insure the personal earnings loss, it cannot prevent disposable household income to drop by a larger amount. We find that on average there are spillover effects to other household members, possibly reflecting a fall in the earnings of the spouse as a result of meeting the care needs of the disabled person. This explanation is supported by the fact that for those

experiencing a more severe shock to health, the concomitant drop in household income is even larger. Whether these actual replacement rates are below or above the desired level, and should provoke any concern for policy-makers, depends. Definitely the concomitant effects on other household members of the disabled individual should be studied further to understand their motives. If the main motive is a voluntary increase in joint leisure (see e.g. Hurd, 1990), the increase leisure might, to a more or lesser extent, compensate for the income loss. Yet if spouses are urged to decrease their labour participation in order to engage into care giving responsibilities, this could be an inefficient solution — potentially detrimental to welfare.

The drop in disposable income after a health shock is not evenly spread over society. Groups in the lowest socio-economic classes — those in the bottom quartile of personal income, and the lower educated — are both the most vulnerable to the health-induced risk of job loss, and their job loss translates into higher relative drops in personal and household disposable income. Consequently, the lower socioeconomic groups suffer from double jeopardy: lower socioeconomic groups are most likely to receive a shock to their health (e.g. Smith, 2005c), and it follows from this thesis that they are least protected against loss of job and income in case a health shock arrives. The average actual replacement rate of 80% faced by individuals in the lowest income quartile compared to the average replacement rates, does also imply that the experience of a health shock by this group would not only raise the observed income inequalities, but also put a non-negligible proportion of individuals in this group at risk of poverty.

Consistent with other studies (e.g. Kunst *et al.*, 2005), the evidence presented in this thesis shows no evidence of falling socioeconomic inequality in health across generations in Europe. While a stable level of socioeconomic inequality may be considered a failure, it should be judged alongside the US evidence of increasing socioeconomic inequality (Pappas *et al.*, 1993; Deaton and Paxson, 1998). Whether stable inequalities are necessarily problematic, is open to debate. An extremely important caveat applies to the goal of reduction of socioeconomic inequalities in health. As Deaton (2002a) argues some Pareto improvements in average population health, such as an innovative medical technology that is adopted sooner by the higher educated, will lead to larger health inequalities. If the sole goal is reducing health inequalities these treatments should be abandoned, effectively resulting in some people dying who could have lived, without preventing any other deaths. Therefore, a policy should never be focused solely on reducing health disparities, it should also account for average health, and wider well-being (Deaton, 2002a; CSDH, 2008).

Overturning these processes of cumulative disadvantage and double jeopardies represents a major challenge even to countries as egalitarian as the Netherlands. The potential

rewards are, however, immense. Getting closer to a goal of equality of opportunity could raise national welfare and wealth by breaking or at least limiting the connection between low education, poor health and early labor force exit. If average health improves, there could also be substantial savings on age-specific health care expenditures and increases can be expected in labor supply, productivity, educational attainment, and savings and investment (Bloom and Canning, 2000). Since education has been shown to causally impact health and mortality, a potentially fruitful way of improving average health and reducing health inequalities is stimulating education among the lower SES groups. Of course, breaking the cycle requires identifying the specific mechanisms that link education with later life health outcomes. The theoretical framework in this thesis provides a first step towards disentangling these mechanisms. Yet, not all educational disparities in health derive from inequality of opportunity in the sphere of health choices. Some must reflect ill-health in childhood and adolescence constraining choices over education. Besides ensuring adequate access to schooling for less healthy children, education policy can do little to address this source of health disparities. In general, if third factors such as cognitive ability or time preference partly contribute to the observed health inequalities, there is little public policy can do to prevent these inequalities.

Policies aimed at improving health conditions at work are potentially important in preventing health related labor force exits, thereby extending working lives and maintaining the financial sustainability of social security systems. Financial protection against health related loss of employment inevitably creates moral hazard, which has been a particular problem in the Netherlands. Therefore, it seems crucial to enact policies in the workplace that can both prevent the development of health problems and reduce the impact of those problems on work capacity, by providing appropriate support to partially disabled workers and those with a diagnosed condition. In that sense, recent Dutch policy changes seem to operate in the right direction. Yet there are obvious limitations on the extent to which workplace interventions can reduce health inequalities. They can do little to halt health deterioration that derives from early life experiences. Nor can they do much to correct differences in health by income that derive from education related differences in life style.

While this thesis has contributed to understanding the main causal pathways, and has proposed several potentially important underlying mechanisms, empirical research is clearly warranted to distinguish between important and less important mechanisms generating the socioeconomic gradient in health. Scholars are invited and encouraged to continue replicating existing studies on the causal effects of SES on health and vice versa, whether based on economic theory or using quasi-natural experiments, both of which have proved extremely valuable in our improved understanding of socioeconomic inequalities.

Also, although this thesis has provided a foundation of a theory of socioeconomic health inequality, the theory needs to be developed further, including more realistic notions such as uncertainty, insights from behavioural economics, and to include potentially other important mechanisms, such as social or family contexts, and other psychosocial mechanisms such as rank. It seems especially important to include labour force withdrawal into the framework, given that this has been identified as one of the major contributors of inequalities in health by socioeconomic status.

Economists are especially invited to take note of recent research in other disciplines, including demography, sociology and epidemiology, which have proposed interesting theories of inequalities in health by SES, and are potentially very useful to include in our existing theories or empirical research. On the other hand, researchers of other disciplines are encouraged to take of note of economic models to provide a different mode of analysis, and to apply the rigour and prudence of economists in claiming and estimating causal effects. A cross-fertilisation of disciplines and modes of analysis is likely to be the only way forward in reducing socioeconomic inequalities in health to an acceptable level, tackling double injustices, and improving welfare.

Nederlandse Samenvatting

Inleiding en motivatie

Gezondheid is van fundamenteel belang. Gezondheid heeft zowel intrinsieke als instrumentele waarde: het is niet alleen van centraal belang voor het menselijk welzijn, maar een goede gezondheid is ook een vereiste om als mens te kunnen functioneren (Anand, 2002; Sen, 2002). Het recht op het hoogst mogelijke gezondheidsniveau is vastgelegd in de statuten van de World Health Organization (WHO) en talloze internationale verdragen (bv. United Nations, 2000). Ondanks dit recht verkeert het grootste gedeelte van de wereldbevolking niet in de goede gezondheid die biologisch gezien mogelijk is.

In het bijzonder is er een sterke relatie tussen gezondheid en sociaaleconomische klasse, waarbij mensen in de lagere sociaaleconomische klassen een veel mindere gezondheid en lagere levensverwachting hebben dan mensen in de hogere sociaaleconomische klassen. In Nederland leven hoger opgeleiden — mensen met een hbo of universitaire opleiding — zo'n 6 tot 7 jaar langer dan lager opgeleiden — mensen die alleen basisschool hebben gevolgd. Het verschil in levensverwachting in goede gezondheid tussen de twee groepen is zelfs nog veel groter: 16 tot 19 jaar (CBS, 2008). Omdat gezondheid noodzakelijk wordt geacht in ons dagelijks functioneren, kunnen deze ongelijkheden in gezondheid als oneerlijk en onrechtvaardig worden bestempeld (Rawls, 1971; Anand, 2002; Sen, 2002). Het terugdringen van ongelijkheden in gezondheid is derhalve een ethische verplichting (CSDH, 2008). Naast deze ethische en sociale redenen is het verbeteren van de volksgezondheid ook economisch gezien aantrekkelijk. Een betere volksgezondheid heeft positieve effecten op arbeidsparticipatie, arbeidsproductiviteit, opleidingsniveau en investeringen (Bloom en Canning, 2000). Gegeven deze sociale en economische motieven zal het geen verbazing oproepen dat het belangrijkste gezondheids-gerelateerde doel van regeringen wereldwijd is om ongelijkheden in gezondheid terug te dringen.

Om dit doel te verwezenlijken is het van vitaal belang te begrijpen waarom sociaaleconomische ongelijkheden in gezondheid bestaan. Dit is echter niet gemakkelijk. Is het zo dat mensen uit lagere sociaaleconomische klassen eerder last krijgen van gezondheidspro-

blemen in hun latere leven? Of is het zo dat gezondheidsproblemen tijdens de kinderjaren invloed hebben op het opleidingsniveau en arbeidsmarktuitkomsten? Het is duidelijk dat gezondheid en sociaaleconomische status elkaar beïnvloeden. Extra complicatie is dat er variabelen kunnen zijn die een invloed hebben op zowel gezondheid als sociaaleconomische status, zoals intelligentie of tijdsvoorkeur (Deary, 2008; Fuchs, 1982). Mede door deze complicaties en ondanks een enorme hoeveelheid literatuur op dit gebied is het begrip van sociaaleconomische verschillen in gezondheid beperkt. Tot op de dag van vandaag ontbreken er grofweg twee essentiële onderdelen van het vraagstuk: (i) robuuste, causale studies naar de belangrijkste relaties tussen gezondheid en sociaaleconomische status, en (ii) kennis van de mechanismen die deze belangrijkste relaties bepalen.

Het doel van dit proefschrift is door middel van economische analyses een beter inzicht te krijgen in deze twee missende onderdelen. De bijdrage van dit proefschrift is tweeledig. Ten eerste benutten we exogene variatie in zowel sociaaleconomische status als gezondheid om causale effecten te schatten van twee belangrijke paden: (i) het effect van opleiding op gezondheid en sterfte, en (ii) het effect van gezondheid op arbeidsparticipatie en inkomen. Door accuraat gebruik te maken van quasi-natuurlijke experimenten omzeilen we variabelen die mogelijk op zowel gezondheid als sociaaleconomische status van invloed zijn. Tweede bijdrage is de formulering van een theoretisch model van gezondheid en sociaaleconomische status over de levenscyclus. Dit model fungeert als een kader waarbinnen we de mechanismen die leiden tot de grote correlatie tussen gezondheid en sociaaleconomische status beter kunnen begrijpen.

Het proefschrift is als volgt opgebouwd. Het eerste deel — hoofdstukken 2 en 3 — omvat een beschrijvende analyse van sociaaleconomische verschillen in gezondheid over de levenscyclus. Deze analyse geeft een idee hoe groot de verschillen zijn en welke relaties nader onderzocht dienen te worden. Uit deze hoofdstukken kunnen echter geen harde beleidsconclusies getrokken worden; een beschrijvende analyse zegt niets over de richting van causaliteit en variabelen als tijdsvoorkeur en intelligentie kunnen de geobserveerde relaties beïnvloeden. Het tweede deel van het proefschrift — hoofdstukken 4 en 5 — benut exogene variatie in gezondheid en sociaaleconomische status om de causale effecten van opleiding op sterfte (hoofdstuk 4) en gezondheid op arbeidsparticipatie en inkomen (hoofdstuk 5) te schatten. Het laatste deel van het proefschrift — hoofdstuk 6 — omvat een theoretisch model om de verschillende mechanismen die leiden van sociaaleconomische status naar gezondheid en vice versa beter te begrijpen. Hoofdstuk 7 concludeert.

Sociaaleconomische verschillen in gezondheid

Hoofdstuk 2 beschrijft hoe gezondheid en ongelijkheid in gezondheid zich over de levenscyclus ontwikkelen in 11 Europese landen. Vanaf een bepaalde leeftijd is een daling van gezondheid onvermijdelijk, maar vanaf welke leeftijd is dit gemiddeld genomen? En hoe snel worden mensen ongezonder? De antwoorden op deze vragen hebben belangrijke implicaties, bijvoorbeeld voor de recente discussie omtrent de ophoging van de pensioengerechtigde leeftijd. Het is verder duidelijk dat er grote verschillen bestaan in gezondheid, maar op welke leeftijd zijn deze verschillen het grootst? Worden de verschillen groter naarmate mensen ouder worden — een proces van *cumulative advantage* — of verdwijnen verschillen in gezondheid juist wanneer mensen ouder worden als de biologische component belangrijker wordt in het gezondheidsproces — *age-as-leveller*?

Gezondheid en ongelijkheid in gezondheid evolueren echter niet alleen over de levenscyclus, maar kunnen ook verschillen over generaties. Technologische vooruitgang en het algemene economische klimaat kunnen hun weerslag hebben op (ongelijkheid in) gezondheid in de populatie. Deze zogenaamde cohort effecten *verstoren* het levenscyclus patroon, maar zijn zelf ook bijzonder interessant, bijvoorbeeld voor het voorspellen van de gevolgen van vergrijzing. Zijn jongere generaties gezonder dan oudere generaties? En is ongelijkheid in gezondheid toe- of afgenomen?

De resultaten laten zien dat zelf-gerapporteerde gezondheid gemiddeld genomen maar licht daalt tot een jaar of 70, waarna een snelle daling inzet. Dit patroon is verrassend consistent over de 11 Europese landen. Dit suggereert dat gezondheid geen significante beperking zou opleveren voor het ophogen van de pensioengerechtigde leeftijd tot een jaar of 70, maar wel een grote beperking kan zijn na deze leeftijd. Opvallend is dat gezondheid in de Zuid-Europese landen en Ierland sterk is verbeterd over generaties, terwijl in de Noord-Europese landen gezondheid niet is verbeterd. Sociaaleconomische ongelijkheid in gezondheid volgt een *omgekeerde U-vorm*, waarbij de verschillen het grootst zijn net voor de pensioenleeftijd. Dit suggereert een grote rol voor arbeidsparticipatie in het genereren van ongelijkheden in gezondheid. Tenslotte vinden we dat sociaaleconomische ongelijkheid in gezondheid niet is afgenomen over de laatste generaties. Ondanks inspanningen van overheden is het niet gelukt deze verschillen in gezondheid te verkleinen.

Ingegeven door de hardnekkige ongelijkheden in gezondheid probeert hoofdstuk 3 deze wat verder te ontrafelen voor Nederland. Meer specifiek probeert hoofdstuk 3 te onderzoeken hoe gezondheid evolueert over de levenscyclus en hoe dit verschilt voor verschillende sociaaleconomische groepen. Het blijkt dat de patronen die we vinden voor Nederland opvallend veel gelijkenissen vertonen met de patronen in de Verenigde Staten (Smith,

2005a; Case en Deaton, 2005). Dit duidt erop dat bepaalde fundamentele relaties tussen gezondheid, opleiding, werk en inkomen niet erg gevoelig zijn voor de inrichting van de gezondheidszorg sector en andere beleidsinstrumenten. De verschillen zijn ook zeker niet verwaarloosbaar: rond de leeftijd van 55 beschouwt 30 % van de mannen in de laagste inkomensgroepen zijn gezondheid als slecht, tegenover slechts 5 % van de mannen in de hoogste inkomensklassen.

Onze interpretatie van de resultaten is dat arbeidsparticipatie een sleutelrol vervult in het genereren van sociaaleconomische ongelijkheden in gezondheid. Als we ons beperken tot de populatie die werkt, verdwijnen ongelijkheden in gezondheid vrijwel in zijn geheel. De grootste verschillen in gezondheid worden gevonden tussen werkende en niet-werkende mensen, wat duidt op een sterk effect van gezondheid op arbeidsparticipatie en uiteindelijk inkomen. Wel is er een effect van opleiding op gezondheid dat niet verdwijnt als er gecorrigeerd wordt voor arbeidsparticipatie. Ten slotte vinden we — net als in hoofdstuk 2 — dat ongelijkheden in gezondheid pieken rond een jaar of 55-60 waarna de verschillen kleiner lijken te worden. Echter, uit onze analyse blijkt dat er ook grote verschillen zijn in sterfte tussen sociaaleconomische klassen rond die leeftijd, zodat de schijnbare vernauwing van ongelijkheden in gezondheid tot op zekere hoogte een artefact is van deze verschillen in sterfte. Immers, alleen de robuuste, relatief gezonde groep uit de lagere sociaaleconomische klassen blijft over na een jaar of 60, zodat het lijkt alsof gezondheidsverschillen tussen groepen kleiner worden.

Causale effecten in de relatie tussen gezondheid en sociaaleconomische status

De beschrijvende analyse suggereert dat de twee belangrijkste paden in het genereren van sociaaleconomische verschillen in gezondheid zijn: 1) het effect van opleiding op gezondheid, en 2) het effect van gezondheid op arbeidsparticipatie en inkomen. Het tweede deel van dit proefschrift benut econometrische technieken om het bestaan van een causale component in deze relaties vast te stellen. Hoofdstuk 4 gebruikt de uitbreiding van de Leerplichtwet van 1928 — kinderen moesten niet langer 6 jaar maar 7 jaar verplicht naar school — als exogene variatie in opleiding om het causale effect te schatten op sterfte in de leeftijdsgroep van 80 tot 88. De aanname van dit zogenaamde *Regression Discontinuity Design* is dat mensen die net niet onder de nieuwe leerplichtwet vallen (geboren in 1916) in alle opzichten vergelijkbaar zijn met mensen die er net wél onder vallen (geboren in

1917). Alle verschillen in sterfte (bovenop het jaar leeftijdsverschil) kunnen nu worden toegeschreven aan het extra jaar opleiding.

De resultaten laten zien dat opleiding een significant effect heeft op sterfte. Een extra jaar opleiding verkleint de kans om te overlijden in de leeftijdscategorie 80 tot 88 met zo'n 4 tot 6 %. De sterke geobserveerde correlatie tussen opleiding en gezondheid komt dus (deels) voort uit een causaal effect van opleiding op gezondheid en sterfte. Onze resultaten impliceren dat de opbrengsten van opleiding verstrekkend zijn. Niet alleen zijn er monetaire opbrengsten in de vorm van een hoger loon (e.g. Card, 1999), maar ook neemt de levensverwachting significant toe, zelfs na 80 jaar. De andere kant van de medaille is dat lager opgeleiden niet alleen korter leven, maar ook minder lang van hun opgebouwde pensioen kunnen genieten.

Hoofdstuk 5 beoogt het causale effect van gezondheid op arbeidsparticipatie en inkomen te schatten. Uitgegaan wordt van een groep gezonde, werkende mensen in de leeftijd 17 tot 64 jaar. Voor deze groep mag ervan uitgegaan worden dat een acute opname in het ziekenhuis een exogene schok voor de gezondheid betekent. Door middel van *Propensity Score Matching* gecombineerd met *Difference-in-Difference* wordt gecorrigeerd voor het feit dat niet iedereen eenzelfde kans heeft op een acute opname. Gegeven een uitgebreide set van achtergrondkenmerken is de kans op een acute ziekenhuisopname random, waarbij een klein deel daadwerkelijk een acute opname ondergaat (de treatment groep) en een groot deel niet (de controle groep). Door mensen in de treatment groep over de tijd te vergelijken met mensen met vergelijkbare kenmerken in de controle groep kan het causale effect van een gezondheidsschok bepaald worden op arbeidsparticipatie en inkomen.

Niet verwonderlijk leidt een gezondheidsschok tot een daling van de arbeidsparticipatie. Aangezien de vervangingsratio's van sociale uitkeringen kleiner zijn dan 100 % is het ook niet opmerkelijk als inkomen uit arbeid lager zou worden. Wellicht interessanter is bekijken of het totale beschikbare inkomen van een persoon en zijn huishouden verandert na alle formele en informele mechanismen van compensatie. Kan de echtgenoot van de arbeidsongeschikte het inkomensverlies compenseren, of is het verlies aan huishoudensinkomen nog groter omdat de echtgenoot (en andere leden van het huishouden) ook hun arbeidsparticipatie moeten verminderen om zorg te dragen voor de arbeidsongeschikte? Ook is het interessant om te bekijken welke groepen in de samenleving slechter zijn beschermd tegen inkomensverlies dan andere na een gezondheidsschok.

De resultaten bevestigen een sterk effect van gezondheidsschokken op zowel arbeidsparticipatie als beschikbaar inkomen. Beschikbaar inkomen daalt met zo'n 8 % drie jaar na de schok. Twee belangrijke conclusies kunnen worden getrokken uit de analyses: 1) arbeidsongeschiktheidsverzekeringen (WAO/WIA) vermijden grote verliezen van per-

soonlijk inkomen, maar kunnen niet voorkomen dat het verlies aan huishoudensinkomen groter is dan het verlies aan persoonlijk inkomen. Ook is het verschil tussen het verlies aan huishoudensinkomen en persoonlijk inkomen groter naarmate de gezondheidsschok zwaarder is. Dit duidt erop dat de overige leden van het huishouden — met name de echtgenoot — hun arbeidsparticipatie verminderen om zorg te dragen voor het arbeidsongeschikte huishoudenslid. Echter, de precieze redenen waarom het verlies aan huishoudensinkomen groter is dan persoonlijk inkomen is niet duidelijk en dient verder onderzocht te worden. De tweede belangrijke conclusie is dat het inkomensverlies na een gezondheidsschok niet bepaald evenredig is verdeeld over de samenleving. Mensen in de lagere sociaaleconomische klassen — de laagste inkomensgroepen en de lager opgeleiden — zijn niet alleen kwetsbaarder om een schok te krijgen in hun gezondheid, maar zijn daarnaast ook nog eens slechter beschermd tegen inkomensverlies als gevolg van zo'n schok.

Onderliggende mechanismen in de relatie tussen gezondheid en sociaaleconomische status

Ondanks het belang van het blootleggen van de belangrijkste causale relaties in de voorgaande hoofdstukken zijn de mechanismen die leiden van sociaaleconomische status naar gezondheid en vice versa grotendeels onduidelijk. Bijvoorbeeld, het is nu bekend dat opleiding een causaal effect heeft op sterfte, maar op welke manier een hogere opleiding wordt vertaald in een hogere levensverwachting is onbekend. Zonder kennis van deze mechanismen is het zeer moeilijk om gedegen beleidsadviezen te geven (Deaton, 2002). Case en Deaton (2005) menen dat het extreem moeilijk is om alle relaties en mechanismen goed te begrijpen zonder een theoretisch kader. Het standaard economische kader in de gezondheidseconomie — het Grossman model (Grossman, 1972a; 1972b) — is bijzonder nuttig in het analyseren van gezondheid en de vraag naar medische zorg, maar schiet tekort in het verklaren van sociaaleconomische verschillen in gezondheid.

Het Grossman model neemt aan dat mensen nut (*utility*) halen uit gezondheid en uit consumptie. Gezondheid wordt geacht elk jaar af te nemen met een bepaalde exogene factor, maar deze daling van gezondheid kan een halt worden toegeroepen door te investeren in medische zorg. Er is echter een grens aan deze uitgaven (de *budget constraint*). De totale uitgaven aan medische zorg en consumptie kunnen niet groter zijn dan het initiële vermogen plus het inkomen dat wordt verdiend (dat ook afhangt van de gezondheid van de persoon). In het kort geeft het model dus aan dat gezondheid zowel directe voldoening

geeft (*consumption benefits*) alsook mensen in staat stelt om te werken en inkomen te verdienen (*production benefits*). Mensen optimaliseren hun nutsfunctie gegeven de restricties in hun inkomen en gezondheid.

Het model dat wordt gepresenteerd in hoofdstuk 6 breidt het standaard Grossman model uit om sociaaleconomische verschillen in gezondheid beter te verklaren. Ten eerste nemen we aan dat investeringen in medische zorg afnemende meeropbrengsten vertonen in tegenstelling tot de constante meeropbrengsten van het standaard model. Het idee hierachter is dat het eerste medicijn dat geslikt wordt van een bepaalde soort waarschijnlijk meer effect heeft dan de daaropvolgende. Tweede uitbreiding is dat ons model mensen in staat stelt om een bepaald beroep te kiezen met hogere gezondheidsrisico's (zwaar lichamelijk werk, monotoon werk) tegen een hoger loon. Mensen met een lagere opleiding hebben vaak weinig andere keus dan zwaar lichamelijk werk te accepteren tegen een bepaald loon, terwijl hoger opgeleiden niet genoodzaakt zijn hun lichaam hiervoor in te zetten. Derde belangrijke uitbreiding is te onderkennen dat de prijs van consumptie niet alleen monetair is, maar dat mensen ook de gezondheidsrisico's in acht nemen van bepaalde consumptiegoederen. Bepaalde soorten drugs geven bijvoorbeeld een bepaald nut aan sommige mensen en de kostprijs is niet extreem hoog. Toch zullen de meeste mensen niet dagelijks deze vorm van nutsvoorziening gebruiken omdat de prijs in termen van gezondheid (potentieel) zeer hoog kan zijn. Ten slotte introduceren we ook preventieve zorg en een realistische relatie tussen opleiding en loon in het model.

Het model voorspelt dat mensen met meer vermogen, hogere inkomens en een hoger opleidingsniveau meer investeren in medische en preventieve zorg, ongezonde consumptie meer en meer inwisselen voor gezonde consumptie en een gezondere werkomgeving kiezen. Deze voorspellingen komen voort uit de hogere marginale kosten van medische zorg voor de hogere sociaaleconomische klassen, zodat de hogere sociaaleconomische klassen meer profiteren van een gezonde leefstijl en werkomgeving. De kosten van een ongezonde leefstijl zijn groter voor deze mensen en ze hebben meer te verliezen bij ziekte (door hogere marginale kosten van medische zorg en hogere lonen). Een grotere consumptie van medische zorg en een gezondere leefstijl en werkomgeving leiden tot een cumulatief gezondheidsvoordeel over de levenscyclus, in lijn met de empirische studies in voorgaande hoofdstukken en de literatuur.

Discussie en conclusie

Er zijn grote gezondheidsverschillen naar sociaaleconomische status. Mensen uit lagere sociaaleconomische klassen beginnen vaak al met een gezondheidsachterstand (in de kinder-

jaren ofwel bij meerderjarigheid) en deze verschillen worden alleen maar groter met de jaren. Ondanks de aandacht en verschillende beleidsmaatregelen zijn sociaaleconomische verschillen in gezondheid niet kleiner geworden over de laatste generaties. Dit proefschrift en de overige literatuur bewijzen dat er twee causale paden extreem belangrijk zijn in het genereren van deze ongelijkheden. Ten eerste, het effect van opleiding op gezondheid. Dit proefschrift bevestigt dat opleiding niet alleen een belangrijke associatie geniet met gezondheid, maar dat opleiding ook een causaal effect heeft op gezondheid en sterfte. Voor Nederlandse mannen die de leeftijd van 80 jaar bereiken betekent een jaar extra onderwijs een daling van de kans op overlijden tussen het 80e en 88e levensjaar van zo'n 4 tot 6 %.

Ten tweede is er een belangrijk en groot effect van gezondheid op arbeidsparticipatie en inkomen. Een schok in gezondheid leidt tot een permanente daling van arbeidsparticipatie en beschikbaar inkomen. Er treedt niet alleen een verlies op van persoonlijk inkomen, maar ook huishoudensinkomen ondergaat een gevoelig verlies, zelfs groter dan dat van persoonlijk inkomen. Waarschijnlijk moeten andere leden van het huishouden hun arbeidsparticipatie ook verminderen om zorg te dragen voor het zieke lid van het huishouden. Ondanks de grote verschillen in gezondheid naar inkomen, is de consensus onder economen aan het groeien dat inkomen zelf geen onafhankelijk effect heeft op gezondheid zodra gecontroleerd wordt voor andere relevante variabelen.

Het theoretisch kader ontwikkeld in dit proefschrift biedt het inzicht dat de belangrijkste reden waarom de hogere sociaaleconomische klassen een betere gezondheid hebben is omdat deze groep hogere marginale kosten van medische zorg ervaart. Dit leidt tot een gezondere leefstijl en een minder risicovolle werkomgeving. De lagere sociaaleconomische klassen, echter, hebben geen of weinig vermogen en weinig opleiding genoten zodat voor hen het een *optimale, doch hevig gerestricteerde* keuze is om ongezond werk te accepteren en een ongezondere leefstijl aan te nemen. Over de levenscyclus leidt dit tot steeds groter wordende verschillen in gezondheid.

Om dit proces van cumulatieve nadelen voor de lagere sociaaleconomische klassen tegen te gaan is een enorme beleidsuitdaging. De potentiële voordelen kunnen echter gigantisch zijn. Als de relatie tussen een lage opleiding, slechte gezondheid en vroegtijdige uittreding uit het arbeidsproces kan worden gelimiteerd of zelfs doorbroken, kan dit leiden tot een enorme stijging van economische groei en maatschappelijk welzijn. Gegeven de analyses in dit proefschrift zou een beleidsstrategie kunnen zijn om onderwijs te stimuleren. Echter, voor een succesvolle strategie is het van belang om ook de onderliggende mechanismen die leiden van een betere opleiding naar een betere gezondheid verder uit te diepen. Een andere strategie is het verbeteren van de gezondheidscondities op de

werkvloer en dan met name het proberen te voorkomen van uittreding als gevolg van gezondheidsschokken. In die zin lijken recente wijzigingen in de arbeidsongeschiktheidwetgeving in de goede richting te gaan.

Echter, de invloed van beleid is maar zeer beperkt wanneer de meeste gezondheidsverschillen al in de kinderjaren (of nog eerder) ontstaan. Ook als andere variabelen zoals tijdsvoorkeur en intelligentie een grote rol spelen is er weinig dat beleid kan doen. Een beter geïnformeerd beleid is gebaat bij verdiepend onderzoek naar de mechanismen die leiden van gezondheid naar sociaaleconomische status en vice versa. Dit proefschrift heeft bijgedragen aan het beter begrijpen van verschillen en een eerste aanzet gegeven tot een theoretisch kader, maar uitbreidingen hierop zijn dringend gewenst om het inzicht te vergroten in gezondheidsverschillen. Dit lijkt de enige manier waarop gezondheidsverschillen naar sociaaleconomische status kunnen worden teruggebracht naar een acceptabel niveau, en de volksgezondheid en welzijn kunnen worden verbeterd.

References

Aarts, L., R.V. Burkhauser, and P.R. de Jong (1996), *Curing the Dutch Disease: an International Perspective on Disability Policy Reform*, Aldershot, U.K: Avebury.

Adams, P., M.D. Hurd, D.M. McFadden, A. Merrill, and T. Ribeiro (2003), "Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status", *Journal of Econometrics*, 112(1):3-56.

Adler N.E., T. Boyce, M.A. Chesney, S. Cohen, S. Folkman, R.L. Kahn, and S.L. Syme (1994), "Socioeconomic status and health. The challenge of the gradient" , *The American Psychologist,* 49: 15-24.

Adler N.E., T. Boyce, M.A. Chesney, S. Folkman, and S.L. Syme (1993), "Socioeconomic inequalities in health: No easy solution", *Journal of the American Medical Association*, 269: 3140-3145.

Adler, N.E. and J.M. Ostrove (1999), "Socioeconomic Status and Health: What We Know and What We Don't" , *Annals New York Academy of Sciences*, 896: 3-15.

Albouy, V. and L. Lequien (2008), "Does compulsory education lower mortality?", *Journal of Health Economics*, 28(1): 155-168.

Anand, S. (2002), "The concern for equity in health", *Journal of Epidemiology and Community Health*, 56: 485-487.

Angrist, J.D. and A.B. Krueger (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples", *Journal of the American Statistical Association*, 87(418):328-336.

Antonovsky, A. (1967), "Social class, life expectancy and overall mortality", *Milbank Quarterly*, 45(2): 31-73.

Arellano, M. and C. Meghir (1992), "Female Labour Supply and On-the-Job Search: An Empirical Model Estimated Using Complementary Data Sets", *Review of Economic Studies*, 59(3): 537-559.

Arendt J.N. (2005), "Does education cause better health? A panel data analysis using school reforms for identification", *Economics of Education Review*, 24(2): 149-160.

Arendt J.N. (2008), "In sickness and in health till education do us part: Education effects on hospitalization", *Economics of Education Review*, 27(2): 161-172.

Arkes, J. (2003), "Does Schooling Improve Health?", Working paper, RAND Corporation DRU-3051.

Atchley R.C. (1976), *The sociology of retirement*, Halstead Press: New York.

Attanasio, O. and S.J. Davis (1996), "Relative Wage Movements and the Distribution of Consumption", *Journal of Political Economy*, 104(6): 1227-1262.

Au, D., W. Han, T.F. Crossley, and M. Schellhorn (2005), "The Effect of Health Changes and Long-Term Health on the Work Activity of Older Canadians", *Health Economics*, 14(10): 999-1018.

Auld, M.C. and N. Sidhu (2005), "Schooling, cognitive ability, and health", *Health Economics*, 14(10): 1019-1034.

Avendano, M., H.C. Boshuizen, F.G. Schellevis, J.P. Mackenbach, F.J. van Lenthe, and G.A. van den Bos, (2006a), "Disparities in stroke preventive care in general practice did not explain socioeconomic disparities in stroke", *Journal of Clinical Epidemiology*, 59(12): 1285-1294.

Avendano, M., I. Kawachi, F. van Lenthe, H.C. Boshuizen, J.P. Mackenbach, G.A. van den Bos, et al. (2006b), "Socioeconomic status and stroke incidence in the US elderly:

the role of risk factors in the EPESE study", *Stroke*, 37(6): 1368-1373.

Bago d'Uva, T., O. O'Donnell, and E. van Doorslaer (2008), "Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans", *International Journal of Epidemiology*, 37(6): 1375-1383.

Balia, S. and A.W. Jones (2008), "Mortality, lifestyle and socio-economic status", *Journal of Health Economics*, 27(1): 1-26.

Banks, J., M. Marmot, Z. Oldfield, and J.P. Smith (2009), "The SES Health Gradient on Both Sides of the Atlantic" in: David A. Wise (ed.), Developments in the Economics of Aging, The University of Chicago Press, Chicago, 359-420.

Bardasi, E., S.P. Jenkins, H. Sutherland, H. Levy, and F. Zantomio (2007), *British Household Panel Survey Derived Current and Annual Net Household Income Variables, Waves 1-14, 1991-2005* [computer file]. *7th Edition,* University of Essex. Institute for Social and Economic Research, [original data producer(s)]. Colchester, Essex: UK Data Archive [distributor], SN: 3909.

Barker, D.J.P. (1995), "Fetal origins of coronary heart disease", *British Medical Journal*, 311(6998): 171-174.

Barker, D.J.P., P.D. Gluckman, K.M. Godgrey, J.E. Harding, J.A. Owens, J.S. Robinson (1993), "Fetal nutrition and cardiovascular disease in adult life", *Lancet*, 341(8850): 938-941.

Barro, R.J. (1997), *Determinants of economic growth: a cross-country empirical study.* Cambridge, MA: MIT Press.

Barsky, R.B., F.T. Juster, M.S. Kimball and M.D. Shapiro (1997), "Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement study", *Quarterly Journal of Economics*, 112(2): 537579.

Baum, C.F., M.E. Schaffer, and S. Stillman (2007), "Enhanced routines for instrumental variables/GMM estimation and testing", Boston College Economics Working Paper no.

667.

Becker, G. (1964), "Human Capital and the Goal Distribution of Income: An Analytical Approach", New York: Columbia University Press.

Beckett, M. (2000), "Converging health inequalities in later life: An artifact of mortality selection?", *Journal of Health and Social Behavior*, 41:106-19.

Behrman, J.R. and M.R. Rosenzweig (2004), "Returns to Birthweight", *Review of Economics and Statistics*, 86(2): 586-601.

Berger, M.C. (1983), "Labour supply and spouse's health: the effects of illness, disability and mortality", *Social Science Quarterly*, 64: 494-509.

Berger, M.C. and B.M. Fleisher (1984), "Husband's Health and Wife's Labor Supply", *Journal of Health Economics*, 3: 63-75.

Blanchflower, D.G., A.J. Oswald, B. van Landeghem (2009), "Imitative Obesity and Relative Utility", *Journal of the European Economic Association*, 7: 528-538.

Blau, D.M. (1998), "Labor Force Dynamics of Older Married Couples", *Journal of Labor Economics*, 16(3): 595-629.

Blau, D.M. and R.T. Riphahn (1999), "Labor force transitions of older married couples in Germany", *Labour Economics*, 6(2): 229-252.

Bloom, D.E. and D. Canning (2000), "The health and wealth of nations", *Science*, 287: 1207-1209.

Blundell, R. and M. Costa Dias (2002), "Alternative approaches to evaluation in empirical microeconomics", *Portuguese Economic Journal*, 1: 91-115.

Blundell, R., L. Pistaferri, and I. Preston (2008), "Consumption Inequality and Partial Insurance", *American Economic Review*, 98(5): 1887-1921.

Bolin, K., L. Jacobson, and B. Lindgren (2001), "The family as the producer of health - When spouses are Nash bargainers", *Journal of Health Economics*, 20: 349-362.

Bolin, K., L. Jacobson, and B. Lindgren (2002a), "Employer investments in employee health - implications for the family as health producer", *Journal of Health Economics*, 21: 563-583.

Bolin, K., L. Jacobson, and B. Lindgren (2002b), "The family as the health producer - When spouses act strategically", *Journal of Health Economics*, 21: 475-495.

Bolin, K., B. Lindgren, M. Lindstrom, and P. Nystedt (2003), "Investments in social capital - implications of social interactions for the production of health", *Social Science and Medicine*, 56(12): 2379-2390.

Borg, V. and T.S. Kristensen (2000), "Social class and self-rated health: can the gradient be explained by differences in life style or work environment?", *Social Science and Medicine*, 51: 1019-1030.

Bound, J. (1991), "Self reported versus objective measures of health in retirement models", *Journal of Human Resources*, 26: 107-137.

Bound, J., and R.V. Burkhauser (1999), "Economic Analysis of Transfer Programs Targeted on People with Disabilities", in Handbook of Labor Economics, Volume 3(c). O. Ashenfelter and D. Card, eds. Amsterdam: Elsevier Science, 3417-3528.

Bound, J., D.A. Jaeger, and R.M. Baker (1995), "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak", *Journal of the American Statistical Association*, 90(430): 443-450.

Bound, J., M. Schoenbaum, T.R. Stinebrickner, and T. Waidmann (1999), "The Dynamic Effects of Health on the Labor Force Transitions of Older Workers", *Labour Economics*, 6(2):179-202.

Bovenberg, L., J.P. Mackenbach, and R. Mehlkopf (2006), "Een eerlijk en vergrijzingbestendig ouderdomspensioen", *Economisch Statistische Berichten*, 91(4500): 648-651.

Bozzoli C., A. Deaton, and C. Quintana-Domeque (2007), "Child mortality, income and adult height", NBER Working Paper No. 12966. NBER: New York.

Burkhauser, R.V., M.C. Daly, and P.R. de Jong. (2008), "Curing the Dutch Disease: Lessons for United States Disability Policy" University of Michigan Retirement Research Center Working Paper WP 2008-188.

Burström, K, M. Johannesson, and F. Diderichsen (2005), "Increasing socio-economic inequalities in life expectancy and QALYs in Sweden 1980-1997", *Health Economics*, 14: 831-850.

Card, D.E. (1999), "The Causal Effect of Education on Earnings" in Orley Ashenfelter and David E. Card, eds., The Handbook of Labor Economics, Volume 3A, Amsterdam: Elsevier.

Case A., A. Fertig, and C. Paxson (2005), "The lasting impact of childhood health and circumstance", *Journal of Health Economics*, 24(2): 365-389.

Case, A. and A. Deaton (2005), "Broken Down by Work and Sex: How our Health Declines", in: David A. Wise (ed.), Analyses in the Economics of Aging, The University of Chicago Press, Chicago, 185-212.

Case, A., D. Lubotsky, and C. Paxson (2002), "Economic status and health in childhood: The origins of the gradient", *American Economic Review*, 92: 1308-1334.

CBS (1931), "Statistiek van het gewoon en uitgebreid lager onderwijs", Centraal Bureau voor de Statistiek. 's-Gravenhage : Algemeene Landsdrukkerij.

CBS (1951), "De ontwikkeling van het onderwijs in Nederland", Centraal Bureau voor de Statistiek. Uitgeversmaatschappij W. de Haan N.V. Utrecht.

CBS (2008), "Hoogopgeleiden leven lang en gezond" in: Gezondheid en zorg in cijfers, Centraal Bureau voor de Statistiek, Den Haag.

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

Charles, K.K. (1999), "Sickness in the family: health shocks and spousal labor supply", Working paper no. 2000-011, University of Michigan.

Charles, K.K. (2002), "Is retirement depressing? Labor force inactivity and psychological well-being in later life", NBER Working Paper No. 9033, Cambridge MA.

Charles, K.K. (2003), "The Longitudinal Structure of Earnings Losses among Work-Limited Disabled Workers", *Journal of Human Resources*, 38(3):618-646.

Chevalier, A. and L. Feinstein (2007), "Sheepskin or Prozac: The Causal Effect of Education on Mental Health" UCD Geary Institute Discussion Paper Series June 2007.

Clark, D. and H. Royer (2007), "The Effect of Education on Adult Mortality and Health: Evidence from the United Kingdom" Mimeo, July.

Cochrane, J.H. (1991), "A Simple Test of Consumption Insurance", *Journal of Political Economy*, 99(5): 957-976.

Cochrane, J.H. (1995), "Time-Consistent Health Insurance", *Journal of Political Economy*, 103: 445-473.

Contoyannis, P. and N. Rice (2001), "The impact of health on wages: evidence from the British Household Panel Survey", *Empirical Economics*, 26: 599-622.

Contoyannis, P., A.M. Jones, and N. Rice (2004), "The dynamics of health in the British Household Panel Survey", *Journal of Applied Econometrics*, 19(4): 473-503.

Contoyannis, P. and A.M. Jones (2004), "Socio-economic status, health and lifestyle", *Journal of Health Economics*, 23(5): 965-995.

Cowell, A.J. (2006), "The relationship between education and health behavior: some empirical evidence", *Health Economics*, 15(2): 125-146

Crimmins, E. (2004), "Trends in the health of the elderly", *Annual Review of Public Health*, 25: 79-98.

Cropper, M.L. (1977), "Health, investment in health, and occupational choice", *Journal of Political Economy*, 85: 1273-1294.

Cropper, M.L. (1981), "Measuring the Benefits from Reduced Morbidity", *American Economic Review*, 71(2): 235-240.

CSDH (2008), "Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health", Final Report of the Commission on Social Determinants of Health, Geneva, World Health Organization.

Cunha, F. and J. Heckman (2007), "The Technology of Skill Formation", *American Economic Review*, 97(2): 3147.

Currie, A., M. Shields, and W. Price (2007), "The Child Health/Family Income Gradient: Evidence from England", *Journal of Health Economics*, 26:213-232.

Currie, J. and B.C. Madrian (1999), "Health, Health Insurance and the labour market", In: Ashenfelter O, Card, D. (eds.) Handbook of labour economics 3: 3309-3415, Amsterdam: Elsevier Science B.V.

Currie, J. and M. Stabile (2003), "Socioeconomic status and child health: why is the relationship stronger for older children?", *American Economic Review*, 93: 1813-1823.

Cutler, D.M., A. Lleras-Muney, and T. Vogl (2008), "Socioeconomic Status and Health: Dimensions and Mechanisms", NBER Working Paper, 14333. Cambridge, NBER.

Cutler, D.M. and A. Lleras-Muney (2008), "Education and Health: Evaluating Theories and Evidence", in Robert F. Schoeni et al., eds., Making Americans Healthier: Social and Economic Policy As Health Policy (National Poverty Center Series on Poverty and Public Policy).

Dardanoni, V. and A. Wagstaff (1990), "Uncertainty and the demand for medical care", *Journal of Health Economics*, 9: 23-38.

De Bruin, A., J. Kardaun, F. Gast, E. de Bruin, M. van Sijl, and G. Verweij (2004), "Record linkage of hospital discharge register with population register: Experiences at Statistics Netherlands", *Statistical Journal of the United Nations*, ECE 21: 23-32

De Graaf, J.H. (2000), *Leerplicht en recht op onderwijs*, Ars Aequi Libri.

De Jong, P. (2008), "Recent changes in Dutch disability policy" APE Report 584, The Hague.

De Walque, D. (2007), "Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education", *Journal of Health Economics*, 26(5): 877-895.

Deary, I. (2008), "Why do intelligent people live longer?", *Nature*, 456: 175-176.

Deaton, A. (1997), *The analysis of household surveys: a microeconometric approach to development* policy, The Johns Hopkins University Press: Baltimore.

Deaton, A. (2002a), "Policy Implications of the Gradient of Health and Wealth", *Health Affairs*, 21(2): 13-30.

Deaton, A. (2002b), *Understanding Consumption*, Oxford: Oxford University Press.

Deaton, A. (2003), "Health, Inequality, and Economic Development", *Journal of Economic Literature*, 41: 113-158.

Deaton, A. (2008), "Income, aging, health and wellbeing around the world: Evidence from the Gallup World Poll", *Journal of Economic Perspectives*, 22(2): 5372.

Deaton, A and C.H. Paxson (1994), "Intertemporal choice and inequality", *Journal of Political Economy*, 102: 437-467.

Deaton, A. and C.H. Paxson, C.H. (1998), "Aging and inequality in income and health", *American Economic Review*, 88:248-53.

Deboosere, P. and K. Neels (2008), "The role of education in the health transition: Evidence from the Belgian census", Extended abstract, EAPS Conference Barcelona. Retrieved from http://epc2008.princeton.edu/download.aspx?submissionId=80397.

Dehejia, R.H. and S. Wahba (1999), "Causal effects in non-experimental studies: Reevaluating the evaluation of training programs", *Journal of the American Statistical Association*, 94: 1053-62.

Dehejia, R.H, and S. Wahba (2002), "Propensity score-matching methods for nonexperimental causal studies", *Review of Economics and Statistics*, 84: 151-161.

Descartes, R. (1637), "Discours de la Méthode, Sixième Partie", In: Bridoux, A., edt. Descartes: Oeuvres et lettres. Paris : Gallimard (Bibliothèque de la Pléiade), 1953:168.

Devereux, P.J. and R.A. Hart (2008), "Forced to be Rich? Returns to Compulsory Schooling in Britain", IZA Working Paper 3305.

Diamond, P. (1992), "Organizing the Health Insurance Market", *Econometrica*, 60: 1233-1254.

Disney, R., C. Emmerson, and M. Wakefield (2006), "Ill Health and Retirement in Britain: A Panel Data-Based Analysis", *Journal of Health Economics*, 25 (4):621-649.

Dodde, N. (2000), "Een geschiedenis van de leerplicht" In: Recht doen aan zorg. 100 jaar Leerplicht in Nederland. Ton van der Hulst & Dolf van Veen. Leuven/Apeldoorn.

Duncan, G.J. and B. Holmlund (1983), "Was Adam Smith Right After All? Another Test of the Theory of Compensating Wage Differentials", *Journal of Labor Economics*, 1(4): 366-379.

Dustmann, C. and F. Windmeijer (2000), "Wages and the Demand for Health - A Life Cycle Analysis", IZA Discussion Paper 171, Germany.

Ehrlich, I. and H. Chuma (1990), "A Model of the Demand for Longevity and the Value of Life Extension", *Journal of Political Economy*, 98(4): 761-782.

Elo, I.T. and S.H. Preston (1996), "Educational differentials in mortality: United States, 1979-85", *Social Science and Medicine*, 42:47-57.

Erreygers, G. (2009a), "Correcting the Concentration Index", *Journal of Health Economics*, 28(2): 504-515.

Erreygers, G. (2009b), "Can a single indicator measure both attainment and shortfall inequality?", *Journal of Health Economics*, 28(4): 885-893.

EUROSTAT (2003), "Construction of weights in the ECHP", Eurostat: Luxembourg.

Fahey, T., B. Nolan, and B. Matre (2004), "Housing Expenditures and Income Poverty in EU Countries", *Journal of Social Policy*, 33: 437-454.

Feeny D., W. Furlong, G.W. Torrance, C.H. Goldsmith, Z. Zhu, S. DePauw, M. Denton, and M. Boyle (2002), "Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System", *Medical Care*, 40(2), 113-128.

Feinstein, L., R. Sabates, T.M. Anderson, A. Sorhaindo, and C. Hammond (2006), *The Effects of Education on Health: Concepts, evidence and policy implications*, Paris: Organisation for Economic Co-operation and Development (OECD).

Ferrie, J.E., M.J. Shipley, G.D. Smith, S.A. Stansfeld, and M. Marmot (2002), "Change in health inequalities among British civil servants: the Whitehall II study", *Journal of Epidemiology and Community Health*, 56: 922-926.

Freedman, V., R. Schoeni, L. Martin, and J. Cornman (2007), "Chronic conditions and the decline in late-life disability", *Demography*, 44(3): 459-477.

Freeman, R. (2007), "Labor Market Institutions Around the World", NBER Working Paper W13242.

Frijters, P., J.P. Haisken-DeNew, and M.A. Shields (2005), "The causal effect of income on health: Evidence from German reunification", *Journal of Health Economics*, 24(5): 997-1017.

Fuchs, V.R. (1982), "Time Preferences and Health: An Exploratory Study", in Economic Aspects of Health. Victor Fuchs, ed. Chicago: U. Chicago Press, 93-120.

Fuchs, V.R. (1986), *The Health Economy*, first ed. Harvard University Press, Cambridge, MA.

Furlong, W.J., D.H. Feeny, G.W.Torrance, and R.D. Barr (2001), "The Health Utilities Index (HUI) System for Assessing Health-Related Quality of Life in Clinical Studies", *Annals of Medicine*, 33(5): 375-384.

Galama, T.J. (2010), "Grossman's Health Capital Model Revived", Mimeo.

Galama, T.J. and A. Kapteyn (2009), "Grossman's missing health threshold", RAND Working Paper, WR-684.

Galama, T.J., A. Kapteyn, R. Fonseca, and P.C. Michaud (2009), "Grossman's Health Threshold and Retirement", RAND Working Paper, WR-658.

Garcia, J. and C. Quintana-Domeque (2007), "The evolution of adult height in Europe: A Brief Note", *Economics and Human Biology*, 5(2): 340-349.

García-Gómez, P. and A. López-Nicolás (2006), "Health Shocks, Employment and Income in the Spanish Labour Market", *Health Economics*, 15: 997-1009.

García-Gómez, P., H-M. von Gaudecker, and M. Lindeboom (2009), "Health, Disability and Work: Patterns for the Working Age Population", Netspar Panel Paper.

García-Gómez, P. (2008), "Institutions, Health Shocks and Labour Outcomes across Europe", Documento de Trabajo 2008-01 FEDEA.

Gerdtham, U.G. and M. Johannesson (2000), "Income-related inequality in life-years and quality-adjusted life-years", *Journal of Health Economics*, 19: 1007-1026.

Glied, S. and A. Lleras-Muney (2003), "Health Inequality, Education and Medical Innovation", NBER Working Paper No. 9738.

Goldman, D.P. and J.P. Smith (2002), "Can Patient Self-Management Help Explain the SES Health Gradient?", *Proceedings of the National Academy of Science*, 99: 10929-10934.

Goldman, D.P. and D. Lakdawalla (2001), "Understanding Health Disparities Across Education Groups" NBER Working Paper W8328.

Goodman, L.A. (1960), "On the Exact Variance of Products", *Journal of the American Statistical Association*, 55: 708-713.

Grimard, F. and D. Parent (2007), "Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking?", *Journal of Health Economics*, 26(5): 896-926.

Grossman, M. (1972a), *The Demand for Health-A Theoretical and Empirical Investigation*, New York: National Bureau of Economic Research.

Grossman, M. (1972b), "On the Concept of Health Capital and the Demand for Health", *Journal of Political Economy*, 80(2): 223-255.

Grossman, M. (2000), "The Human Capital Model", in Handbook of Health Economics, Volume 1, 347-408, A.J. Culyer and J.P. Newhouse (Eds.), Elsevier Science.

Grossman, M. (2003), "Education and nonmarket outcomes", In: E. Hanushek and F. Welch (eds.), Handbook of the Economics of Education. North-Holland, Elsevier Science,

Amsterdam.

Grossman, M. and R. Kaestner (1997), "Effects of Education on Health", in J. R. Berhman and N. Stacey (eds.) The Social Benefits of Education (Ann Arbor: University of Michigan Press).

Gudex, C. and G. Lafortune (2000), "An inventory of health and disability-related surveys in OECD countries", OECD Labor Market and Social Policy Occasional Papers, 44. Paris: OECD Publishing.

Hagan, R, A.M. Jones, and N. Rice (2008), "Health shocks and the hazard rate of early retirement in the ECHP", *Swiss Journal of Economics and Statistics*, 144: 323-335.

Hahn, J., P. Todd, and W. van der Klaauw (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design", *Econometrica*, 69(1): 201-209.

Hanlon, P., D. Walsh, and B. Whyte (2006), *Let Glasgow flourish*. Glasgow: Glasgow Centre for Population Health.

Haurin, D.R. (1989), "Women's Labor Market Reactions to Family Disruptions", *Review of Economics and Statistics*, 71(1): 54-61.

Heckman, J. (2000), "Policies to foster human capital", *Research in Economics*, 54: 356.

Heckman J., H. Ichimura, and P. Todd (1997), "Matching as an Econometric Evaluation Estimator", *Review of Economic Studies*, 64: 605-654.

Heckman, J. and B. Jacobs (2010), "Policies to Create and Destroy Human Capital in Europe", in: Edmund Phelps and Hans-Werner Sinn (eds) (2010), Perspectives on the Performance on the Continent's Economies, Cambridge-MA: MIT-Press.

Hentzen, C. (1928), "De financieele gelijkstelling 1920-1925", In: Politieke Geschiedenis Lager Onderwijs in Nederland, R.K. Centraal Bureau voor Onderwijs en Opvoeding te 's-Gravenhage

Hentzen, C. (1932), "De financieele gelijkstelling 1926-1929", In: Politieke Geschiedenis Lager Onderwijs in Nederland, R.K. Centraal Bureau voor Onderwijs en Opvoeding te 's-Gravenhage

Herd, P. (2006), "Do functional health inequalities decrease in old age?", *Research on Aging*, 28(3):375-392.

Herd, P., R.F. Schoeni, and J.S. House (2008), "Upstream solutions: does the supplemental security income program reduce disability in the elderly?", *Milbank Quarterly*, 86(1): 5-45.

Hernández-Quevedo, C, A.M. Jones, and N. Rice (2008), "Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey", *Cuadernos Económicos de ICE*, 75: 63-97.

Horsman, J., W. Furlong, D. Feeny, and G. Torrance (2003), "The Health Utilities Index (HUI): concepts, measurement properties and applications", *Health and Quality of Life Outcomes*, 1: 54.

House, J.S., K.R. Landis, and D. Umberson (1988), "Social relationships and health", *Science*, 241(4865): 540-545.

House, J.S., R.C. Kessler, and A. Regula Herzog (1990), "Age, Socioeconomic Status, and Health", *Milbank Quarterly*, 68(3): 383-411.

House, J.S., J.M. Lepkowski, A.M. Kinney, R.P. Mero, R.C. Kessler, and A.R. Herzog (1994), "The social stratification of aging and health", *Journal of Health and Social Behavior*, 35: 213-34.

HSG (1946/1947), *Handelingen der Staten-Generaal*, Bijlagen 1946-1947. 459.3: p.4.

HTK (1946/1947), *Handelingen Tweede Kamer*, 1946-1947. 459: p. 1968

Huisman, M., F. van Lenthe, M. Avendano, and J.P. Mackenbach (2008), "The contribution of job characteristics to socioeconomic inequalities in incidence of myocardial

infarction", *Social Science and Medicine*, 66(11): 2240-2252.

Hurd, M.D. (1990), "The joint retirement decision of husbands and wives", In: Issues in the Economics of Aging. David A. Wise edt. The University of Chicago Press.

Hurd, M.D. and A. Kapteyn (2003), "Health, Wealth, and the Role of Institutions", *Journal of Human Resources*, 38(2): 386415.

Idler, E.L. and Y. Benyamini (1997), "Self-rated health and mortality: a review of twenty-seven community studies", *Journal of Health and Social Behavior*, 38: 21-37.

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

Inoue, A. and G. Solon (2009), "Two-Sample Instrumental Variables Estimators", *Review of Economics and Statistics*, Forthcoming.

Jacobson, L. (2000), "The family as producer of health-an extension of the Grossman model", *Journal of Health Economics*, 19: 611-637.

Jiménez-Martín S., J.M. Labeaga, and M. Martínez Granado (1999), "Health status and retirement decisions for older European couples", CEPS working paper.

Jiménez-Martín, S., J.M. Labeaga, and C. Vilaplana Prieto (2006), "A sequential model of older workers' labor force transitions after a health shock", *Health Economics*, 15: 1033-1054.

Jones, A.M., X. Koolman, and N. Rice (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*, 169(3): 543-569.

Kamrul Islam, M., U.G. Gerdtham, P. Clarke, and K. Burström (2009), "Does income-related health inequality change as the population ages? Evidence from Swedish panel data", *Health Economics*, 10.1002/hec.1479.

Kapteyn, A., J.P. Smith, and A. van Soest (2007), "Vignettes and self-reports of work disability in the United States and the Netherlands", *American Economic Review*, 97(1): 461-473.

Kawachi, I., and L. Berkman (2003), *Neighborhoods and health*, Oxford, U.K.: Oxford University Press. I. Kawachi and L. Berkman Eds.

Khang, Y.H., J.W. Lynch, S. Yang, S. Harper, S.C. Yun, K. Jung-Choi, et al. (2009), "The contribution of material, psychosocial, and behavioral factors in explaining educational and occupational mortality inequalities in a nationally representative sample of South Koreans: relative and absolute perspectives", *Social Science and Medicine*, 68(5): 858-866.

Kim, J. and K. Durden (2007), "Socioeconomic status and age trajectories of health", *Social Science and Medicine*, 65: 2489-2502.

Kitagawa, E.M. and P.M. Hauser (1973), *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*, Cambridge: Harvard University Press.

Kunst, A., V. Bos, E. Lahelma, M. Bartley, I. Lissau, E. Regidor, A. Mielck, et al. (2005), "Trends in socioeconomic inequalities in self-assessed health in 10 European countries", *International Journal of Epidemiology*, 34(2): 295-305.

Kunst, A. and J.P. Mackenbach (1994), "Mortality differences associated with educational level in nine industrialized countries", *American Journal of Public Health*, 84: 932-37.

Lafortune, G. and G. Balestat (2007), "Trends in Severe Disability Among Elderly People: Assessing the Evidence in 12 OECD Countries and the Future Implications" OECD Health Working Papers 26, OECD Directorate for Employment, Labour and Social Affairs.

Laibson, D. (1998), "Life-cycle consumption and hyperbolic discount functions", *European Economic Review*, 42: 861-871.

Lantz, P.M., J.S. House, J.M. Lepkowski, D.R. Williams, R.P. Mero, and J. Chen (1998), "Socioeconomic Factors, Health Behaviors, and Mortality - Results From a Nationally

Representative Prospective Study of US Adults", *Journal of the American Medical Association*, 279(21): 1703-1708.

Le Galès, C., C. Buron, N. Costet, S. Rosman, and G. Slama (2002), "Development of a preference-weighted health status classification system in France: The Health Utilities Index 3", *Health Care Management Science*, 5(1): 41-51.

Lechner, M. and R. Vázquez-Álvarez (2004), "The effect of disability on labour market outcomes in Germany: evidence from matching", CEPR Discussion Paper No. 4223.

Lee, D.S. (2008), "Randomized experiments from non-random selection in U.S. House elections", *Journal of Econometrics*, 142(2): 675-697.

Lee, D.S. and D. Card (2008), "Regression Discontinuity Inference with Specification Error", *Journal of Econometrics*, 142(2): 655-674.

Lee, D.S. and T. Lemieux (2009), "Regression Discontinuity Designs in Economics", NBER Working Paper 14723.

Liljas, B. (1998), "The demand for health with uncertainty and insurance", *Journal of Health Economics*, 17(2): 153-170.

Liljas, B. (2000), "Insurance and imperfect financial markets in Grossman's demand for health model - a reply to Tabata and Ohkusa", *Journal of Health Economics*, 19(5): 821-827.

Lindeboom, M. and E. van Doorslaer (2004), "Cut-point shift and index shift in self-reported health", *Journal of Health Economics*, 23: 1083-1099.

Lindeboom, M., A. Llena-Nozal, and B. van der Klaauw (2007), "Health Shocks, Disability and Work", Netspar Discussion Paper 2007 - 026.

Link, B.G. and J. Phelan (1995), "Social Conditions As Fundamental Causes of Disease", *Journal of Health and Social Behavior*, 35: 80-94.

Lleras-Muney, A. (2005), "The Relationship Between Education and Adult Mortality in the United States", *Review of Economic Studies*, 72: 189-221.

Lleras-Muney, A. and F. Lichtenberg (2002), "The Effect of Education on Medical Technology Adoption: Are the More Educated More Likely to Use New Drugs", NBER Working Paper, No. 9185.

Lynch, J.W., G.A. Kaplan, and S.J. Shema (1997), "Cumulative Impact of Sustained Economic Hardship on Physical, Cognitive, Psychological, and Social Functioning", *New England Journal of Medicine*, 337(26): 1889-1895.

Lynch, S.M. (2003), "Cohort and life-course patterns in the relationship between education and health: A hierarchical approach", *Demography*, 40: 309-31.

MacFarlan, M. and H. Oxley (1996), "Social transfers: spending patterns, institutional arrangements and policy responses", OECD Economic Studies, 27, 147-94.

Mackenbach, J.P., A.E. Kunst, A.E. Cavelaars, F. Groenhof, and J.J. Geurts (1997), "Socioeconomic inequalities in morbidity and mortality in Western Europe", *Lancet*, 349:1655-1659.

Mackenbach, J.P., M.J. Bakker, A.E. Kunst, and F. Diderichsen (2002), "Socioeconomic inequalities in health in Europe: An overview", In: J.P. Mackenbach, & M.J. Bakker (Eds.), Reducing Inequalities in Health: A European Perspective (3-24). London and New York: Routledge.

Mackenbach, J.P., M. Huisman, O. Andersen, M. Bopp, J.K. Borgan, C. Borrell, et al. (2004), "Inequalities in lung cancer mortality by the educational level in 10 European populations", *European Journal of Cancer*, 40(1): 126-135.

Mackenbach, J.P., I. Stirbu, A-JR. Roskam, M.M. Schaap, G. Menvielle, M. Leinsalu, and A.E. Kunst (2008), "Socioeconomic inequalities in health in 22 European countries", *New England Journal of Medicine*, 358(23): 2468-2481.

Mandemakers, C.A. (1996), Gymnasiaal en middelbaar onderwijs: ontwikkeling, structuur, sociale achtergrond en schoolprestaties, Nederland, ca. 1800-1968. Doctoral disser-

tation, Erasmus University Rotterdam.

Manton, K.G., X.L. Gu, and V.L. Lamb (2006), "Change in chronic disability from 1982 to 2004/2005 as measured by long term changes in function and health in the U.S. elderly population", *Proceedings of the National Academy of Sciences*, 103(48): 18374-18379.

Marmot, M. (1999), "Multi-level approaches to understanding social determinants", in Lisa Berkman and Ichiro Icawachi (eds.), Social Epidemiology, Oxford: Oxford University Press.

Marmot, M. (2002), "The Influence Of Income On Health: Views Of An Epidemiologist", *Health Affairs*, 21(2): 31-46.

Marmot, M. (2007), "Achieving health equity: from root causes to fair outcomes", *Lancet*, 370: 1153-1163.

Marmot, M., M.J. Shipley, and G. Rose (1984), "Inequalities in death - specific explanations of a general pattern?" *Lancet* i: 1003-1006.

Marmot, M., G. Davey Smith, S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, and A. Feeney (1991), "Health inequalities among British civil servants: the Whitehall II study", *Lancet*, 337(8754): 1387-1393.

Marmot, M. and M. Shipley (1996), "Do socioeconomic differences in mortality persist after retirement? 25 Year follow up of civil servants from the first Whitehall study", *British Medical Journal*, 313: 1177-1180.

Marmot, M., H. Bosma, H. Hemmingway, E. Brunner, and S. Stansfeld (1997a), "Contribution of job control and other risk factors to social variations in coronary hearth disease incidence", *Lancet*, 350(9073): 235-239.

Marmot, M., C.D. Ryff, L.L. Bumpass, M. Shipley, and N.F. Marks (1997b), "Social inequalities in health: next questions and converging evidence", *Social Science and Medicine*, 44(6): 901-910.

Mazumder, B. (2008), "Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws", *Federal Reserve Bank of Chicago Economic Perspectives*, Q2: 2-16.

McWhinnie, J.R. (1982), "Measuring disability", OECD Social Indicator Development Programme, Special Studies, 5. Paris: OECD.

Meijsen, J.H. (1976), "Lager onderwijs in de spiegel der geschiedenis", 's Gravenhage, the Netherlands

Mincer, J.A. (1974), *Schooling, Experience, and Earnings*, Columbia University Press.

Moller Dano, A. (2005), "Road Injuries and Long-Run Effects on Income and Employment", *Health Economics*, 14(9): 955-970.

Murasko, J.E. (2008), "An evaluation of the age-profile in the relationship between household income and the health of children in the United States", *Journal of Health Economics*, 27(6): 1489-1502.

Muurinen, J-M. (1982), "Demand for health: A generalised Grossman model", *Journal of Health Economics*, 1(1): 5-28.

Muurinen, J-M. and J. Le Grand (1985), "The Economic Analysis of Inequalities in Health", *Social Science and Medicine*, 20(10): 1029-1035.

Neuman, K. (2008), "Quit Your Job and Get Healthier? The Effect of Retirement on Health", *Journal of Labor Research*, 29(2): 177-201.

OECD (2003), *Transforming Disability into Ability: Policies to Promote Work and Income Security for Disabled People*, OECD: Paris.

OECD (2009), *Sickness, disability and work. Keeping on track in the economic downturn*, OECD: Paris.

Olson, C.A. (1981), "An Analysis of Wage Differentials Received by Workers on Dangerous Jobs", *Journal of Human Resources*, 16(2): 167-185.

Oreopoulos, P. (2006), "Estimating average and local average treatment effects of education when compulsory schooling laws really matter", *American Economic Review*, 96(1): 152-175.

Pappas, G., S. Queen, W. Hadden, and G. Fisher (1993), "The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986", *New England Journal of Medicine*, 329: 103-109.

Parsons, D. (1977), "Health, Family Structure, and Labor Supply," *American Economic Review*, 67(4): 703-712.

Pauly, M.V., H. Kunreuther, and R. Hirth (1995), "Guaranteed renewability in insurance", *Journal of Risk and Uncertainty*, 10: 143-156.

Perri, T.J. (1984), "Health Status and Schooling Decisions of Young Men", *Economics of Education Review*, 3(3): 207-213.

Peto, R., A.D. Lopez, J. Boreham, M.J. Thun, and C.W. Heath (1992), "Mortality from tobacco in developed countries: indirect estimation from national vital statistics", *Lancet*, 339: 1268-1278.

Phelan J.C. and B.G. Link (2005), "Controlling disease and creating disparities: a fundamental cause perspective", *Journal of Gerontology*, series B 60(2): 27-33.

Preston S.H. and I.T. Elo (1995), "Are educational differentials in adult mortality increasing in the United States?", *Journal of Aging and Health*, 7: 476-496.

Rawls, J. (1971), *A theory of justice*, Cambridge, MA: Harvard University Press.

Rice, N., J. Roberts, and A. Jones (2006), "Sick of work or too sick to work? Evidence on health shocks and early retirement from the BHPS", Health, Econometrics and Data Group Working Paper 2006/13. University of York.

Riphahn, R. (1999), "Income and employment effects of health shocks. A test case for the German welfare state", *Journal of Population Economics*, 12:363-389.

Robert, S.A. (1998), "Community-Level Socioeconomic Status Effects on Adult Health", *Journal of Health and Social Behavior*, 39(1): 18-37.

Rosenbaum, P. and D.B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70: 41-55.

Ross, C.E. and C.L. Wu (1996), "Education, age, and the cumulative advantage in health", *Journal of Health and Social Behavior*, 37:104-20.

Ruhm, C.J. (2003), "Good times make you sick", *Journal of Health Economics*, 24: 637-658.

Ruhm, C.J. (2005), "Healthy Living in Hard Times", *Journal of Health Economics*, 24(2); 341-363.

Schrijvers, C.T.M., H.D. van de Mheen, K. Stronks, and J.P. Mackenbach (1998), "Socioeconomic inequalities in health in the working population: the contribution of working conditions", *International Journal of Epidemiology*, 27: 1011-1018.

Seierstad, A. and K. Sydsaeter (1977), "Sufficient Conditions in Optimal Control Theory", *International Economic Review*, 18(2): 367-391.

Seierstad, A. and K. Sydsaeter (1987), "Optimal Control Theory With Economic Applications", In: Advanced Textbooks in Economics, Volume 24, Series Editors: C.J. Bliss and M.D. Intriligator, Elsevier, North Holland.

Sen, A.K. (1985), *Commodities and Capabilities*, Amsterdam: North-Holland.

Sen, A.K. (2002), "Why Health Equity?", *Health Economics*, 11: 659-666.

Shan, T.P., J.K. Wendt, R.P. Donnelly, G. de Jong, and F.S. Ahmed (2005), "Age at Retirement and Long Term Survival of an Industrial Population: Pospective Cohort Study",

British Medical Journal, 331:995, doi:10.1136/bmj.38586.448704.EO.

SHARE (2007), "Documentation of generated variables in SHARE release 2.0.1" Retrieved from http://www.share-project.org/t3/share/index.php?id=74

Silles, M. (2009), "The causal effect of education on health: evidence from the United Kingdom", *Economics of Education Review*, 28(1): 122-128.

Skalicka, V., F. van Lenthe, C. Bambra, S. Krokstad, and J.P. Mackenbach (2009), "Material, psychosocial, behavioural and biomedical factors in the explanation of relative socio-economic inequalities in mortality: evidence from the HUNT study", *International Journal of Epidemiology*, 38(5): 1272-1284.

Smith, A. (1776), "An inquiry into the nature and causes of the wealth of nations", Reprinted Roy H. Campbell and Andrew S. Skinner, eds. Oxford: Clarendon Press, 1976.

Smith J.P. (1998), "Socioeconomic status and health", *American Economic Review*, 88(2): 192-196.

Smith, J.P. (1999), "Healthy bodies and thick wallets: the dual relation between health and socioeconomic status", *Journal of Economic Perspectives*, 13: 145-166.

Smith, J.P. (2005a), "Unraveling the SES-health connection", *Population and Development Review*, 30: 108-132.

Smith, J.P. (2005b). The impact of SES on health over the life course, RAND working paper, WR-318. Santa Monica: RAND.

Smith, J.P. (2005c), "Consequences and Predictors of New Health Events" In Analyses in the Economics of Aging, ed. David A. Wise, 213-37. Chicago: University of Chicago Press and National Bureau of Economic Research.

Smith, J.P. (2007), "The impact of SES on health over the life-course", *Journal of Human Resources*, 42(4): 739-764.

Smith, R.S. (1978), "Compensating Wage Differentials and Public Policy: A Review", *Industrial and Labor Relations Review*, 32(3): 339-352.

Smith, J.A. and P.E. Todd (2005), "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?", *Journal of Econometrics*, 125(1): 305-353.

Smith, J.P. and R. Kington (1997), "Demographic and Economic Correlates of Health in Old Age", *Demography*, 34: 159-170.

Spasojevic, J. (2003), *Effects of Education on Adult Health in Sweden: Results from a Natural Experiment*, PhD thesis. New York: City University of New York Graduate Center.

Staiger, D. and J.H. Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65(3): 557-586.

Sterringa, G. (1934), *Leerplicht en leerplichtwet*, J.B. Wolters Uitgevers-Maatschappij, Groningen, The Netherlands.

Stock, J.H. and M. Yogo (2002), "Testing for weak instruments in linear IV Regression" NBER Technical Working Paper 284 (Revised 2004).

Suhrcke, M., M. McKee, D. Stuckler, R. Sauto Arce, S. Tsolova, and J. Mortensen (2006), "The contribution of health to the economy in the European Union", *Public Health*, 120: 994-1001.

Sutton, M. (2004), "The effect of income on health: evidence from a British pseudo-panel" ECuity III Project Working paper 10, University of Glasgow.

Thaler, R. and S. Rosen (1975), "The value of saving a life: evidence from the labor market", in: N. Terleckyj (ed.), Household production and consumption, New York: National Bureau of Economic Research.

Thistlethwaite, D.L. and D.T. Campbell (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment", *Journal of Educational Psychology*, 51:

309-317.

Trochim, W.M.K. (1984), *Research Design for Program Evaluation: the Regression-Discontinuity Approach*, Beverly Hills, CA.

United Nations (2000), General Comment 14: The right to the highest attainable standard of health. Geneva: Office of the United Nations High Commissioner for Human Rights.

US Department of Health and Human Services (1989), "Reducing the health consequences of smoking: 25 years of progress", A report of the Surgeon General, Rockville, Maryland: Public Health Service, Centers for Disease Control, Office on Smoking and Health, 1989, (DHHS Publication No (CDC) 89-8411.).

Van den Berg G., M. Lindeboom, and F. Portrait (2006), "Economic Conditions Early in Life and Individual Mortality", *American Economic Review*, 96(1): 290-302.

Van den Berg, G. and M. Lindeboom (2007), "Birth is the messenger of death - but policy may help to postpone the bad news" Netspar panel paper, 4. Tilburg: Netspar.

Van der Klaauw, W. (2002), "Estimating the effect of financial aid offers on college enrollment: a Regression Discontinuity Approach", *International Economic Review*, 43(4): 1249-1287.

Van der Klaauw, W. (2008), "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics", *Labour*, 22(2): 219-245.

Van der Lecq, F. and A. Oerlemans (2009), "Zelfstandingen zonder pensioen", Netspar NEA Paper 24.

Van Doorslaer, E. and A.M. Jones (2003), "Inequalities in self-reported health: validation of a new approach to measurement", *Journal of Health Economics*, 22: 61-87.

Van Doorslaer, E. and X. Koolman (2004), "Explaining the differences in income-related health inequalities across European countries", *Health Economics*, 13: 609-628.

Van Doorslaer, E., A. Wagstaff, H. Bleichrodt, S. Calonge, U-G. Gerdtham, et al. (1997), "Income-related inequalities in health: some international comparisons", *Journal of Health Economics*, 16(1): 93-112.

Van Doorslaer, E., H van Kippersluis, O. O'Donnell, and T. Van Ourti (2008), "Socioeconomic Differences in Health over the Life Cycle: Evidence and Explanations", Netspar Panel paper prepared for the Netspar Panel Day 16 October, 2008, Tilburg.

Van Kippersluis, H., T. Van Ourti, O. O'Donnell, and E. van Doorslaer (2008), "Health and income across the life cycle and generations in Europe", Tinbergen Institute Discussion Paper Series 08-009/3.

Van Kippersluis, H., T. Van Ourti, O. O'Donnell, and E. van Doorslaer (2009), "Health and income across the life cycle and generations in Europe", *Journal of Health Economics*, 28(4): 818-830.

Van Kippersluis, H., O. O'Donnell, E. van Doorslaer, and T. Van Ourti, (2010), "Socioeconomic Differences in Health over the Life Cycle in an Egalitarian Country", *Social Science and Medicine*, 70(3): 428-438.

Van Lenthe, F.J., E. Gevers, I.M. Joung, H. Bosma, and J.P. Mackenbach (2002), "Material and behavioral factors in the explanation of educational differences in incidence of acute myocardial infarction: the Globe study", *Annals of Epidemiology*, 12(8): 535-542.

Van Oort, F.V., F.J. van Lenthe, and J.P. Mackenbach (2005), "Material, psychosocial, and behavioural factors in the explanation of educational inequalities in mortality in The Netherlands", *Journal of Epidemiology and Community Health*, 59(3): 214-220.

Van Ourti, T. (2003), "Socio-economic inequality in ill-health amongst the elderly. Should one use current or permanent income?", *Journal of Health Economics*, 22: 219-241.

Viscusi, K.W. (1978), "Wealth Effects and Earnings Premiums for Job Hazards", *Review of Economics and Statistics*, 60: 408-416.

Viscusi, K.W. (1979), *Employment hazards: An investigation of market performance*, Cambridge: Harvard U. Press.

Viscusi, K.W. (1993), "The Value of Risks to Life and Health", *Journal of Economic Literature*, 31(4): 1912-1946.

Wagstaff, A., P. Paci, and E. van Doorslaer (1991), "On the measurement of inequalities in health", *Social Science and Medicine*, 33(5): 545-557.

Wagstaff, A. (1986), "The demand for health: Some new empirical evidence", *Journal of Health Economics*, 5: 195-233.

Weiss, Y. and L.A. Lillard (1978), "Experience, vintage and time effects in the growth of earnings: American Scientists, 1960-1970", *Journal of Political Economy*, 86: 427-47.

Willson, A., K. Shuey, and G. Elder (2007), "Cumulative advantage processes as mechanisms of inequality in life course health", *American Journal of Sociology*, 112(6): 1886-1924.

Wolfe, J.R. (1985), "A Model of Declining Health and Retirement", *Journal of Political Economy*, 93(6): 1258-1267.

Wu, S. (2003), "The Effects of Health Events on the Economics Status of Married Couples", *Journal of Human Resources*, 38(1): 219-230.

Yen, I.H. and G.A. Kaplan, G.A. (1999), "Neighborhood Social Environment and Risk of Death: Multilevel Evidence from the Alameda County Study", *American Journal of Epidemiology*, 149(10): 898-907.

Zweifel, P. and F. Breyer (1997), *Health Economics*, Oxford University Press, New York.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

419. L. RATNOVSKI, *A Random Walk Down the Lombard Street: Essays on Banking.*

420. R.P. NICOLAI, *Maintenance models for systems subject to measurable deterioration.*
421. R.K. ANDADARI, *Local clusters in global value chains, a case study of wood furniture clusters in Central Java (Indonesia).*
422. V. KARTSEVA, *Designing Controls for Network Organizations: A Value-Based Approach.*
423. J. ARTS, *Essays on New Product Adoption and Diffusion.*
424. A. BABUS, *Essays on Networks: Theory and Applications.*
425. M. VAN DER VOORT, *Modelling Credit Derivatives.*
426. G. GARITA, *Financial Market Liberalization and Economic Growth.*
427. E. BEKKERS, *Essays on Firm Heterogeneity and Quality in International Trade.*
428. H. LEAHU, *Measure-Valued Differentiation for Finite Products of Measures: Theory and Applications.*
429. G. BALTUSSEN, *New Insights into Behavioral Finance.*
430. W. VERMEULEN, *Essays on Housing Supply, Land Use Regulation and Regional Labour Markets.*
431. I.S. BUHAI, *Essays on Labour Markets: Worker-Firm Dynamics, Occupational Segregation and Workplace Conditions.*
432. C. ZHOU, *On Extreme Value Statistics.*
433. M. VAN DER WEL, *Riskfree Rate Dynamics: Information, Trading, and State Space Modeling.*
434. S.M.W. PHLIPPEN, *Come Close and Co-Create: Proximities in pharmaceutical innovation networks.*
435. A.V.P.B. MONTEIRO, *The Dynamics of Corporate Credit Risk: An Intensity-based Econometric Analysis.*
436. S.T. TRAUTMANN, *Uncertainty in Individual and Social Decisions: Theory and Experiments.*
437. R. LORD, *Efficient pricing algorithms for exotic derivatives.*
438. R.P. WOLTHOFF, *Essays on Simultaneous Search Equilibrium.*
439. Y.-Y. TSENG, *Valuation of travel time reliability in passenger transport.*
440. M.C. NON, *Essays on Consumer Search and Interlocking Directorates.*
441. M. DE HAAN, *Family Background and Children's Schooling Outcomes.*
442. T. ZAVADIL, *Dynamic Econometric Analysis of Insurance Markets with Imperfect Information*

443. I.A. MAZZA, *Essays on endogenous economic policy*
444. R. HAIJEMA, *Solving large structured Markov Decision Problems for perishable-inventory management and traffic control*
445. A.S.K. WONG, *Derivatives in Dynamic Markets*
446. R. SEGERS, *Advances in Monitoring the Economy.*
447. F.M. VIEIDER, *Social Influences on Individual Decision Making Processes.*
448. L. PAN, *Poverty, Risk and Insurance: Evidence from Ethiopia and Yemen.*
449. B. TIEBEN, *The concept of equilibrium in different economic traditions: A Historical Investigation.*
450. P. HEEMEIJER, *Expectation Formation in Dynamic Market Experiments.*
451. A.S. BOOIJ, *Essays on the Measurement Sensitivity of Risk Aversion and Causal Effects in Education.*
452. M.I. LÓPEZ YURDA, *Four Essays on Applied Microeconometrics.*
453. S. MEENTS, *The Influence of Sellers and the Intermediary on Buyers' Trust in C2C Electronic Marketplaces.*
454. S. VUJIĆ, *Econometric Studies to the Economic and Social Factors of Crime.*
455. F. HEUKELOM, *Kahneman and Tversky and the Making of Behavioral Economics.*
456. G. BUDAI-BALKE, *Operations Research Models for Scheduling Railway Infrastructure Maintenance.*
457. T.R. DANIËLS, *Rationalised Panics: The Consequences of Strategic Uncertainty during Financial Crises.*
458. A. VAN DIJK, *Essays on Finite Mixture Models.*
459. C.P.B.J. VAN KLAVEREN, *The Intra-household Allocation of Time.*
460. O.E. JONKEREN, *Adaptation to Climate Change in Inland Waterway Transport.*
461. S.C. GO, *Marine Insurance in the Netherlands 1600-1870, A Comparative Institutional Approach.*
462. J. NIEMCZYK, *Consequences and Detection of Invalid Exogeneity Conditions.*
463. I. BOS, *Incomplete Cartels and Antitrust Policy: Incidence and Detection.*
464. M. KRAWCZYK, *Affect and risk in social interactions and individual decision-making.*
465. T. C. LIN, *Three Essays on Empirical Asset Pricing.*
466. J.A. BOLHAAR, *Health Insurance: Selection, Incentives and Search.*
467. T. FARENHORST-YUAN, *Efficient Simulation Algorithms for Optimization of Discrete Event Based on Measure Valued Differentiation.*
468. M.I. OCHEA, *Essays on Nonlinear Evolutionary Game Dynamics.*